Training Researchers in the Use of Statistics

IASE Round Table Conference

TOKYO 2000

Edited by

Carmen Batanero

International Association for Statistical Education

International Statistical Institute

Training Researchers in the Use of Statistics

IASE Round Table Conference, Tokyo 2000

Training Researchers in the Use of Statistics

Edited by:

Carmen Batanero

University of Granada, Spain

International Association
for Statistical Education

International
Statistical Institute

ISBN 90-73592-19-4

Editor: Carmen Batanero

Published by: International Association for Statistical Education
International Statistical Institute

ISI Permanent Office
428 Prinses Beatrixlaan
PO Box 950
2270 AZ Voorburg
The Netherlands

Cover Design: René Keijser, International Statistical Institute

Printed by: Servicio de Reprografía de la Facultad de Ciencias
Universidad de Granada
Avenida Fuentenueva s/n 18071 Granada, Spain

TABLE OF CONTENTS

FOREWORD

There is no doubt that the training of researchers in the use of statistics is very important to improve the quality of empirical research and to foster technical and economical development. Empirical sciences, in general, rely heavily on the statistical analysis of data, in particular on inferential statistics. However, since the logic of statistical inference is difficult to grasp, its use and interpretation are not always adequate and have been criticised for nearly 50 years.

Books, such as the classical work by Morrison and Henkel, *The Statistical Tests Controversy* published in 1970 by Aldine or the more recent book by Harlow, Mulaik and Steiger, *What if There Were No Significance Tests?*, published in 1997 by Erlbaum, and a wide related literature suggest that scientists often apply statistics in a mechanical way, that they are paying too much attention to the significance tests' results and are forgetting other statistical methods, as well as the estimation of the magnitudes of the effects they are investigating.

Given the increasing complexity and new methods emerging in the field of statistics, it is difficult for researchers to bring themselves up-to date in the new statistical procedures and, therefore, there is a clearly increasing need for the statistician's support in empirical research. However, in many cases, the statistician is only consulted after the data has already been collected, with the inherent consequences of poor experimental design, biased sampling methods or inadequate data. The ready accessibility of statistical computer programs may even provide an excuse for not consulting a statistician at all in some cases, to the subsequent detriment of the scientific quality of the research.

The controversy about the adequate use of statistics has recently increased within professional organisations such as the American Educational Research Association or the American Psychological Association, which are suggesting important shifts in their editorial policies regarding the use of statistics and are recommending better use of statistical language in reported research. Research journals in medicine such as the *British Medical Journal* or *Statistics in Medicine* have highlighted the poor quality of methodology and statistics in medical research. This debate is also reflected in research journals in psychology, and education (e.g., in *Educational Researcher, Mathematics Education Research Journal, Theory and Psychology, Educational and Psychological Measurement* or *Educational Psychological Review).*

Some explanations suggested in these debates for the persistence of an inadequate or inappropriate use of statistics include inertia, conceptual confusion, lack of better alternative tools, and psychological mechanisms such as invalid generalisation from deductive logic to inference under uncertainty.

Within the International Association for Statistical Education it was felt that there was an educational problem at the root of this dilemma and that the training of researchers should be analysed by statistics educators from its various perspectives and approaches. As a result a *Round Table Conference on The Training of Researchers in the Use of Statistics* was organised at the Institute of Statistical Mathematics in Tokyo, August 7-11, 2000, where researchers from around the world met to discuss the topic and to produce suggestions about possible ways in which statistical education might contribute to the

better understanding and application of statistics in applied research. This conference on the training of researchers also served to illustrate the need to understand statistical concepts and to develop statistical thinking at all educational levels.

The present book includes the works and discussions at the Conference and it is one of a series of publications resulting from the International Statistical Institute and International Association for Statistical Education Round Table Conferences. The first Round Table Conference was organised by the *ISI Education Committee* in 1968 in The Hague on the theme "The University Teaching of Statistics in Developing Countries". Subsequent ISI Round Table Conferences dealt with "New Technologies in Teaching Statistics" (Oisterwijjk, The Netherlands, 1970), "Statistics at the School Level", (Vienna, 1973), "The Teaching of Statistics in Schools" (Warsaw, 1975), "Teaching of Statistics in the Computer Age" (Canberra, 1984), "Training Teachers to Teach Statistics" (Budapest, 1988), and "Teaching Data Analysis in Schools. Who should Teach it and How?" (Quebec, 1992). The first IASE Round Table Conference was organised in 1996 in Granada, Spain on "Research on the Role of Technology on Research in Statistics Education" and it is planned to continue this series of conferences focusing in new educational topics every four years.

The Tokyo conference was not an isolated event, but a part of a long international collaborative process, which started in 1997, when the IASE Executive Committee decided the topic for this particular Round Table Conference and later nominated Carmen Batanero to organise its scientific committee. In early 1998 the Scientific Committee was in place and they started to produce a Discussion Document describing the aims of the conference and suggesting the main points to be discussed. Members of the Scientific Committee and the IASE Executive Committee met during the ICOTS-V Conference in Singapore, August 1998 to revise the document.

The Discussion Document, which is reprinted in Part I of this book, was published in the *Newsletter of the International Study Group for Research in Statistics Education* and located at the Conference web site in October, 1998. Shorter versions were also published in the *ISI Newsletter*, *Teaching Statistics* and *IASE Review*. From October 1998 to November 1999, a Call for papers for the conference was announced through the IASE and ISI publications, and through a number of statistics and mathematics education journals; its theme attracted wide interest on the part of statisticians and statistics educators. Proposals for paper presentations were solicited for November1999 by the Scientific Committee.

The IASE organised a refereeing process with the participation of specialists from different countries to assure fairness and quality in the process of reviewing and selecting the papers to be presented at the conference among the many proposals received. The authors of each accepted proposal were required to complete a preliminary paper by May 1, 2000. These preliminary papers were then put on the web, so that those attending the Round Table Conference could download and read the papers of the various presenters before the actual meeting. The Scientific Committee classified the papers accepted in a number of categories to produce the conference programme and invited some additional participants (both statisticians and statistics educators) to act as reactors for every set of related papers.

The conference was sponsored by the IASE and the ISI, the Institute of Statistical Mathematics in Tokyo and the Japan Statistical Society. The 48 participants who met during five days, included professional and official statisticians, lecturers, researchers and statistics educators with experience in teaching, research or consultancy in different

areas of application (see the list of participants in part I). The participants represented different countries of the five continents, as well as developed and developing countries.

At the conference, the papers were presented and debated and, after the conference, a summary of the discussions was sent to the authors who were given some additional time and were asked to produce a revision of the papers, taking into account the suggestions and discussions held.

As a result of this process we are happy to present today this set of contributions. "Training Researchers in the Use of Statistics" is not a simple topic. In this book the reader will find various analysis of the problems related to this training, and a number of views of ways in which some of these problems might be solved: The controversies on the use of statistics in research; the researchers' attitudes towards statistics; the challenges set by technology; the particular needs of training in specific research fields; the problems of communication among statisticians and researchers are just a few examples of didactical problems that are discussed.

The book is organised in six parts. After the introductory section, which contains the materials related to the conference, several chapters offer a broad view of the didactical problems related to the training of researchers: Part 2 describes the problems related to the training of researchers in particular statistical topics. In Part 3 the challenges set by technology and how this affects the training of researchers is discussed. Part 4 deals with the particular training needs of researchers in areas such as education, social sciences, medicine or biology. International successful experiences to solve challenging problems in the training of researchers are described in Part 5 and Part 6 discusses the didactical problems that underlay statistical consultation. A final chapter presents a synthesis of the main conclusions in the Conference.

The themes and ideas that emerge from these papers should be considered as suggestions for further research more than as definitive answers. There are so many unanswered questions about what the best ways of training future and current researchers are, about how can we change their views and attitudes towards statistics and about how best we can collaborate with them best.

We hope this book will serve as a starting point for other lecturers, researchers and statistics educators to reflect on the statistical training of researchers in empirical sciences, to change their teaching approaches, to improve the interest to collaborate in applied research and to start new didactical research on some of the problems described.

ACKNOWLEDGEMENTS

We would like to express our recognition to the different people and institutions that made the conference and the publication of this book possible.

A very important role was played by the members of the Scientific Committee, professors Theodore Chadjipadelis (Greece), Joan B. Garfield (USA), Yuki Miura (Japan), David Ospina (Colombia) and Brian Phillips (Australia) who helped the Chair Carmen Batanero (Spain), and supported her decisions throughout the whole planning period, and during the conference. Other members of the IASE Executive Committee M. Gabriella Ottaviani, Dani Ben-Zvi, Gilberte Schuyten, and Lionel Pereira-Mendoza also contributed with their advice and experience to important decisions about the conference.

Both committees are very grateful to the referees who were willing to contribute to this collective study with their expertise in different fields and topics and who provided many valuable comments to help the authors in developing their papers.

The IASE is also grateful to the ISI for supporting this conference and also to the Institute of Mathematical Statistics, and its director Dr. Shimizu by offering to hold this event and for their generous support. It was a great honour for participants to be hosted by this Institute, which is a leader in the statistical practice and research in Japan and at an international level. The IASE also thanks the Japan Statistical Society, in particular the Education Section, and Professors Yuki Miura, Masakatsu Murakami, Toshiro Shimada, and Kensey Araya and the members of the Institute who attended the meeting and the Institute staff for their work in arranging all the details of the local organisation, which served to make the conference more enjoyable and productive.

We are especially grateful to all the authors, discussants and observers, who shared their experiences, research, international projects surveys and reflections on the teaching of statistics and on consultancy in different fields of research and in different statistical topics.

I also wish to thank René Keijser of the International Statistical Institute for designing the cover. And finally and not least important we are indebted to Angela Barnie for her work and patience in helping to revise this manuscript.

Carmen Batanero

PART 1

THE IASE ROUND TABLE CONFERENCE

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 11-19.
ã 2001, International Association for Statistical Education and International Statistical Institute.
Printed in Granada, Spain.

THE IASE ROUND TABLE CONFERENCE

Since 1968, a number of Round Table Conferences have been organised on statistics education topics, initially by the Education Committee of the International Statistical Institute and, since 1996, by IASE (the International Association for Statistical Education). Since 1988 these conferences have been held as satellite meetings to each ICME (International Congress on Mathematics Education).

The goal of the Round Table Conferences is to bring together a small number of experts, representing as many different countries as possible, to share their scholarly work on a given topic area. The round table meetings provide opportunities for developing better mutual understanding of common problems, and for making recommendations concerning the topic area under discussion. A main outcome is a monograph containing a set of refereed papers, which have been prepared for, and discussed during, the conference.

In this book, we present the monograph resulting from the IASE Round Table Conference on Training Researchers in the Use of Statistics, which was held in Tokyo in August 2000. This collection of papers includes an overview of the conference subject, and recommendations for future work in this area.

In this chapter we include the original conference Discussion Document, a list of committee' members and participants and a brief description of the Institute of Statistical Mathematics, where the roundtable was held. The papers in the following chapters were initially prepared in response to the Discussion Document, and were revised both before the conference as a consequence of the referees' comments and after the roundtable to take into account the discussions and suggestions at that meeting.

DISCUSSION DOCUMENT

The following people contributed to this document: Carmen Batanero, Theodore Chadjipadelis, Joan B. Garfield, Yuki Miura, David Ospina, M. Gabriella Ottaviani, and Brian Phillips.

Researchers in different sciences need to collect and analyse data about the phenomena which they study. Conducting empirical research is an exercise that requires conceptual, practical, and also applied statistical skills. The arrival of computers has led to diffusion and intensive use of new statistical techniques, and their application to the analysis of progressively more diverse data sets in a growing number of disciplines. As a result, statistics has become a fundamental tool for experimental researchers, many of whom lack the necessary training in statistics.

Researchers frequently bring statisticians into their research teams. This has two advantages. It helps to ensure that complex data are correctly analysed. It also enables the researcher to learn of new developments in statistical procedures and software tools. However, experimental data analysis and solving practical problems cannot be considered to be solely the responsibility of statisticians. Some fundamental research design issues, and decisions within the data analysis process, as well as the final interpretation of results, require a knowledge of the specific discipline area that is

generally deeper than that which a statistician is able to contribute.

Inappropriate attitudes or lack of knowledge on the part of researchers about the central role that statistics can play in research may affect the communication and collaboration between statisticians and researchers in other fields. Psychological heuristics and biases also affect the processes of decision taking and interpreting of random experiment results. Researchers, for example, sometimes overestimate the power of their research methods, or attribute too much confidence to the reliability (replicability) of their findings.

As a consequence, statistics courses can be seen to be a crucial part of the general training received by new researchers. Within many masters and doctoral programmes in other disciplines, courses on data analysis concepts and procedures and, more generally, on statistical reasoning and research methodology are included. Therefore, the study of difficulties and obstacles that new researchers face when learning the contents of these courses should be of fundamental concern to statistical educators. Key issues are the existing concepts and attitudes that the researchers bring to these courses, as well as the methods by which they are expected to develop their statistical understanding.

That there are inadequacies in such training programmes may be inferred from the degree of controversy that exists within some research communities about the use of statistical tests and other statistical procedures. This controversy has increased in recent years within some professional research organisations. Important shifts have been proposed, and in some cases implemented, in editorial policies regarding the publication of statistical analysis results.

However, there is still a great deal of evidence that researchers do not have the necessary understanding of statistical concepts and processes to appreciate what methods are really appropriate to their needs. It is therefore crucial that statistical educators consider empirical findings about researchers' uses and misuses of statistics from a didactic perspective, so that they can contribute significantly to the development of appropriate training programmes for researchers.

IASE, the International Association for Statistical Education, is convening a Research Round Table on this topic. It is hoped that statistical educators across the world will wish to contribute to the international discussion and thereby extend understanding about the problems of training researchers in the use of statistics.

SOME RESEARCH QUESTIONS

The overall intention is to study the problem of training researchers in the use of statistics, in all its many dimensions. One aspect of this relates to the considerable variation that pertains in the statistical methods that the trainee researchers require, dependent upon which scientific context they will be working in. A wide-ranging discussion is needed within the statistical education community itself in order to identify prevailing views on how researchers may be optimally trained.

There is, however, also a need to develop mutual understanding of the problems by encouraging dialogue between statistical educators and representatives of other scientific disciplines. The following broad areas are expected to form the main structure of the Round Table programme and its deliberations.

1. *What are the specific statistical competencies that researchers in different disciplines should acquire in their postgraduate training?*

   Statistics is an important component in the training of new researchers within

masters and doctorate courses. There are a number of reasons for including a statistical component in such training. Scientists need: a basic knowledge of statistics to carry out simple statistical tasks themselves, a level of understanding that allows them to identify statistical errors in research literature, and the ability to recognise situations where they should seek a statistician's help.

2. *What are the particular needs and problems in the statistical training of researchers in specific fields? Of what good examples of successful teaching experiments in specific fields do we know?*

Many domains of scientific knowledge have become highly specialised. Scientists in different disciplines, have different needs and opportunities to apply statistics. Their views about statistics may differ, according to their previous statistical and mathematical training and research backgrounds. This may be reflected in their having different training needs. Analysis of these issues, and examples of course design and evaluation could provide examples for future developments.

3. *What are the main learning problems, misconceptions and errors concerning particular advanced statistical concepts?*

Research into the learning of advanced statistics is starting to emerge. Students' difficulties on topics such as hypothesis testing, estimation, correlation, the normal distribution, are now starting to receive attention from researchers in statistical education. However, much more research on these problems, and on students' difficulties in other advanced statistics topics, is needed. It is also important to communicate the outcomes of such research to statistical educators, so that they can be taken into account in the design of research training courses.

4. *How should we design/evaluate courses for training researchers in particular statistical topics? What good examples do we have of successful teaching experiments in advanced statistics?*

Because of the speed at which statistical techniques develop and diversify, researchers frequently need to update their understanding of specific procedures or learn about new techniques that they should be employing in their research projects. One main problem is the design and evaluation of courses on topics such as multivariate analysis, Bayesian inference, sampling design, categorical data analysis, experimental design, quality control, etc., directed at new and senior researchers.

5. *What are the effects of technology on the statistical training of researchers?*

Research on the Role of Technology in Teaching and Learning Statistics was the theme of the previous Round Table Conference held in Granada in 1996. The conference demonstrated that educational technology does afford us with a large variety of strategies for teaching statistics. However, a number of problems were also identified, e.g. inadequate preparation of teachers, inappropriate curriculum content and structure, or lack of appropriate assessment methods. The extent to which these possibilities and problems could affect the training of researchers in the use of statistical software, as well as the training needs of researchers, will be addressed.

6. *What errors are frequent in the use of statistics by researchers?*

Because the logic of statistical inference is difficult to grasp, its use and interpretation is not always adequate. It is important for the scientific community to be made aware of common errors and misinterpretations in the use of statistics.

7. *How the researchers attitudes towards statistics affect the current role of data analysis in experimental research?*

The role of statistics in research is sometimes conditioned by researchers' views of statistics and the utility of statistics, including overestimating the power of data analysis, considering statistics as a mechanistic process, or complete rejection of the quantitative approach to research. The task of assessing the role of statistical methods in empirical research is also complicated by the facts that; (a) researchers use different research methods to achieve the same goals, (b) the theoretical, practical and statistical constraints on researchers differ when the researchers have different objectives, and (c) the relationship between the substantive and statistical hypotheses is often oversimplified by researchers.

8. *Could we compare consultation to a teaching/ learning process? Are there problems of communication in the collaborations between statisticians and experts in different disciplines?*

Statisticians who frequently collaborate with researchers need to understand enough about their colleagues' discipline areas as well. They must also learn to communicate statistical issues and results effectively to researchers. Problems can arise because of differences between statistical, technical scientific and everyday use of language. Statistical educators can, and should, be involved in improving the learning processes underlying the skills associated with providing statistical consultation.

9. *What statistical concepts and procedures are informally learned from reading research literature?*

Statistical language and statistical reasoning allow a specific way of communicating information and reaching conclusions. The way in which this resource is used in research literature to support researchers' results, and the learning processes involved in reading published research studies, need to be examined from the perspective of statistical education. A related topic is the need to develop criteria that can be used by researchers to evaluate the statistical methods found in research reports.

THE INSTITUTE OF STATISTICAL MATHEMATICS

Since its found as the only national institute of statistics in Japan, the Institute of Statistical Mathematics has exerted a powerful influence on the study and research of statistical science. The ever-increasing needs for statistical methods and ideas in various fields of science and technology led the Institute to reorganise itself in 1985 as an inter-university research institute which puts a major emphasis on research collaboration with all disciplines of science. In 1997, in order to foster collaborative research projects even more effectively and to intensify the impact of statistical science in academia, industry, and government, the Institute restructured its two attached centres and established positions for foreign researchers as visiting professors.

At present, the Institute consists of four departments, two centres, two councils, and a committee. The four departments: Fundamental Statistical Theory, Statistical Methodology, Prediction and Control, and Interdisciplinary Statistics form the active core of the Institute with its 55 academic staff, carrying out research either on statistical theory or on its application to other fields of science and industry.

The Department of Fundamental Statistical Theory and its four divisions address the

fundamental aspects of theoretical statistics. In the six divisions of the Department of Statistical Methodology, efforts are concentrated on improving statistical methods and creating new techniques for analysis. The seven divisions in the Department of Prediction and Control are specifically concerned with development of innovative approaches to understand and possibly control phenomena of a stochastic nature. The Department of Interdisciplinary Statistics, with its three divisions, is set up in order to transfer methodological developments to other disciplines and to receive, in turn, impetus that urges the creation of new statistical methods.

The two centres attached to the Institute have the aim of supplementing its activities. As of April 1997, their names and structures have changed in order to stimulate cross-disciplinary statistical research and to provide researchers both inside and outside the Institute with adequate computational and informational resources. The Centre for Development of Statistical Computing consists of two divisions. Together they undertake research on statistical computing, and also facilitate the use of computers and network connections. The Centre for Information on Statistical Sciences encourages research collaboration, publicises research findings through journals and the internet, and seeks to cultivate statistical thinking in the general public. The Centre has an additional division for foreign visiting professors.

In addition to the departments and centres, the Institute has a section of 12 technical staff that work on special jobs including maintenance of computer systems and bibliographical service. The Institute has an excellent library of books and journals, not only in pure statistics, but also in fields of specific interest to researchers (e.g., physics, genetics, and social sciences). Lastly there is also a division of 18 officials who take care of general affairs. The Institute devotes itself to educating young statisticians as well. As a constituent of the Graduate University for Advanced Studies (Department of Statistical Science, School of Mathematical and Physical Science), the Institute offers graduate programs leading to a Ph.D. degree.

COMMITTEES

The International Association for Statistical Education is grateful to the following people, who contributed to the organisation of the conference, the selection and revision of papers.

SCIENTIFIC COMMITTEE

Carmen Batanero, *University of Granada, Spain, Chair*
Theodore Chadjipadelis, *University of Thessaloniki, Greece*
Joan B Garfield, *University of Minneapolis, USA*
Yuki Miura, *Surugadai University, Japan*
David Ospina, *National University, Colombia*
Brian Phillips, *Swinburne University of Technology, Australia*

LOCAL ORGANISING COMMITTEE

Yuki Miura, *Surugadai University, Chair*
Kensei Araya, *Fukushima University*
Masakatsu Murakami, *The Institute of Statistical Mathematics*
Toshiro Shimada, *Professor Emeritus, Meiji University*

REFEREEING COMMITTEE

In addition to the speakers, and the members of the committees, the following people participated in the revision and selection of papers:

Helena Barcelar, Faculty of Psychology and Education, University of Lisbon, Portugal
John Bibby, QED MatheMagic, University of York, UK
Sinsup Cho, Department of Statistics, Korea National University, Korea
Regis Gras, Professor Emeritus, Institute en Recherches des Mathématiques, University of Rénnes, France
Delwyn Harnisch, Educational Psychology, University of Illinois at Urbana, Champain, USA
Tjaart Imbos, Methodology and Statistics, University of Maastrich, The Netherlands
Petter Laake, Medical Statistics, University of Oslo, Norway
Helmut Maier, Berlin School of Economics, Germany
Carlos Mate, Departament of Industrial Management, Catholic University of Comillas, Madrid, Spain
Raul Pedro Mentz, Instituto de Investigaciones Estadisticas, Universidad Nacional de Tucuman, Tucumán, Argentina
Vitali Muba, Eastern Africa Statistical Training Centre, Dar es Salaam, Tanzania
Parimal Mukhopadhyay, Department of Mathematics and Computing Science, University of South Pacific, Suva, Fiji
Joseph S. Oke, School of Business Studies, Federal Polytechnic University, Nigeria
M. Gabriella Ottaviani, Dipartimento di Statistica, Probabilita' e Statistiche Applicate, Universita' di Roma "La Sapienza", Roma, Italy
Lionel Pereira- Mendoza, National Institute of Education, Nanyang Technological University, Singapore
Eric Sowey, School of Economics, The University of New South Wales, Sidney, Australia
Susan Starkings, South Bank University, London, UK
Jeff Witmer, Department of Mathematics, Oberlin College, Oberlin, USA

ADDRESSES OF PARTICIPANTS

The following people took part in the conference:

Kensei Araya
Fukushima University,
7-9 Higyodan Kitasawamata,
Fukushima-Shi 960-8251, Japan.
E-mail: araya.k@oregano.ocn.ne.jp

Shrikant I Bangdiwala
University of North Carolina,
School of Public Health, Biostatistics,
CB# 7400 Chapel Hill, NC 27599-7400, USA
E-mail: kbangdiw@bios.unc.edu

Carmen Batanero
Universidad de Granada,
Departamento de Didáctica de la Matemática,
Campus de Cartuja, 18071 Granada, Spain.
E-mail: batanero@goliat.ugr.es

Carlos Marcos Batista
Universidade Federal de Goias - UFG FCHF,
Departamento de Ciencias Sociais,
Campus 2 - Goiania - GO Brazil.
E-mail: carlosmb@fchf.ufg.br

Gabriella Belli
Virginia Polytechnic Institute & State
University, 7054 Haycock Road,
Falls Church, VA 22043-2311, USA.
E-mail: gbelli@vt.edu

Martha, B. Bilotti-Aliaga
The University of Michigan,
Department of Statistics, 4077 Frieze
Building, Ann Arbor Michigan, 48109, USA.
E-mail: aliaga@umich.edu

Glenys Bishop
Australian Bureau of Statistics,
PO Box10, Belconner 2616, Australia.
E-mail: glenys.bishop@abs.gov.au

Carol Joyce Blumberg
Winona State University,
Department of Mathematics and Statistics,
Winona, MN 55987-5838, USA.
E-mail: cblumberg@winona.edu

Theodore Chadjipadelis
Aristotle University of Thessaloniki,
Department of Education,
54006 Thessaloniki, Greece.
E-mail: chadji@eled.auth.gr

Megan Clark
Victoria University, School of Mathematical
and Computing Sciences, P.O. Box 600,
Wellington, New Zealand.
E-mail: megan.clark@vuw.ac.nz

Antonio Estepa-Castro
Universidad de Jaen,
Facultad de Humanidades y Educación,
Campus las Lagunillas, 23071 Jaen, Spain.
E-mail: aestepa@ujaen.es

Gianfranco Galmacci
Dipartimento di Scienze Statistiche,
Università degli Studi,
Casella Postale 1315 / Succ. 1,
06123 Perugia, Italy.
E-mail: glm@stat.unipg.it

Michael J. Glencross
University of Transkei,
Research Resource Centre, Private Bag X1,
Unitra Umtata 5117, South Africa.
Email: glencros@getafix.utr.ac.za

Juan D. Godino
Universidad de Granada,
Departamento de Didáctica de la Matemática,
Campus de Cartuja 18071 Granada, Spain.
E-mail: jgodino@goliat.ugr.es

John Harraway
University of Otago,
Department of Mathematics and Statistics,
P.O. Box 56, Dunedin,
New Zealand.
jharraway@maths.otago.ac.nz

Kosaburo Hirakawa
21-2, 5-chome, Ayase, Adachi-ku,
Tokyo, 120-0005, Japan.

Chihiro Hirotsu
Faculty of Science & Technology,
Meisei University, 2-1-1 Hodokubo,
Hino-City, Tokyo 191-8506, Japan.
E-mail: hirotsu@ge.meisei-u.ac.jp

Gudmund Iversen
Department of Mathematics and Statistics,
Swarthmore College, 500 College,
Av. Swarthmore, PA 19081, USA.
E-mail: iversen@swarthmore.edu

Flavia Jolliffe
Computing and Mathematical Sciences,
University of Greenwich, Maritime
Greenwich Campus, Old Royal Naval
College, Greenwich, London, SE10 9LS U.K
E-mail: F.R.Jolliffe@gre.ac.uk

Tae Rim Lee,
Korea National Open University,
Depatrment of Applied Statistics,
169 Dongsung dong, Jongro Ku,
Seoul 110-791 Korea.
E-mail: trlee@av9500.knou.ac.kr

Kumiko Maruyama
Seigakuin University, Faculty of Humanities,
Department of Human Welfare,
1-1,Tozaki Ageo-shi, Saitama-Ken,
362-8585 Japan.
E-mail: k_maruyama@seigakuin-univ.ac.jp

Tadashi Matsunawa
The Institute of Statistical Mathematics,
4-6-7, Minami-Azabu, Minato-ku, Tokyo
106-8569, Japan.
E-mail: matunawa@ism.ac.jp

Kameo Matusita
5-28-7, Kamiuma, Setagaya-ku, Tokyo,
154-0011, Japan.

Alan McLean
Monash University, Department of
Econometrics and Business Statistics,
900 Dandenong Road, Caulfield East,
Victoria 3150, Australia.
E-mail: alan.mclean@buseco.monash.edu.au

Hiroshi Midzuno
1720-3, Higashinaganuma, Inagi,
Tokyo, 206-0802, Japan.

Yuki Miura
Faculty of Economics, Surugadai University,
2-43-12, Kichijoji Higashi-cho,
Musashino-shi, Tokyo 180-0002, Japan.
E-mail: miura@parkcity.ne.jp

Andile Mji
University of Transkei,
Research Resource Centre, Private Bag X1,
Unitra Umtata 5117 South Africa.
E-mail:mji@getafix.utr.ac.za

S. P. Mukherjee
University of Calcutta, Department of Statistics,
35 Ballugunge Circular Road,
Calcutta 700019, India.
E-mail: csri@satyam.net.in

Masakatsu Murakami
The Institute of Statistical Mathematics,
4-6-7 Minami Azabu,Minato-Ku,
Tokyo 106-8569, Japan.
E-mail: murakami@ism.ac.jp

Kenzi Nakamura
The Institute of Statistical Mathematics,
Tokyo, Japan.
E-mail:Kenzi@dreamtheater.zzn.com

David Ospina
Universidad Nacional de Colombia,
Departamento de Estadística,
Apartado Postal 75230,
Santa Fe de Bogotá, Colombia.
E-mail: dospina@matematicas..unal.edu.co

Maria-Gabriella Ottaviani
Universita' di Roma "La Sapienza", Dip. di
Statistica, Probabilita' e Statistiche Applicate,
P.le A. Moro, 5 - 00185 Roma.
E-mail: Mariagabriella.Ottaviani@uniroma1.it

Brian Phillips
Swinburne University of Technology,
School of Mathematical Sciences,
P.O. Box 218, Hawthorn 3122, Australia.
E-mail: bphillips@groupwise.swin.edu.au

Naoko Sakurai
Waseda University
Institute of Asia-Pacific Studies,
1-21-1, Nishiwaseda, Shinjuku-ku,
Tokyo, 169-0051, Japan.
E-mail: sakurai@wiaps.waseda.ac. jp

David J. Saville
Biometric Unit Agresearch, Canterbury
Agriculture & Science Centre, P.O. Box 60,
Lincoln, Canterbury, New Zealand.
E-mail: savilleD@agresearch.cri.nz

Gilberte Schuyten
University of Gent,
Department of Data Analysis,
Henri Dunantlaan 1, 9000 Gent, Belgium.
E-mail: gilberte.schuyten@rug.ac.be

Ben-Chang Shia
Department of Statistics, Fu Jen University,
510 Chung Chen Rd. Hsin Chuang, 24205,
Taipei Hsien, Taiwan, China.
E-mail: stat1001@mails.fju.edu.tw

Toshiro Shimada
Professor Emeritus Meiji University,
8-9-3-314, Kokuryo-machi Chofu-shi,
Tokyo, 182-0022, Japan.
E-mail: simfuji0@mtj.biglobe.ne.jp

Ryoichi Shimizu
The Institute of Statistical Mathematics,
4-6-7, Minami-Azabu, Minato-ku,
Tokyo 106-8569, Japan.
E-mail: shimizu@ism.ac.jp

Dalene Stangl
Duke University,
Institute of Statistics and Decision Sciences,
P.O. Box 90251, 223C Old Chemistry,
Durham, NC, 27708-0251, USA.
E-mail: dalene@stat.duke.edu

Elisabeth Svensson
Chalmers and Göteborg University,
Biostatistics Branch, Department of Statistics,
Örebro University
SE-701 82 Orebro, Sweden.
E-mail: eliss@math.chalmers.se

Ann-Lee Wang
University of Malaya,
Institute of Mathematical Sciences,
50603 Kuala Lumpur Malaysia.
E-mail: j2wang@ns1.umcsd.um.edu.my

Jinfang Wang
The Institute of Statistical Mathematics,
4-6-7, Minami-Azabu, Minato-ku,
Tokyo 106-8569, Japan.
E-mail: wang@ism.ac.jp

Yuan Wei
Renmin University of China,
175 Haidian Road, Beijing, 100872, China.
E-mail:wyuan@mail.ruc.edu.cn

Ryozo Yoshino
The Institute of Statistical Mathematics,
4-6-7, Minami-Azabu, Minato-ku, Tokyo,
106-8569, Japan.
E-mail:yoshino@ism.ac.jp

Yuejun Zheng
The Institute of Statistical Mathematics,
4-6-7, Minami-Azabu, Minato-ku, Tokyo,
106-8569, Japan.
E-mail: zheng@ism.ac.jp

PART 2

TRAINING RESEARCHERS IN PARTICULAR STATISTICAL TOPICS

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 23-35.

ELISABETH SVENSSON

IMPORTANT CONSIDERATIONS FOR OPTIMAL COMMUNICATION BETWEEN STATISTICIANS AND MEDICAL RESEARCHERS IN CONSULTING, TEACHING AND COLLABORATIVE RESEARCH, WITH A FOCUS ON THE ANALYSIS OF ORDERED CATEGORICAL DATA

This paper focuses on problems encountered in the teaching of statistics to applied researchers working particularly with rating scales and questionnaires. Examples of teaching strategies that are designed to remove misconceptions and the misuse of statistics will be presented. Such strategies should increase the level of understanding about the relationship between study design, measurement processes and the choice of statistical methods of analysis. A survey among applied researchers showed that tradition, the need to compare the results with other studies and a lack of knowledge of novel statistical methods were the major factors determining the choice of methods for evaluation of questionnaires. Besides pedagogic skills, professional competence and an open-minded inter-disciplinary understanding were the most important qualifications for optimal inter-professional communication.

1. BACKGROUND

Various reviews of medical journals have highlighted the poor quality of methodology and statistics in medical research (Altman, 1991; 1994; Hand, 1994; Coste, Fermanian, & Venot, 1995; Feinstein, 1997). The increasing use of complex methods, such as survival analysis and multiple regression analysis, and the use of questionnaires and rating scales, also creates problems. Therefore, there is a clear need for statisticians to be involved in applied research at an early stage (Altman, 1998; Nelder, 1999).

However, the accessibility of statistical computer programs may provide an excuse for not consulting a statistician, to the subsequent detriment of the scientific quality of the research. A reliance on statistical software without enough statistical knowledge could result in incorrect statistical treatment of data (Shimada, 2001; Jolliffe, 2001).

Recently, Hand (1996) drew attention to the fact that little consideration is given to the relationship between measurement theory and assessment, although this is fundamental to the choice of statistical approach to the data. The published comments to his paper, given by several statisticians, illustrate the various opinions that are prevalent concerning the importance of theories of measurements. Clearly, the impact of the measurement process on the correct choice of statistical analysis must be considered for each study (McPherson, 1989; Altman, 1991; Svensson, 1998a).

Questionnaires and rating scales are commonly used to measure qualitative variables, such as feelings, attitudes, preferences and health-related variables. The response values from rating scales indicate only an ordered structure and not a numerical value in a mathematical sense (Stevens, 1946; Merbitz, Morris, & Grip, 1989;

Agresti, 1990; Hand, 1996). The rank-invariant properties of data from rating scales mean that the statistical methods used for their analysis differ completely from the traditional statistical methods for quantitative variables (Svensson, 1993). The rank-invariant properties of ordered categorical data are well known, even if there is still controversy about the measurement properties of data from rating scales and a misuse of statistical methods and misinterpretation of results from qualitative measurements. As categorical responses are often transformed into numerical scores, there is also a temptation to treat such quantified data as numbers with the same arithmetic properties as quantitative data (Feinstein, Josephy, & Wells, 1986; Agresti, 1990; Altman, 1991; Coste, Fermanian, & Venot, 1995).

2. INTRODUCING NOVEL STATISTICAL METHODS FOR ORDINAL DATA

With the popularity of questionnaires, there is an increased demand for statistical methods for dependent ordinal data. My research concerns development of statistical methods that take account of the rank-invariant properties of ordinal data. A family of methods for a comprehensive evaluation of reliability and also of change in ordinal assessments has been proposed (Svensson, 1993; 1997; 1998a; 1998b). The basis of this novel statistical approach is the bivariate ranking procedure that makes it possible to measure the systematic component of change in paired assessments separately from the individual variations.

The demand for rank-invariant statistical methods for dependent ordinal data among applied scientists has led to the early introduction of these methods in courses and collaborative research projects, and also to reformulation of the mathematical description of the measures (Sonn & Svensson, 1997; Gosman-Hedström & Svensson, in press; Claesson & Svensson, 2001). Therefore much experience concerning the consequences of introduction of novel statistical methods on applied research is gained.

3. FOCUS OF THE PAPER

This paper will focus on the teaching and learning processes associated with the statistical treatment of data from questionnaires and rating scales. The link between the teaching and learning processes is the inter-professional communication based on a mutual understanding of the problems from both the applied science and statistical perspective (McPherson, 1989; Altman, 1991; Greenfield, 1993).

My own experience of the importance of creating good inter-disciplinary communication and of the similarities and differences in the teaching processes between education, consultation and research collaboration will be reported.

Furthermore, the factors identified by applied researchers in their choice of methods for statistical analysis of data from rating scales will be reported from a questionnaire, which attempted to define the researcher´s attitudes towards rank-invariant statistical methods and the reasons behind the choice of appropriate or inappropriate methods of analysis. The results will form the basis of recommendations for approaches to achieve optimal inter-disciplinary communication in teaching, consultation and collaborative research with regard to statistical methods for ordered categorical data.

4. TEACHING STATISTICS TO RESEARCHERS

Fundamental to the teaching of practical statistics is the mutual recognition of the complexity of the applied research problem in relation to the statistical possibilities and restrictions. Therefore, the main criterion for a successful learning process is to create optimal communication between the statistician and the applied researcher. The researcher should gain scientific and statistical knowledge and confidence in order to be able to choose appropriate statistical methods for the research project. Therefore, it is important to find a common language and to make the statistical theories and approaches understandable and relevant to the researcher's own field of interest.

This is the basic approach to the teaching model used by the author for research courses in practical statistics for applied scientists in Sweden. The teaching model is interactive and focuses on statistical strategy rather than on statistical technique (Svensson, 1998c). The measurement process, including the operationalisation process of the variables, and the identification of the measurement properties of the data, are important issues. The participants are encouraged to apply appropriate methodological and statistical theories to their own research problems and to discuss the research process during the course. A model for teaching the measurement process was presented at the ICOTS 5 meeting in 1998 (Svensson, 1998d).

Another important issue to take into account in the teaching process is the potential conflict between members within a research group when introducing new statistical approaches. In order to avoid communication problems and scientific conflicts, courses in scientific methodology and applied biostatistics have been given to research groups that have included all the researchers, their supervisor and others sharing the same research problem. The experiences gained from such courses have also been presented at the ICOTS 5 meeting in 1998 (Svensson 1998c).

The consultation procedure provides an ideal learning situation, as the statistician and the researchers can concentrate their discussions on a specific applied problem. The main teaching approach in research courses on applied statistics is to create a climate of mutual understanding, which is very similar to the consultation situation.

In a course, there is often a broad range of fields of interests represented. By means of interactive learning, the researchers must apply the methodological and statistical theories to their own research problems. In the discussions with other participants, similarities and differences in statistical solutions, in the measurement processes and comparisons between approaches for qualitative and quantitative data will increase their understanding of problems and also shed light on the need for different statistical methods for different types of data. This means that it is sometimes advantageous to discuss statistical approaches in a course, as all the researchers contribute to the understanding of the statistical solutions to the complex problems encountered in reality.

5. THE SURVEY

Between 1994 and 1999, courses in scientific methodology and practical statistics for applied researchers in medical and health sciences were offered to doctoral students and others involved in research projects at Göteborg University and the Sahlgrenska University Hospital, Sweden. As mentioned above, some of the courses were aimed at clinical research groups including the supervisors and post-doctorate scientists. In 1999,

a questionnaire was mailed to all 108 individuals, with known addresses, who had participated in the courses, in seminars or consultations with the author. This means that the participants of this survey have a good basic knowledge of statistical methods, and they were all aware of the link between the measurement properties of data and the choice of appropriate statistical methods of analysis.

6. RESULTS

Responses to the questionnaire were obtained from 73 individuals (68% of the questionnaires sent), who were involved in ongoing research projects as researchers or supervisors. Sixty had participated in a research course, and the others had participated in shorter courses, seminars and/or statistical collaboration with the author. The largest professional group were physicians (n=25), but nurses, occupational therapists, physiotherapists, social workers and laboratory technicians were also represented.

As part of the survey concerned various aspects of being dependent on a supervisor, three subgroups were identified. There were 42 doctoral students, who had supervisors (“doctoral student“). Twenty individuals were involved in research projects as researchers, assistant researchers or as student supervisors to master’s degree at university (“others“). Finally, eleven post-doctorate researchers were included, of whom six were also supervisors (“post-doc“).

6.1. ROLE OF THE STATISTICIAN IN APPLIED RESEARCH

Table 1 shows which parts of the research process commonly involved a statistician at the research department according to the 54 individuals, who responded to the question. One main reason for the 19 non-responders was that they did not know the common routines at the research department. According to nine of the 54 responders (17%) a statistician was never involved in the projects at the research department. In general, according to 30 (56%) responders, a statistician will be involved after all data have been collected, and this was often the first reason for involving a statistician.

Three of the supervisors were aware of that contacting a statistician first when all data were collected was too late. One of them proposed a biostatistical centre at the hospital with access to free statistical advice before starting a study.

Table 1. Frequency (and Percent) of Responses about which Research Process Stage Involved a Statistician

Research process stage (n=54 responses)	Stage involving a statistician	First stage of statistical contact
Planning	18 (33)	18 (33)
Design of materials, sample size	16 (30)	5 (9)
Design of methods	12 (22)	2
After collecting data	30 (56)	18 (33)
Interpretation of calculations by computers	10 (18)	1 (2)
When writing the report/article	14 (26)	1 (2)
After the referee´s review	7 (13)	
None	9 (17)	9 (17)

A similar question concerning the involvement of a statistician in ongoing research among the 42 doctoral students showed that 29 (69%) had involved a statistician. The

first reason for contacting a statistician was design issues (41%) or statistical treatment of data (38%). Some of the doctoral students pointed out that since they attended the statistical research course at the beginning of their research project they had to apply the measurement process and other design issues to their own research. Three of the non-responders were not aware of the importance of a statistical contact, and the supervisors of four doctoral students judged that there was no need to involve a statistician.

Four of the post-doctorate researchers stated that supervisors are supposed to have sufficient knowledge of statistics. Therefore, contacting a statistician was a low priority. However, some of the doctoral students were recommended to contact a statistician even when the supervisor had had a negative experience of consulting a statistician because they failed to focus on the research problem. Another comment was that there was a lack of statisticians who were well acquainted with nursing research. Therefore, well-known statistical methods, even though inappropriate for the research problem, were chosen, despite the fact that this would lead to unreliable conclusions. One supervisor commented that it was easier to keep to tradition than to hear from the statistician that the approach used was inappropriate.

6.2. THE MEASUREMENT PROCESS AND THE CHOICE OF STATISTICAL METHODS

According to 20 of the 42 doctoral students, there was generally no discussion between the supervisor and the doctoral student concerning the link between the properties of data, the design and the choice of statistical methods. The main reason mentioned (n=14) was the lack of knowledge among the supervisors, and that this kind of question had a low priority (n=7) in discussions concerning the research project. Four doctoral students mentioned that there was a statistician involved in the project, but a statistician was generally not involved in the discussions. According to seven post-doctorate researchers, there was an intention to discuss design problems with a statistician, but this had a low priority and was not normally included in the research process.

Table 2 shows that tradition and the statistician´s advice were the two most common external reasons for the choice of statistical methods. The purpose of the study and the properties of data also determined the choice of statistical methods according to five doctoral students and three post-doctorate researchers.

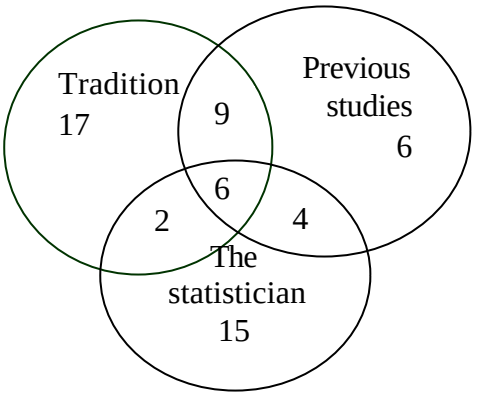
Table 2. Frequency (and Percent) of Usual Considerations Behind the Choice of Statistical Methods in the Different Groups of Individuals Involved in Scientific Projects

Considerations for the choice of statistical method	Doctoral students (n = 42)	Post-doc (n =11)	Others (n =20)	Total (n=73)
Tradition	22	5	9	36 (49)
Statistician´s advice	20	4	4	28(38)
Previous studies	13	3	8	24(33)
The journal	6	0	3	9(12)
The statistical software	2	1	4	7(10)
No response/ do not know	2	1	4	7 (10)

Figure1 shows the relationship between the three main considerations behind the choice of statistical methods according to 59 individuals. In seven additional cases the

purpose of the study or the properties of data were mentioned as single reasons for the choice of statistical methods.

Figure1. Relationship Between the Three Main Reasons Behind the Choice of Statistical Methods in 59 Responses



Sixty-two responders rated the level of importance of keeping to the tradition in the choice of statistical methods. One third stated that it was “very important“(8) or “important“ (12), while the traditional choice of statistical methods was “not so important“ (23) or “unimportant“ (19) according to 68% of the responders. The most frequent reasons for opinion among the doctoral students (n=41, one non-responder) are listed in Table 3. The numbers of similar responses are given in parentheses.

Table 3: Relative Importance among 41 Doctoral Students to Keep to the Traditional Choice of Statistical Methods and (Frequency) of their Reasons

Reasons for considering Very important (n=4) or Important (n=8) to keep to tradition	Reasons for considering Not so important (n=16) or Unimportant (n=13) to keep to tradition
Communication with the supervisor and the research group (3)	It is more important to choose appropriate methods (7)
Comparability with previous studies (3)	The research field is new with no traditions (4)
The traditions maintain the quality of research (2)	Traditional methods are not appropriately updated (3)
The supervisor is stuck to the tradition and scientific journals (1)	Confidence gained in statistical advice (1)
Acceptance (1)	A possibility to influence the tradition (1)

6.3. THE CHOICE OF NOVEL STATISTICAL METHODS

The statistical treatment of data from rating scales and questionnaires should take into account the non-metric properties of ordinal data. The research groups of three doctoral students used statistical methods appropriate for ordinal data. Sixteen of the 30 doctoral students, who dealt with rating scales, and eleven of the others involved in projects knew that some of the research group members used statistical methods that assume quantitative data, when qualitative data were analysed. The main reasons mentioned for this were that it allowed comparison with other studies and a lack of knowledge among supervisors and other applied scientists (Table 4).

Table 4. Frequency of Researchers' Main Reasons to Use Statistical Methods Suitable for Quantitative Data in the Evaluation of Qualitative Sata

Main reasons	Doctoral students (n=20)	Others (n=11)
Comparability with other studies	12	6
Lack of knowledge by the supervisor	10	8
Better acceptance with traditional methods	8	3
Treatment according to the manual of the instrument	5	3
Lack of knowledge among researchers	4	5
The fear of going against the current	4	3
Disagreement between statisticians	4	4
The same results with different methods	5	1
The journal wants well-known statistical methods	4	1

6.4. STATISTICAL ANALYSIS OF DATA FROM RATING SCALES

The majority, 50 out of 73, of the individuals in the present study used rating scales in their research. For 19 of them the rating scales were the main instruments in the research.

Among eight of the doctoral students and five of the other researchers, there had been conflicts with their supervisors concerning the choice of statistical methods for analysis of data from rating scales and questionnaires. The reasons were the choice of the novel rank-invariant statistical method instead of the traditional treatment (n=8), the choice of scaling (n=3) and the statistical description of ordinal data. Five of the doctoral students had to use parametric methods in their latest research in order to be able to compare results with previous studies and also because this was demanded by the supervisor.

Table 5. Reasons Given by Doctoral Students for not Using Novel Statistical Methods for Ordinal Data (and their Frequency)

The lack of knowledge concerning new statistical methods (n=13)
Among supervisors
Among research group members
Among statisticians
Alternative statistical methods are unknown
Publication delay when using novel methods
Lack of communication between statisticians and applied scientists
Lack of biostatisticians
The tradition in the research group (n=9)
Disagreement amongst statisticians
The presence of various approaches to analysing data from rating scales
The need to compare with other/previous studies (n=8)
A matter of acceptance
The refereeing system
Difficulty of changing established behaviour
Difficulties in learning new techniques
Resistance in research group
Lack of confidence
New statistical methods are not accepted

The major reasons why novel statistical methods for ordinal data were not used, according to the doctoral students, were lack of knowledge, the tradition within research groups and that traditional methods were demanded in order to compare results with those of previous studies (Table 5). Seven doctoral students have applied the rank-invariant approach in their research. One scientific journal required the traditional parametric approach, before it would accept the paper.

Table 6 presents the pros and cons of choosing a novel statistical method of analysis according to the experiences of these doctoral students. The numbers of similar responses in parentheses.

Table 6. Advantages and Disadvantages of Choosing a Novel Statistical Method Experienced by Seven Doctoral Students (and their Frequencies)

Advantages	Disadvantages
It was the appropriate method for the research problem (n=6)	Conflict with the other researchers (n=2)
The results are reliable and interpretable (n=5)	Lack of confidence in novel methods among reviewers (n=2)
The possibility of performing a comprehensive evaluation of data (n=4)	Difficult to compare results from other studies (n=1)
Ethical reasons (n=1)	Time to learn new methods (n=1)
Scientific challenge (n=1)	
Scientific relevance, quality (n=1)	
Honest reporting of the results from subjective assessments on rating scales (n=1)	

6.5. FACTORS OF IMPORTANCE FOR THE OPTIMAL COMMUNICATION BETWEEN STATISTICIANS AND APPLIED RESEARCHERS

The main open question of this study concerned important qualifications among statisticians and the applied scientists, and other considerations, for obtaining meaningful communication. Table 7 shows the suggestions from 65 of the 73 individuals. The numbers of similar responses in parentheses

Additional considerations of inter-disciplinary importance, mentioned by 34 individuals, when the research involves rating scales and questionnaires, were:

- Collaborative research projects with a competent biostatistician;
- Open-minded discussions;
- To offer seminars, workshops and courses for applied scientists including supervisors;
- To offer seminars, workshops and courses for biostatistician;
- To have an open dialogue with the supervisor;
- Ability of breaking the resistance against novel statistical methods in research groups;
- To show inter-disciplinary respect, humility, understanding;
- Inter-disciplinary communication in all research projects;
- To make statistical methods understandable for applied scientists.

Table 7. The Most Important Qualifications and Considerations for Effective Inter-Disciplinary Communication Suggested by 65 Responders

Qualifications of the statistician	Qualifications of the applied researchers
Competence (n=24):	Statistical knowledge/understanding (n=25)
Broad knowledge of applied statistics	Knowledge in basic practical statistics (n=25)
Experienced in applied statistical research (n=4)	Interest in learning (n=12)
	Open-minded to statisticians and statistical advice (n=19)
Interdisciplinary knowledge/interest (n=19)	
Knowledge of the applied research field (n=19)	Scientific competence (n=32)
Understanding of the applied research problem/context (n=17)	Ability to present the research problem and clearly defined questions (n=32)
Awareness of clinical/practical difficulties in applied research (n=8)	To have a scientific approach to the research project (n=8)
Interest in applied research problems (n=15)	A true interest in the applied research problem (n=5)
Ability to enter into the applied research problem (n=7)	Ability to define the measurement process (n=3)
Open-mindedness, flexibility (n=10)	To be honest (n=2)
Pedagogic ability to: (n=21)	
Explain understandably (n=12)	Inter-disciplinary communication (n=20)
Listen (n=7)	To make early contact (n=13)
Motivate into accepting the relevance of suggested methods(n=5)	To be well-prepared (n=9)
Collaborate (n=4)	Open minded, flexible, curious, unafraid of questioning (n=18)
Be distinct, honest, confident (n=10)	Ability to explain the use of traditional methods (n=1)
Accessibility, continuity (n=14)	Ability to break the use of traditional methods (n=3)
Other(n=6)	Ability to collaborate (n=3)
Persuasive powers, a sense of humour, Enthusiasm, ability to enjoy his/her work, patience	Accessibility

7. DISCUSSION

This study among researchers with good basic knowledge of statistical methods showed that about half the doctoral students discussed the relationship between the properties of data and the choice of statistical methods with their supervisors, but commonly without a statistician present. The main reasons for involving a statistician in applied research were, according to this study, after collecting the data and when planning the study, but, in general, statistical contacts had a low priority, especially among supervisors. The lack of experienced statisticians and lack of a common language were reasons for applied scientists to keep to well-known statistical approaches, disregarding the appropriateness, without involving statisticians.

Comparability with other studies, communication with other researchers and acceptance were important factors behind the preference for well-known statistical methods. On the other hand, statistical traditions were not so important for 29 of the 41 doctoral students, who preferred the choice of appropriate methods. However, lack of knowledge concerning new statistical methods among supervisors, researchers and statisticians is still a hindrance to the choice of novel statistical methods. The lack of

understanding of the relationship between the measurement properties of data and the choice of statistical methods of analysis among statisticians and applied researchers is a global problem (Nelder, 1986; McPherson, 1989; Greenfield, 1993; Hand, 1996; Bishop, 2000; Jolliffe, 2001). There is therefore a need to train both statisticians and applied researchers in order to produce good-quality research.

This study also confirmed that there is a potential conflict between the use of non-standard statistical methods in applied research and the well-known traditional methods, in terms of acceptance by referees and journals. According to Lesser and Parker (1995) the best intentions of biostatisticians to provide a thorough statistical analysis could be counter-productive and result in unfavourable reviews by journals. It is common that editors of journals in applied research fields suggest that description of statistical methods used in a study, even when the methods are uncommon, should be minimised or replaced by a reference to a statistical paper (Jolliffe, 2001). I have as yet experienced very few exceptions. For pedagogical reasons editors have accepted a comprehensive demonstration of new statistical methods for analysis of data from rating scales that were reformulated, demonstrated and explained in an understandable way for applied scientists (Sonn & Svensson 1997; Gosman-Hedström & Svensson, in press; Claesson & Svensson, 2001).

The lack of knowledge concerning the research process and statistical methods for the various applied problems is also common among editors and referees. Altman (1998) recommends that biostatisticians should review papers in medical journals in order to increase the quality of medical research.

The need for knowledge in basic statistics and scientific competence among the applied scientists was one of the most important factors for communication with statisticians. Important factors for the optimal interdisciplinary communication suggested in this study were seminars, workshops and courses for research groups and statisticians and inter-disciplinary communication in all research projects. My experience of the research courses in practical statistics for study groups, including the supervisors, is that the group members gained confidence in statistics and developed a higher level of awareness concerning the choice of statistical methods appropriate to the measurement level of the outcome measurements (Svensson, 1998c). The teaching model of the International Clinical Epidemiology Network (INCLEN), as experienced by Bangdiwala (2001), provides training of applied medical researchers in statistics and statisticians in clinical epidemiology methods, but could be applied to health related research problems as well. Bishop and Talbot (2001) propose an approach to training applied researchers in statistical thinking with attention to the entire research process.

Another way of eliminating a potential conflict between members in a research group is to give seminars concerning the relationship between the applied research problem, the measurement process, other design issues and the appropriate statistical methods. This approach would also improve the communication skills among statisticians and applied researchers and stimulate the interdisciplinary knowledge. Targeted seminars reflect the teaching-learning process of consulting, when the members of the research group and the supervisor are present. According to my experience, supervisors appreciated this possibility of up-dating their statistical knowledge with focus on the relationship between their complex research problems and the statistical possibilities and restrictions of application. Seminars and workshops focusing on specific applied research problems would meet the need for continuous training in statistics. New statistical methods and software might have an influence on applied research problems and vice versa, as practical problems stimulate statistical methodological research as well (Greenfield, 1993; Svensson, 1993; McPherson, 1989;

Jolliffe, 2001).

As a result of this study, students in statistics, statisticians and applied scientists are invited to a new series of symposia concerning statistical problem solving and interactive communication in medical and health sciences arranged by me at the Örebro University in Sweden. Each symposium will have a theme, such as "statistical aspects on medical diagnostic tests", "repeated measurements", "the measurement process and statistics", "statistical aspects on rating scales and questionnaires", "on significances", and will normally contain two half-day lectures held by invited statisticians and applied scientists.

The lack of biostatisticians, especially with experience in statistical evaluation of data from rating scales and questionnaires, is a major hindrance in inter-professional communication concerning qualitative assessments. From my experience, most teaching of theoretical and practical statistics are focused on methods and models appropriate for quantitative data. Therefore, well-educated statisticians may be virtually unaware of the fact that there are statistical methods that take into account the rank-invariant properties of data from rating scales. The controversy concerning the choice of statistical approach for data from questionnaires and rating scales is not only a sign of the lack of knowledge of appropriate methods for ordered categorical data, but also reflects the statistical and methodological complexity of subjective assessments. A review of the scientific literature in which statistical methods have been applied to rating scales and data would certainly reveal a high level of ignorance of the non-metric measurement properties of ordered categorical data.

This study showed that there is a need for more biostatisticians with an interest in collaborative research, not only for the improvement of the applied research but also for the development of the bio-statistical science. The statistics departments should therefore inform statisticians about the applied research fields that provide both statistical and educational challenges. All biostatisticians should have experience of collaborating with research groups, and should be familiar with the importance of the measurement process for the choice of statistical approach. There is not only a need for statistical knowledge, but also a need to be able to listen, to show interest in applied problem solving and to be able to transform abstract statistical descriptions into an understandable applied context. Statistical consultancy offers a good practice in the communication skills and so does the participation in workshops and applied research (Greenfield, 1993; Nelder, 1986; Preece, 1987; Belli 2001, Jolliffe 2001, Godino, Batanero, & Gutiérrez- Jáimez 2001).

REFERENCES

Agresti, A. (1990). *Categorical data analysis.* New York: Wiley.

Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman.

Altman, D. G. (1994). The scandal of poor medical research. *British Medical Journal,* 308, 283-284.

Altman, D. G. (1998). Statistical reviewing for medical journals. *Statistics in Medicine,* 17, 2661-2674.

Bangdiwala S. I. (2001). Training of statisticians worldwide to collaborate as co-investigators within country clinical epidemiology units: The experience of the International Clinical Epidemiology Network (INCLEN). In Batanero, C. (Ed.), *Training researchers in the use of statistics* (pp. 265-275). Granada: International Association for Statistical Education and International Statistical Institute.

Belli, G. (2001).The teaching/ learning process in University statistical consulting labs in the United States. In Batanero, C. (Ed.), *Training researchers in the use of statistics* (pp. 325-338). Granada: International Association for Statistical Education and International Statistical Institute.

Bishop, G., & Talbot, M. (2001). Statistical thinking for novice researchers in biological sciences. In Batanero, C. (Ed.), *Training researchers in the use of statistics* (pp. 215-226). Granada: International Association for Statistical Education and International Statistical Institute.

Coste, J, Fermanian, J, & Venot, A.(1995). Methodological and statistical problems in the construction of composite measurement scales: A survey of six medical and epidemiological journals. *Statistics in Medicine,* 14, 331-345.

Feinstein, A. R. (1997). Beyond statistics: What is really important in medicine? *Cleveland Clinical Journal of Medicine,* 64,127-128.

Feinstein, A. R., Josephy B. R., & Wells, C. K. (1986). Scientific and clinical problems in indexes of functional disability. *Annals of Internal Medicine,* 105, 413-420.

Godino, J. D., Batanero, C., & Gutiérrez-Jáimez, R. (2001) The statistical consultancy workshop as a pedagogical tool. In Batanero, C. (Ed.), *Training researchers in the use of statistics* (pp. 339-353). Granada: International Association for Statistical Education and International Statistical Institute.

Gosman-Hedström, G., & Svensson, E. (In press). Parallel reliability of the functional independence measure and the Barthel ADL index. *Disability and Rehabilitation*.

Greenfield, T. (1993). Communicating statistics. *Journal of the Royal Statistical Society, Series A,* 156, 287-297.

Hand, D. J (1994). Deconstructing statistical questions. *Journal of the Royal Statistical Society, Series A* 157, 317-356

Hand, D. J. (1996). Statistics and the theory of measurement. (incl. discussion). *Journal of the Royal Statistical Society, Series* A, 159, 445-492.

Jolliffe F. (2001) Learning from experience. In Batanero, C. (Ed.), *Training researchers in the use of statistics* (pp. 355-370). Granada: International Association for Statistical Education and International Statistical Institute.

Lesser, M. L., & Parker, R. (1995). The biostatistician in medical research: Allocating time and effort. *Statistics in Medicine,* 14, 1683-1692.

McPherson, G. (1989). The scientists´view of statistics. A neglected area. *Journal of the Royal Statistical Society, Series A,* 152, 221-240.

Merbitz, C., Morris J., & Grip J. C. (1989). Ordinal scales and foundations of misinference. *Archives of Physical Medical Rehabilitation*, 70, 308-312.

Nelder, J. A. (1986). Statistics, science and technology. *Journal of the Royal Statistical Society, Series A,* 149, 109-121

Nelder J. A. (1999). Statistics from the millenium. From statistics to statistical science. *The Statistician,* 48, 257-269.

Preece, D. A. (1987). Good statistical practice. *The Statistician, 36, 397-408*.

Shimada T. (2001) Precaution against errors in using stochastic software. In Batanero, C. (Ed.), *Training researchers in the use of statistics* (pp. 129-137). Granada: International Association for Statistical Education and International Statistical Institute.

Sonn U., & Svensson, E. (1997). Measures of individual and group changes in ordered categorical data: Application to the ADL Staircase. *Scandinavian Journal of Rehabilitation Medicine,* 29, 233-242

Stevens, S. S. (1946). On the theory of scales of measurement. *Science,* 103, 677-680.

Svensson E. (1993). *Analysis of systematic and random differences between paired ordinal categorical data* Ph.D. Thesis. Stockholm: Almqvist & Wiksell International.

Svensson E. (1997). A coefficient of agreement adjusted for bias in paired ordered categorical data. *Biometrical Journal,* 39, 643-657.

Svensson E. (1998a). Application of a rank-invariant method to evaluate reliability of ordered categorical assessments. *Journal of Epidemiology and Biostatistics*, 3 (4), 403-409.

Svensson E. (1998b). Ordinal invariant measures for individual and group changes in ordered

categorical data. *Statistics in Medicine*, 17(24), 2923-2936.

Svensson E. (1998c). Teaching biostatistics to clinical research groups. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statist*ics (pp. 289-294). Singapore: International Association for Statistical Education and International Statistical Institute.

Svensson E. (1998d). Teaching the measurement process in biostatistics. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 1257- 1262). Singapore: International Association for Statistical Education and International Statistical Institute.

Elisabeth Svensson
Mathematical Statistics,
Chalmers and Göteborg University, Göteborg,
and Department of Statistics,
Örebro University, SE-701 82 Örebro, Sweden
E-mail: elisabeth.svensson@esa.oru.se

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 37-51.

Printed in Granada, Spain.

ANTONIO ESTEPA & FRANCISCO T. SÁNCHEZ COBO

EMPIRICAL RESEARCH ON THE UNDERSTANDING OF ASSOCIATION AND IMPLICATIONS FOR THE TRAINING OF RESEARCHERS

In this paper we summarise the main research findings on the understanding of association carried out in psychology and mathematics education and we present results from an assessment study on the understanding of correlation and regression by university students. We finally discuss the implications of these results for designing courses directed to train researchers in the use of statistics

1. INTRODUCTION

Association has great relevance for the training of researchers since it is essential for many statistical methods and techniques frequently used by researchers, such as simple and multiple regression, variance and covariance analysis, log-linear models and LISREL, in addition to the majority of multivariate methods. Association plays a main role in educational research (Blumberg, 2001), and its understanding is also needed to read research literature (Bangdiwala, 2001). On the other hand, there is a close connection between the concept of association and the idea of cause, on which scientific knowledge and decision making are based, since causal explanations allow us:

> *to explain the past, control the present and predict the future* (Crocker, 1981, p. 272).

This notion has been debated by mathematicians, statisticians and philosophers of science, and, even when research methodology is based to a great extent on studying the association between variables, causality and association do not always coincide.

Besides this epistemological difficulty, psychological and mathematical education research has shown that the ability to judge association well is not developed intuitively. Numerous difficulties and obstacles related to association might imply serious misinterpretations and misuses of statistics methods in research. Below, we summarise the main research in this area, and present the results of an assessment study on the meaning correlation and regression by undergraduates. We finally reflect on the implications of these studies for the training of researchers.

2. PREVIOUS RESEARCH ON THE DEVELOPMENT AND USE OF ASSOCIATION

2.1. RESEARCH FROM PSYCHOLOGY

The perception of covariation between stimuli, behaviour and outcomes is a main component of human adaptable behaviour, and this explains the interest towards association from clinical, social and development psychologists. Decision making under

uncertainty has been studied in these fields, where heuristics have been described (Kahneman, Slovic, & Tversky, 1982), using the metaphor of the intuitive statistician (Peterson & Beach, 1967).

Although the pioneering work by Inhelder and Piaget (1955) on the development of the idea of correlation was carried out with adolescents (12-15 year olds), most research in psychology has focussed on adults, such as undergraduates or professionals, and therefore its findings can be valid for the training of researchers. This research shows that adult's reasoning on association is, as a rule, very poor (Nisbett & Ross, 1980).

Strategies in the interpretation of association in contingency tables

Most psychological research on association studies adult's interpretation of association from 2x2 contingency tables, like table 1, by analysing the strategies employed, performance of association judgements and the factors that intervene in the same.

Table 1. Data in a 2x2 Contingency Table

	B	Not B	Total
A	a	b	a+b
Not A	c	d	c+d
Total	a+c	b+d	a+b+c+d

A first set of papers study the strategies employed in solving association problems. Pérez Echeverría (1990) distinguishes seven types of strategies found in these papers, according to the cells employed in Table 1 and the form in which they are combined:

- Using only cell [a] (present - present) where the two variables simultaneously occur, estimating a positive, negative or null association if this value is greater than, smaller than or equal to the other three cells.
- Finding the difference [a-b] between cell a (present - present) and cell b (present - absent). This was the most widely used strategy in one experiment by Arkes and Harkness (1983).
- Using the difference between the absolute frequencies of cells a and c (absent - present).
- Comparing the differences found in cells [a, b, c], which, in terms of causal judgement, determine the need and sufficiency of causes.
- Comparing the differences [a-b] and [c-d] between the differences of the frequencies in two rows (or columns).
- Comparing [a+d] and [b+c], which are the cases favouring and contradicting the possible relationship. A variation of this strategy, which consists of computing the difference between the sum of the diagonals $\delta D = [(a+d)-(b+c)]$ was studied by Shaklee and Mims (1982), Arkes and Harkness (1983) and Allan and Jenkins (1983). Jenkins and Ward (1965) observed that this strategy is limited to the cases in which the marginal frequencies are equal.
- The last strategy consists of relating the absolute frequencies of the four cells using multiplication, that is, in comparing the conditional probabilities of a variable, given the alternative values of the other variable. Though this is the correct strategy, few students used this strategy in a spontaneous way.

The influence of previous theories

Association judgements are also influenced by previous theories or expectations,

which according to Shanks (1987) depend on the individual's experience of contingencies between actions and outcomes. Jennings, Amabile and Ross (1982) concluded that correlation is overestimated when previous theories exist. However, a strong correlation between the data is necessary to detect the association when previous theories do not exist, and in this case correlation is underestimated. Chapman and Chapman (1969) studied *illusory correlation*, which can be defined in the following words:

> *"When correlation is perceived on the basis of one's theories, but is not based on empirical facts"* (Murphy, & Medin, 1985, p. 301).

Tversky and Kahneman (1982b) argued that illusory correlation might be explained by the heuristic of accessibility, since:

> *"Lifelong experience has taught us that, in general, instances of large classes are recalled better and faster than instances of less frequent classes; that likely occurrences are easier to imagine than unlikely ones; and that the associative connections between events are strengthened when the events frequently co-occur"* (Tversky, & Kahneman, 1982b, p. 14).

Another related term is the *"illusion of control"*, or

> "*expectancy of a personal success probability inappropriately higher than the objective probability would warrant"* (Langer, 1975, p. 232).

Accuracy of association judgements

Association judgements are influenced by several factors, such as the data format or the strength of correlation. Erlick and Mills (1967) studied the influence of positive or negative correlation and found that some subjects judged negative correlated variables to be independent. Lane, Anderson and Kellan (1985) concluded that graphical format induces judgements of stronger correlation than tabular format.

Finally, Beyth-Marom (1982) differentiated between symmetrical variables, where the values are given the same weight by an individual (e.g., being male – being female) and asymmetrical variables, when its values are not given the same weight by an individual, (e.g., being male – not being male) and found a better perception of correlation in symmetrical variables. According to Nisbett and Ross (1980) negative events have a lesser impact on people's attention than positive events, therefore in asymmetrical variables the two values are given different weights.

2.2. RESEARCH FROM STATISTICS EDUCATION

Within statistics education, research on association has been carried out with pre-University or University students. Of particular interest for researchers' training is the study of preconceptions, since these preconceptions might also be held by future researchers when starting their training in statistics, and therefore, they should be taken into account to organise the teaching of association. This research has focused on the subjects's interpretation of association in contingency tables, their perception of correlation between numerical variables, the comparison of numerical variables in two or more samples and the effect of instruction.

Association in contingency tables

Batanero, Estepa, Godino and Green (1996) studied pre-university students' preconceptions about statistical association by analysing students' strategies in judging the association from a mathematical point of view. They identified three misconceptions of statistical association:

- *Determinist conception of association*. Some students expect a correspondence that assigns only a single value in the dependent variable for each value of the independent variable. When this is not so, they consider there is no dependency between the variables. That is, the correspondence between the variables must be, from the mathematical point of view, a function.
- *Unidirectional* conception of association. Sometimes students perceive dependence only when the sign is positive (direct association), and they consider an inverse association (negative sign) as independence.
- *Local* conception of association. Students form their judgements using only part of the data provided in the problem. If this partial information serves to confirm a given type of association, they adopt this type of association in their answer.

Correlation between numerical variables

The determinist and local conceptions were confirmed by Estepa and Batanero (1996) in their research on the intuitive strategies used by pre-university students (18 years old) upon evaluating correlation between numerical variables represented in a scatter plot. In addition, a new misconception was identified:

- *Causal* conception of association*:* Some students only considered the association between the variables if this could be attributed to a causal relationship between them.

Some of the above conceptions were also found by Morris (1997) in her Psychology students who deduced causality from correlation. She also found that some students believed that positive correlation was always stronger than negative correlation and that a negative correlation indicates independence between the variables.

Truran (1997) evaluated economics and business students' learning of regression. He studied the students' interpretation of the slope and intercept in the regression line, their interpretation of correlation and determination coefficients, and their predictions from the regression equation. Almost all the students in Truran's study identified moderate and negative correlation. Students' answers in making extrapolations were reasonable, as they took into account the strength of correlation and the sample size. However, the author observed a routine learning of the determination coefficient and found the determinist conception of association.

Sánchez (1999) studied the meaning of correlation and regression in undergraduates, and found misconceptions related to correlation and regression, difficulties in translating between the different representations of correlation (verbal description, table, scatter plot and correlation coefficient), difficulties in solving correlation problems and in computing and interpreting the two regression lines.

Comparison of two samples

Estepa, Batanero and Sanchez (1999) described the intuitive strategies employed by pre-University students in comparing two samples (independent and related samples). They found that the most commonly used strategies were different in related and

independent samples and also found the correct and incorrect strategies about association that we have described in the previous paragraphs.

Effect of instruction

Batanero, Estepa and Godino, (1997) and Batanero, Godino and Estepa, (1998) analysed the learning process of University students in some teaching experiences. They identified nine elements of the meaning of association that emerged in specific moments of the process of learning association. They also found that the determinist and local conceptions of association were overcome by the majority of students. The unidirectional conception improved only in some students and the causal conception hardly improved at all in the students taking part in the experiments.

4. THE UNDERSTANDING OF CORRELATION AND REGRESSION IN FUTURE RESEARCHERS AND PROFESSIONALS.

Below we describe an assessment study on the understanding of correlation and regression in future researchers, which is based on the theoretical framework described in Godino and Batanero (1997). In this theoretical framework the *meaning of a mathematical or a statistical object* (for example, association) is conceived as a complex entity, consisting of several interrelated *elements of meaning*. A distinction is also made between the *institutional meaning*, presented to a student in a given teaching institution, and the *personal meaning* of the object which is in fact acquired by the student. Understanding a mathematical object is conceived as the progressive matching of personal and institutional meanings.

The research described in Section 2 has only analysed partial aspects of the meaning of association, with special emphasis in the study of preconceptions before teaching. Our work aimed to evaluate the meaning that a sample of undergraduates give to correlation and regression, after an introductory statistics course, and to study interrelated elements of meaning. Since, in the specialities analysed, the students only receive this statistics course, we also characterise the knowledge about association in researchers entering doctorate and postgraduate courses as well as future professionals. This might be useful when predicting misinterpretations and misuses of association in these future researchers and professionals and when designing courses directed at postgraduates and future researchers.

4.1. METHODOLOGY

The questionnaire and the students' responses to the same are presented in the appendix. The questionnaire was made up of 11 items, each one comprising several subitems in which the students had to choose all the true responses. An additional task involves the ordering of some correlation coefficients (item 1).

Some of the items have been taken from previous research works, such as Tversky and Kahneman (1982 a) and Morris (1997); other items were taken from the book by Cruise, Dudley and Thayer (1984), and the remaining items were prepared by ourselves. The number of correct options in the items is variable in order to reduce random choices. A pilot sample was used to try the questionnaire, adjust wording problems, and study the questionnaire reliability.

The mean frequency (percentage) of correct answers in the items was 102.6 (53.2%),

s = 35.6, which indicates the difficulty that these concepts have for the students. Consequently, we can predict difficulties and errors in the use of correlation and regression in research projects, in analysis of real data, and in interpreting the results. The papers by Bishop and Talbot (2001), Godino, Batanero and Jáimez (2001), and Shimada (2001) also describe researchers' potential errors in applying statistics.

The population intended were the students majoring in business studies or in nursing at the University of Jaén (Spain). The sample of students who completed the questionnaire included 193 students, 104 (37 men and 67 women) from Business Studies and 89 (20 men and 69 women) from Nursing Studies. The average age of these students was 20. All the students had finished the introductory statistics course in the University that year.

In this paper association is studied only from a descriptive point of view and therefore, we do not discuss inferential problems. All our questions concerning association only refer to the specific sample data presented to the students and we never ask them to infer the existence of association in the population where the samples have been taken. In the course followed by the students who answered our questionnaire only a short time was devoted to the study of inference at the end of the year. Even then, the study was restricted to inference for means and proportions and the chi-square test of association. They were not introduced to inference for correlation or regression coefficients. This is an important point to understand our analysis of students' responses and strategies and how we classify them as correct or incorrect.

As regards previous statistics studies, we found that 117 students (60.6 %) had not studied statistics before entering the University; 62 students (32.1%) had studied some statistics in secondary education and/or pre-university studies, and 14 students (7.2 %) had studied statistics in other University courses, in the previous academic year.

Below we analyse the data, describing the meaning given by the subjects to different concepts that intervene in the overall meaning of association. These concepts are: dependence/independence, covariance, correlation and regression.

4.2. DEPENDENCE

We first analyse how the subjects conceived functional and random dependence and independence, as well as types of dependence.

Functional dependence. We analysed whether the students conceive random dependence as an extension of functional dependence, and whether they assign correct values to the correlation coefficient that corresponds to these types of relationships. In our questionnaire (item 7c), 112 students, 58%, asserted that if a functional relationship exists the correlation coefficient takes a value of ±1, and thus not all of them were conscious that a non-linear functional relationship (e.g., parabolic or exponential function) could give a different value to the correlation coefficient.

Linear dependence. Less than 10% of the students in Truran's research (1997) questioned the assumption of linearity. In response 6d, only 33.2% of our students replied that when the correlation coefficient is null the relationship might be non-linear. A related answer is choosing item 7a (independence), in which 138 (71.5%) students replied that if r=0, the variables are independent.

Consequently, many of these students might consider that there is dependence only if the correlation coefficient is different to zero, when analysing dependence between two variables in their professional life. If the correlation coefficient is zero, they might also consider that there is no dependence, and forget non-linear relationships or the possible influence of other variables (item 2b). There also was a low rate of correct

answers in item 2b, where only 11.4% of the students replied that if the covariance is positive the correlation could be non-linear. Although non-linear regression was taken into account in the teaching of these students, linear correlation received more attention, which probably can explain the incorrect answers in these items.

4.3. COVARIANCE

According to Wild and Pfannkuch (1999), recognising the role of variation is a basic element of statistical thinking. Understanding the covariance and related concepts is a first step in the study of correlation and regression. In the questionnaire we included an item (item 2) and two options (4d and 6b) to study the understanding of this concept. In the teaching of the theme to these students covariance was defined as the measure of the combined variability between X and Y, and the relation between the sign of covariance and the type of dependence was also studied. The correlation coefficient was defined by (1).

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \qquad (1).$$

M ore than a half of the students (54.3%) understood that a positive covariance corresponds to positive correlation (item 2) and to a positive correlation coefficient (item 2e). However, in spite of the fact that the slope of the regression line Y/X was defined as σ_{xy} / σ_x^2 in the instruction, only 45.1% of the students of this sample replied that when the covariance is greater than zero the slope of the regression line is positive. The remaining students did not relate the mathematical equations studied (item 2d).

We also emphasise the low rate of correct answers (43.5%) for option 6b, which is an immediate consequence of equation (1), as if $\rho = 0$, $\sigma_{xy} = 0$. Finally, only 17.1% of the students replied that when the strength of the relationship decreases the absolute value of covariance increases (item 4d), where again equation (1) is forgotten.

4.4. CORRELATION

With respect to the understanding of correlation, we analysed the following additional aspects:

The strength of the relationship between two variables depends on the scatter plot spread (item 3d and 4c). We can observe a substantial percentage of students that do not seem to understand this relationship.

The correlation coefficient is non-dimensional (item 3). 44.6 percent of the students in this sample did not recognise this property, consequently they thought that the correlation coefficient depends on the measurement units.

Positive correlation and variability (item 5). In item 5, options a) and c) are equivalent, however the students replied in a different way to both options. The students preferred positive cases to negative ones, a fact observed by Nisbett and Ross (1980) as we have described in section 2.

Strength of correlation (item 1). Ranking several values of the correlation coefficient makes the knowledge of the various types of dependence explicit and serves to assess the students' understanding of strength and sign of correlation. In Table 2 we show the frequencies and percentages of the different students' orderings of correlation coefficients. Morris (1997) gave this same item to her students, although she only asked them to find the correlation coefficient that represented the greatest strength of

correlation. We preferred to ask the students to rank all the values. Taking this difference into account, our results were very close to those of Morris. The percentage of correct answers was 46.1%. 137 students put zero in the last place (71%). These students recognised that r=0 indicates the smallest correlation between two variables.

Table 2. Frequency of Different Rankings of Correlation Coefficients in Item 1

Ranking of correlation coefficient		Frequency	Percentage
(I)	-0.8, 0.5, -0.4, 0.2, 0	89	46.1
(II)	0.5, 0.2, 0, -0.4, -0.8	33	17.1
(III)	0.5, 0.2, -0.4, -0.8, 0	26	13.5
(IV)	0.5, 0.2, -0.8, -0.4, 0	9	4.7
(V)	0, 0.2, -0.4, 0.5, -0.8	3	1.6
(VI)	-0.8, -0.4, 0.5, 0.2, 0	3	1.6
(VII)	Other	23	11.9
No answer		7	3.6
Total		193	100.0

In response [II] the students classify the values applying the usual ordering in R, that is to say, they only take into account the numerical meaning of the correlation coefficient and not its statistical meaning. The knowledge of the order of negative numbers is an obstacle (Brousseau, 1997) when these students tried to rank the strength of correlation. Similar results were obtained by Batanero, Estepa and Godino (1997) and Batanero, Godino and Estepa (1998) when analysing the learning process by some students, and observing their comparison of two negative correlation coefficients. These students did not consider the greatest absolute value to represent the greatest correlation, and they also ordered negative correlation coefficients as if they were just negative numbers.

The students using the rankings (III), (IV) (VI) classified all the values, except zero, according to different types of ordering. In (III) and (IV), the students wrote the positive numbers before the negative numbers. In (III) they classified each subset of numbers according to the usual numerical order, while in (IV) the students consider –0'8 to be greater than –0'4, which shows that these students have not mastered the order in R. The same difficulties are observed in (VI).

Correlation and proportionality. A frequent activity when teaching and learning proportionality is to ask students to discriminate between direct and inverse proportionality. The aim is providing students with tools for dealing with the different types of proportionality. When considering direct proporcionality in a functional setting, we usually characterise a *linear function f(x) = kx*, as a function that verifies the properties (2) and (3).

$$f(x + y) = f(x) + f(y) \quad (2)$$
$$f(\lambda x) = \lambda f(x) \quad (3)$$

With the aim of teaching students how to distinguish direct proportionality, it is frequent to tell them that, in equation (3) an increase (decrease) of a variable ($x \rightarrow \lambda x$) corresponds to a proportional increase (decrease) of the other ($f(x) \rightarrow \lambda f(x)$).

The analogy with correlation, and the use of the similar wording (direct and inverse correlation / direct and inverse proportionality), may lead some students to confuse proportionality and correlation. It is not surprising that some students compare the correlation coefficient r with a proportionality constant. This comparison can be done in

standardised variables, where the regression equation is reduced to a homogeneous linear function and r is the constant of proportionality that relates the value X to the average of the distributions Y conditioned to X, but not in the general case. However, 22.8 % of our students answered that *"if r = 0'6 the correlation between the variable X and Y is double that when r = 0'3"* (item 7b), and thus these students extended proportionality to the general case where it is not true. Our interpretation is reinforced by the number of correct answers in items 5a and 5c.

Confusion between r and r^2. The determination coefficient is a measure of the goodness of fit of the scatter plot to a straight line. It also expresses the reduction of variance when predicting the *y* value by using the regression line Y/X instead of the equation $y = \bar{y}$. Some of our students (10.4%) confused the correlation coefficient r with the determination coefficient r^2, since they chose item 7d, where it is said that r can be interpreted as a percentage of the variance.

Value of the correlation coefficient and its relationship with both regression lines (item 6, 11 and 12). Only half the students in the sample recognised the perpendicularity of the regression lines when the correlation coefficient is zero, while a seventh of the students (14.5%) replied that in the case of a null correlation coefficient, both regression lines have the same slope, that is to say, they are parallel (option a of items 6 and 11). Only 38.9% of the students recognised that if the two regression lines have the same slope then the correlation coefficient is +1 or -1, choosing the two correct options (options b and c, item 11). However, 37.3% of the students only accepted a correlation coefficient value of +1 in this case. We confirm here some resistance to accept the inverse correlation in agreement with Batanero, Estepa, Godino and Green, (1996) and Morris, (1997). Finally, 2/3 of the students recognised that when correlation is perfect the two regression lines are parallel (item 12).

4.5. CORRELATION AND PREDICTION

A main objective in both scientific research and decision making is to find causal relationships between variables through the analysis of association. However, there are different types of relationships that can explain correlation, such as unilateral dependence, interdependence, concordance, or spurious correlation. In the first case we need to distinguish between the dependent variable and the independent variables. Below we study the students' understanding of these questions, which are essential in many types of research (items 8, 9 and 10).

Correlation and causality (item 8). The causal conception of association (identifying association and causality) has been described in Estepa and Batanero (1996) and Morris (1997) and it might be influenced by the instruction that the students receive in mathematics, science and other areas, where all the phenomena are given a causal explanation. The students transfer their ideas from other topics to the study of association, and when finding a strong association between two variables they might infer there is a causal dependence between the variables.

In item 8c, a high percentage of students (89,1%) recognised that a strong relationship is expected when there is a high value of the correlation coefficient; however this percentage is considerably reduced (29.0%) when students have to recognise that other factors can influence the results (option a). There were also 22.3% of the students who accepted with certainty that double the surface sown corresponded to double the crop, expressing a causal conception of association. This fact is reinforced when we observe that 36 students (18.6%) chose simultaneously answers c) and d) in

item 8, which seems to indicate that these students reasoned that, as the correlation coefficient was high, it could be asserted with certainty that twice the planted surface will correspond to twice the crop.

Distinction between independent and dependent variables (item 10). A previous step in a prediction problem is to distinguish between the independent and the dependent variables. This distinction is essential to discriminate the two regression lines, which according to Sánchez (1999) are only correctly distinguished by 32.0% of undergraduates. The low percentage of correct answers (36.3%) in this item confirms Sánchez's (1999) results as 73.7% of the students in our sample did not discriminate between the variables and 34.2% confused the two regression lines. This fact will have a negative influence in the use of the regression equations for predictions in research projects.

Interdependence (item 9). In functional dependence independent and dependent variables are univocally determined. Association extends the idea of functional dependence although, in this case, the variables are not always univocally determined, since in situations of interdependence, both variables can play either the role of dependent or independent variable. This situation is presented in item 9, where almost 60% of the students accepted interdependence, although 41.4% preferred to fix one variable as independent and the other as dependent. We note that while height was chosen as an independent variable by 34.7% of the undergraduates, weight was chosen as an independent variable only by 6.7% of students. This possibly can be due to the fact that height does not usually vary in an adult, while weight normally does.

5. IMPLICATIONS FOR THE TRAINING OF RESEARCHERS

In this paper we have summarised the main research results on the understanding of association, including our empirical study carried out with undergraduate students, whose results can be applied to future researchers and professionals. Many research projects are intended to find related variables and to establish causal relationships between them, in such a way that a response variable can be explained or predicted by one or several explanatory variables. Consequently, a correct understanding of statistical association and of all its interrelated elements of meaning is basic in research methodology. In spite of this, undergraduates do not always acquire a correct understanding of association and show misconceptions, some of which are not always overcome after instruction. Association judgements are influenced by previous theories and misconceptions, and incorrect strategies are sometimes used to carry out association judgements.

Our research results suggest that some elements of the meaning of association are only acquired by a few students. The difficulty of the concepts related to association (such as correlation, covariance, and regression line) was greater than predicted, since the mean rate of correct answers in the questions dealing with these concepts is slightly superior to fifty percent (53.2%).

The undergraduate's difficulties in relating linear regression, correlation coefficient and covariance should warn us about the need to emphasise these fundamental relations in the teaching of association and base this teaching on the understanding of covariance, as a measure of the joint variation. Few undergraduates in our study showed a correct knowledge of the properties of the correlation coefficient. In particular non-dimensionality, strength of correlation, and negative correlation were scarcely understood. The existence of possible obstacles associated to these mistakes suggests

that these difficulties could possibly be overcome through didactical situations that produce cognitive conflicts.

We have also found confusion between association and causality; lack of distinction between interdependence and unilateral dependence, problems in adequately choosing the dependent and independent variables; and excessive emphasis on linear dependence. Linear dependence is sometimes an oversimplification, and this should be emphasised in teaching, where students should be encouraged to explore other models, which nowadays is easy with the help of computers.

Without a full integration of the elements of meaning, the conceptions of correlation and regression acquired by future researchers will be biased and incomplete, and will produce improper uses of statistics and, consequently, incorrect research conclusions. Therefore, these points should be taken into account in the researcher's statistical instruction. The planning of the researchers' training in statistics should take into account the elements of meaning of correlation and regression and psychological and didactic research results, in order to improve the results of the instruction.

ACKNOWLEDGEMENTS:

This research was supported by the grant PB97-0851, Secretaría de Estado de Universidades, Investigación y Desarrollo, Spain.

APPENDIX: QUESTIONNAIRE AND RESULTS

(Correct answers are marked with x)

Item 1. Order the 5 following correlation coefficients from the one that indicates the highest amount of correlation to that which indicates the lowest amount of correlation or no correlation: 0'5, -0'8, 0'2, -0'4, 0.

	Item 2. When the covariance between X and Y is greater than 0, then:	Frequency	Percentage
x	a) The correlation between X and Y is positive	127	65.8
x	b) The relationships between X and Y might be non-linear	22	11.4
	c) The variables might be not interrelated	17	8.8
x	d) The regression line has a positive sign slope	87	45.1
x	e) The correlation coefficient is positive	115	59.6

	Item 3. John correlated height and weight of graduate male students. He used meters and kilograms as his measures. Angela also correlated height and weight on the same group of subjects using centimetres and grams as her measures. John and Angela computed correlation coefficients between their two sets of measures.	Frequency	Percentage
	a) Angela's correlation coefficient will tend to be greater than John's	23	11.9
x	b) The two correlation coefficients will be approximately equal	107	55.4
	c) John's correlation coefficient will tend to be greater than Angela's	17	8.8
x	d) The value of the correlation coefficient depends on the data spread	102	52.8

Item 4.	When the strength of the relationship between two variables decreases:	Frequency	Percentage
	a) The slope of the regression line of Y/X increases	23	11.9
	b) The slope of the regression line of X/Y increases	32	16.6
x	c) There is greater spread in the scatter plot	124	64.2
	d) The absolute value of the covariance increases	33	17.1

Item 5.	If two variables are positively correlated:	Frequency	Percentage
x	a) As one increases, the other increases	152	78.8
	b) As one decreases, the other increases	8	4.1
x	c) As one decreases, the other decreases	90	46.6
	d) There is a linear relationships between the two variables	91	47.2

Item 6.	If the correlation coefficient between two variables is zero:	Frequency	Percentage
	a) Both regression lines Y/X and X/Y are parallel	28	14.5
x	b) The value of the covariance is zero	84	43.5
	c) Both regression lines Y/X and X/Y coincide	19	9.8
x	d) The variables may have a non-linear relationship	64	33.2
x	e) Both regression lines Y/X and X/Y are perpendicular	96	49.7

Item 7.	If *r* is the correlation coefficient of two variables:	Frequency	Percentage
	a) If *r* = 0 the variables are independent	138	71.5
	b) When *r* = 0.6 there is twice as much correlation between the variables X and Y than when r = 0.3	44	22.8
x	c) When there is a perfect lineal relationship between the variables, *r* is +1 or - 1	112	58.0
	d) The correlation coefficient can be interpreted as a percentage of the variance	20	10.4

Item 8.	A farmer studied the wheat surface sown in thousands of hectares and the crops obtained in millions of metric quintals, over five consecutive years, and he obtained a correlation coefficient of 0.91. Consequently,	Frequency	Percentage
x	a) There may be other factors that make the results vary	56	29.0
	b) We should have to take a larger sample to be able to express the relationship between the planted surface and the obtained crop	6	3.1
x	c) There is a strong correlation between the crop and the planted surface	172	89.1
	d) If we plant double the surface, we will double the crop with absolute certainty	43	22.3

Item 9.	In which forecast do you have more confidence?	Frequency	Percentage
	a) Estimating the height of a man from his weight	13	6.7
	b) Estimating the weight of a man from his height	67	34.7
x	c) The two estimations are equally reliable	114	59.1

		Frequency	Percentage
Item 10.	In a study incomes are used to predict savings. Both variables are measured in thousands of pesetas. Which of the following statements is true if y = 1000 + 0.1x is the regression equation?		
	a) y is the income, x is the saving, the income is the independent variable	26	13.5
	b) y is the income, x is the saving, the saving is the independent variable	66	34.2
	c) y is the saving, x is the income, the saving is the independent variable	56	29.0
x	d) y is the saving, x is the income, the income is the independent variable	70	36.3

		Frequency	Percentage
Item 11.	If both regression lines have the same slope, what is the value of r ?		
	a) 0	29	15.0
x	b) 1	149	77.2
x	c) –1	80	41.5
	d) 0.5	7	3.6

		Frequency	Percentage
Item 12.	If X and Y have perfect correlation (r=1 or r=-1), the angle between the two regression lines is:		
	a) 120º	5	2.6
	b) 90º	32	16.6
	c) 45º	24	12.4
x	d) 0º	129	66.8

REFERENCES

Allan, L. G., & Jenkins, H. M. (1983). The effect of representations of binary variables on judgement of influence. *Learning and Motivation*, 14, 381-405.

Arkes H. R., & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General*, 112(1), 117-135.

Batanero, C., Estepa, A. Godino, J. D., & Green, D. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, 27(2), 151-169.

Batanero, C., Estepa, A., & Godino, J. D. (1997). Evolution of students' understanding of statistical association in a computer based teaching environment. In J. B. Garfield, & G. Burril (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 191-205). Voorburg, The Netherlands: International Association for Statistical education and International Statistical Institute.

Batanero, C., Godino, J., & Estepa, A. (1998). Building the meaning of statistical association through data analysis activities. In A. Olivier, & K. Newstead (Eds.), *Proceedings of the 22nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 221-236). Stellenbosh, South Africa: University of Stellenbosh.

Bangdiwala S. I. (2001). Training of statisticians worldwide to collaborate as co-investigators within country clinical epidemiology units: The experience of the International Clinical Epidemiology Network (INCLEN). In Batanero, C. (Ed.), *Training researchers in the use of statistics* (pp. 265-275). Granada: International Association for Statistical Education and International Statistical Institute.

Beyth-Marom, R. (1982). Perception of correlation reexamined. *Memory and Cognition*, 10(6),

511-519.

Bishop, G, & Talbot, M. (2001). Statistical thinking for novice researchers in biological sciences. In Batanero, C. (Ed.), *Training researchers in the use of statistics*. (pp. 215-226). Granada: International Association for Statistical Education and International Statistical Institute.

Blumberg, C. J. (2001). Training regular education and special education teachers in the use of research methodology and statistics. In Batanero, C. (Ed.), *Training researchers in the use of statistics*. (pp. 231-244). Granada: International Association for Statistical Education and International Statistical Institute.

Brousseau, G. (1997). *Theory of didactical situations in mathematics*. Dordrecht: Kluwer.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an abstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 74, 271-280.

Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin*, 90(2), 272-292.

Cruise, R. J., Dudley, R. L., & Thayer, J. D. (1984). *A resource guide for introductory statistics.* Dubuque, IO: Kendall / Hunt.

Erlick, D. E., & Mills, R. G. (1967). Perceptual quantification of conditional dependency. *Journal of Experimental Psychology*, 73(1), 9-14.

Estepa, A., & Batanero, C. (1996). Judgments of correlation in scatter plots: An empirical study of students' intuitive strategies and preconceptions. *Hiroshima Journal of Mathematics Education*, 4, 25-41.

Estepa, A., Batanero, C., & Sanchez, F. T. (1999). Students' intuitive strategies in judging association when comparing two samples. *Hiroshima Journal of Mathematics Education*, 7, 17-30

Godino, J. D., & Batanero, C. (1997). Clarifying the meaning of mathematical objects as a priority area for research in mathematics education. In A. Sierpinska, & J. Kilpatrick (Eds.), *Mathematics education as a research domain: A search for identity* (pp. 177-195). London: Kluwer.

Godino, J. D., Batanero, C., & Gutiérrez Jáimez, R. (2001). The statistical consultancy workshop as a pedagogical tool. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 339-353). Granada: International Association for Statistical Education and International Statistical Institute.

Inhelder, B., & Piaget, J. (1955). *De la logique de l'enfant á la logique de l'adolescent* [From the child's logic to the adolescent's logic]. Paris: Presses Universitaires de France.

Jenkins, H. M., & Ward, W. C. (1965). Judgment of de contingency between responses and outcomes. *Psychological Monographs*, 79, 1-17.

Jennings, D. L., Amabile, T. M., & Ross, L. (1982). Informal covariation assessment: Data-based versus theory-based judgments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 211-230). New York: Cambridge University Press.

Kahneman, P. Slovic, & A. Tversky (Ed.), (1982). *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Lane, D. M., Anderson, C. A., & Kellam, K. L.(1985). Judging the relatedness of variables: The psychophisics of covariation detection. *Journal of Experimental Psychology. Perception and Performance*, 11(5), 640-649.

Langer, E. J. (1975). The illusion of control. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 231-238). New York: Cambridge University Press.

Morris, E. J. (1997). An investigation of students' conceptions and procedural stills in the statistical topic correlation. *Centre for Information Technology in Education*, Report n. 230. Milton Keynes, U.K: The Open University.

Murphy G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92(3), 289-316.

Nisbett, R., & Ross, L. (1980). *Human inference: strategies and shortcomings of social judgment.* New Jersey: Prentice Hall.

Pérez Echeverría, M. P. (1990). *Psicología del razonamiento probabilístico*. [Psychology of probabilistic reasoning]. Madrid: Ediciones de la Universidad Autónoma.

Peterson C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, 68(1), 29-46.

Sanchez, F. T. (1999). *Significado de la correlación y regresión para los estudiantes universitarios* [Meanings of correlation and regression for undergraduates]. Unpublished PhD. University of Granada.

Shaklee, H., & Mims, M. (1982). Sources of error in judging event covariations: Effects of memory demands. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 8 (3), 208-224.

Shanks, D. R. (1987). Acquisition functions in contingency judgment. *Learning and Motivation*, 18, 147-166.

Shimada T. (2001) Precaution against errors in using stochastic software. In Batanero, C. (Ed.), *Training researchers in the use of statistics.* (pp. 129-137). Granada: International Association for Statistical Education and International Statistical Institute.

Truran, J. M. (1997). Understanding of association and regression by first year economics students from two different countries as revealed in responses to the same examination questions. In J. Garfield, & J. M. Truran, (Eds.), *Research papers on stochastics educations from 1997* (pp. 205-212). Minneapolis, MN: University of Minnesota.

Tversky, A., & Kahneman, D. (1982). Causal schemas in judgments under uncertainty. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 117-128). New York: Cambridge University Press.

Tversky, A., & Kahneman, D. (1982b). Judgment under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*, (pp. 3-20). New York: Cambridge University Press.

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223-248.

Antonio Estepa,
Facultad de Humanidades y Ciencias de la Educación
Universidad de Jaén, Spain
E-mail: aestepa@ujaen.es

Francisco Tomás Sánchez Cobo
Escuela Politécnica Superior
Universidad de Jaén, Spain
E-mail: fsanchez@ujaen.

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 53-63.

Printed in Granada, Spain

CHIHIRO HIROTSU

STATISTICAL TRAINING OF RESEARCHERS IN TOTAL QUALITY MANAGEMENT: THE JAPANESE EXPERIENCE

A training system for statistical methods in Total Quality Control or Total Quality Management is discussed and we suggest what and how to teach. It is stated that we have no department of statistics in the universities in Japan and stressed that applied statistics is most efficiently taught to those who have their own problems and motivations to apply these statistical methods. It is then essential for a company to have their own training systems for the TQM researchers although some extra company training courses may also be efficiently utilised. As an example we introduce in some detail the seminars provided by JUSE as well as in-company training systems of Toyota Motor Corporation and Takenaka Corporation.

1. INTRODUCTION

In this paper we consider a training system for statistical methods in TQC (Total Quality Control) or TQM (Total Quality Management). Two important aspects of the system are what and how to teach. The success of quality control in Japan is due to the company-wide activities, which involve all the staff and departments in a company and do not just depend on a few experts. It is also due to the natural tendency of the Japanese to be very diligent, generally clever and willing to devote themselves to the company.

Each company has a statistics section as a part of the QM promotion section. Ideally a company should have a TQM promotion team involving several advisors and trainers who are expert in the area and can teach these statistical methods. However, some elementary courses may be more efficiently taught in Japan by an external institution such as JSA (Japanese Standards Association) or JUSE (Japanese Union of Scientists and Engineers). Such institutions are particularly useful in Japan since there is no department of statistics in the universities and statistical methods are very poorly taught.

Now I describe five courses to learn the statistical methods that are most useful in practice:

1. *Elementary statistics:* Basic idea of variations in data, statistical estimation and tests, concept of TQM, basic tools such as QC seven tools and control charts;
2. *Design of Experiments:* One- and two-way layouts, split plot design, hierarchical design, orthogonal array, analysis of variance (ANOVA), reliability analysis;
3. *Multivariate Analysis*: Regression analysis, discriminant analysis, principal component analysis, correspondence analysis, cluster analysis, contingency tables;
4. *Advanced:* Beyond ANOVA techniques, graphical modelling, GLM, GAM, Multiple correspondence analysis, Taguchi method;
5. *Applications*: Problem solving by integrated use of various statistical methods.

The first three courses might be taught by some external institution, but the last two should be taught within the company and should be based on the researcher's own problems. It is then desirable to have convenient tailor-made software for statistical analysis and the database of the company's past achievements. It should be stressed here that the CWQC (Company-Wide Quality Control) in Japan has been successfully developed by all the people within a company, by applying statistical methods to his or her own problem even though the methods used might be very elementary. It should also be noted that a recent trend is to apply statistical approaches not only to the manufacturing processes but also to the planning, marketing and management processes of the company. It is also essential to have the hierarchical education system in a company for maintaining its statistical activities. One of the most successful examples in Japan is the Toyota System.

Finally an annual company-wide conference is very useful to give people in the company an opportunity to present their statistical activities to the top management of the company and to promote their statistical activities. A presidential award might be given to the best achievement.

2. GENERAL STATISTICAL BACKGROUND IN JAPAN

We will begin by describing the general background of statistics education in Japan. One of the most prominent characteristics is that there is no department of statistics at Japanese universities and that statisticians are scattered around various faculties forming very small research teams.

There was a very hot discussion on this subject a long time ago, when it was decided to distribute the statistics offices (called koza in Japan) over the various faculties requiring the study of statistics within their own field, instead of having a concentrated statistical department. A koza has been composed of one professor, one associate professor and two research associates.

To give an example, at the University of Tokyo about 15 professors and associate professors of statistics are working in the Faculties of Economics, Engineering, Medicine, Agriculture, Education, Mathematics, and Culture. In my experience as a Professor of the Department of Mathematical Engineering at the University of Tokyo, I took charge of a laboratory composed of one associate professor, one research associate and about ten doctoral and master students including a few from companies. There is only one statistics laboratory among more than two hundred laboratories in the Faculty of Engineering at the University of Tokyo. It may be surprising that we have only one professor and one associate professor among approximately 400 faculty members in the very big Faculty of Engineering. We have, however, several additional statistics laboratories in the Faculties of Economics, Medicine, Science, Agriculture, Education and Culture and we organise an inter-faculty statistics meeting once a week and collaborate to educate graduate students. In this sense the University of Tokyo is rather favoured and I am afraid that the case will not be the same for other many universities.

Professional statisticians are usually brought up in the statistics laboratories scattered in various faculties in the universities as in the example of the University of Tokyo. The number and the range of lectures are usually not enough and students read books themselves or in small groups, attend seminars and discuss their notes with their supervisors. There is no particular external consulting service for researchers in the universities. Of course we give advice on their request, though this is not often needed since, at least in the Faculty of Engineering, researchers are usually capable enough to

solve their statistical problems by themselves with the aid of some statistical package.

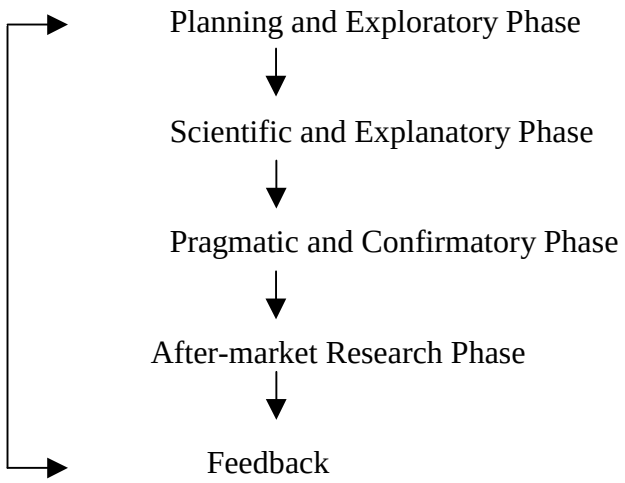
Therefore, we think it is important to have a weekly inter-faculty statistics seminar for graduate students. We have many opportunities to present our respective problems and ideas to our colleagues and obtain suggestions from them, and sometimes this naturally leads to collaborative work. Those opportunities include seminars and symposiums.

Most undergraduate students, however, take only an elementary statistics course during their studies except those students who belong to particular departments where there is statistics staff. They only have a poor concept of variations in data and an elementary knowledge of statistical tests and estimation. The general backgrounds of the researchers who perform the Total Quality Management in the company in Japan will be mechanical, civil and electronical engineering, chemistry, architecture and so on. Even when I give advice to graduate students from other departments on their requests this is far from sufficient. It is thus essential to have the statistical training courses outside universities for researchers in companies who did not receive any proper statistics courses in universities.

However, this is not a major defect in Japan since applied statistics can be most efficiently taught when students have their own problems and motivations. In my experience, for example, it is much more difficult to teach the idea of multiple comparisons procedure to students in a classroom than to explain those ideas to researchers in pharmaceutical companies who are dealing with various types of multiplicity problems in their ordinary research work, such as multiple endpoints, subgroup analyses and interim analyses.

It is therefore possible for a researcher to learn statistics methods after he or she has been involved in some department of a company and has realised the problems to be solved there. We also note that the Deming Prize Application has been useful in Japan to motivate people in companies to learn statistics (see the special issue: The Deming Prize edited by Okuno, 1990-1991).

Fig.1 The Four Phases of R & D Activities



One thing I should stress here is that a researcher in a company should not be an individual data analyst, but should relate his or her research to preceding and succeeding works. Any research and development (R&D) activity has four steps of exploration, explanation, confirmation and after-market research, and thus the information obtained by the after-market research should give feedback promptly to the first step of planning,

as it is shown in Figure1.

In each phase the type of data might be different and even with the same data the approach to the data and the decision based on the data might be different (Hirotsu, 1992). An example of this could be the difference between Phase II and Phase III of clinical trials in the stream of new drug development, which are sometimes referred to as explanatory and pragmatic phases. To perform his or her role appropriately, it is therefore essential for a researcher to be aware of the stage he or she is in the stream of R & D. This implies the necessity of an in-company training course at least in the final stage of education of applied statistics, and also suggests a need for a general manager to supervise the whole process of R & D.

Now under the circumstances of Japan and the characteristics of applied statistics, the need of some extensive training system for people to perform TQM in companies is obvious.

3. TQM EDUCATION COURSES HELD OUTSIDE COMPANIES

In Japan we have many TQM education courses outside companies. Typical and extensive examples are the courses provided by JUSE and JSA (see Ishikawa, 1969 and Mizuno & Kume, 1978). There have been, however, several changes since these papers and the current status of JUSE is described in some detail below.

A variety of systems of education courses exist, such as post-oriented, division-oriented, theme-oriented, methodology-oriented courses, statistical software courses and a correspondence course. There are also various levels from elementary to advanced, which include also rather philosophical seminars to introduce the concept of TQM as well as more technical statistical seminars. Since it is important to maintain the training system successfully in a company, top management of the company should be aware of the relevance of applying statistics fully in the R & D activities. It should also be noted that there are courses provided not only for the manufacturing processes but also for the planning, marketing and management processes.

3.1. POST ORIENTED COURSES

1. *Top Management Course* (intensive, 9 hrs.×4 days): Introducing the managing director to management and TQM for the promotion of company-wide quality management activities.

2. *Executive Management Course* (intensive, 9 hrs.×4 days): Introducing the general manager to planning and implementing TQM.

3. *Senior Management Course* (6 hrs.×3 days): Introductory course for senior managers to the basic principles of TQM and TQC.

4. *Middle Management Course* (6 hrs.×9 days): Practical course for middle managers to promote TQM in their respective departments.

5. *Chief Basic Course* (6 hrs.×6 days): Role of chief staff in the ordinary quality control activities.

3.2. DIVISION ORIENTED COURSES

1. *TQM Instructor courses* (6 hrs.×6 days): Methods of introduction and promotion of TQM for TQM instructors with basic knowledge of TQM and TQC.

2. *Procurement Department Course* (6 hrs.×4 days): Purchasing and logistics service control for value engineering and cost reduction.

3. *Elementary Course for Sales Department* (6 hrs.×4 days): Concept of TQM and QA (Quality Assurance) in sales department.

4. *Advanced Course for Sales Department* (6 hrs.×8 days): Roles of sales department for TQM and the current method of QA for customer satisfaction.

5. *QC Seminar for Good Manufacturing Practice* (6 hrs.×3 days): Necessary knowledge of GMP (Good Manufacturing Practice) to promote TQM and QA in manufacturing and selling foods and drugs.

3.3. THEME ORIENTED COURSES

1. *Policy and Planning Seminar* (6 hrs.×3 days): Method and organisation for determining the management, quality and quality control policies of the company and for transmitting them throughout all the company sectors.

2. *Introductory Course for TQM* (6 hrs.×3 days): Basic concept of TQM, quality and control; Method of problem solving and approaching a project.

3. *Cost Down Seminar* (6 hrs.×6 days): Basic concept, promotion and method of cost down in manufacturing planning and purchase departments.

4. *QC Story Seminar for Achieving a Management Project:* An approach and know-how for innovating the business based on the company top management policy.

5. *Introductory Course for Product Liability* (6 hrs. ×3 days): Current status of the law and system for product liability; Experiences and measures to solve the product liability problems.

6. *Advanced Course for Product Safety:*
   A. *Product Safety Technology Course* (6 hrs.×2 days): Guidelines of product liability for engineers in planning, design, research and development, quality assurance and quality control.
   B. *Product Safety Co-ordinator Course* (6 hrs.×2 days): Roles of the product safety co-ordinator in product safety; Designing the product safety review system and the document safety system.

7. *R & D Management Seminar:* Management of research and development; Method of new product development, market research and new product planning.

3.4. METHODOLOGY ORIENTED COURSES (ELEMENTARY)

1. *QC Seminar Basic Course* (6 hrs.×30 days): Seminar of quality control concepts and theory and application of statistics for engineers and staff with at least 3 years business experience; Lectures, practice with personal computer and QC games for basics statistics methods, statistical test and estimation, design of experiments, regression analysis, reliability engineering, sensory test, feeling evaluation and so on.

2. *QC Seminar Elementary Course* (6 hrs.×8 days): Basic concept of quality control and elementary statistics methods including QC seven tools, collecting and summarising data, test and estimation, analysis of variance and correlation and regression analyses.

3. *QC New Seven tools* (6 hrs.×3 days): Affinity chart method, relation chart method, system chart method, arrow diagram method, process decision program chart (PDPC), matrix chart and matrix data analysis.

4. *Seminar for Computer Application for Problem Solving* (6 hrs.×2 days): Problem solving, decision making and information system.

5. *Quality Function Deployment (QFD) Seminar*
   5.1. *QFD Practice Course* (6 hrs.×2 days): Practice of QFD application, making two-way tables and problem solving.
   5.2. *QFD Introductory Course* (6 hrs.×4 days): Outline and utility of QFD.

6. *Strategy Planning Seminar for Policy Management* (6 hrs.×2 days): Framework of planning strategy, environmental analysis, product analysis, market analysis, allocating resources, analysis of strategy factors; case studies.

7. *Product Planning Seven Tools*
   7.1. *Introductory Course* (6 hrs.×4 days): Seven tools for producing hit product; Group interview, questionnaires, positioning analysis, imaginary method, joint analysis, product planning based on marketing; case studies.
   7.2. *Quick Course* (6 hrs.×1 days): Outline of seven tools for product planning.

3.5. METHODOLOGY ORIENTED COURSE (ADVANCED)

1. *Design of Experiment Seminar* (1) (7 hrs.×8 days): Role experimental design, mean and variance, test and estimation, 1-way layout, 2-way layout, split plot design, orthogonal array, theory of ANOVA, correlation analysis, simple regression analysis.

2. *Design of Experiment Seminar* (2) (7 hrs.×12 days, 4 days per a month): Multi-way layout, advanced orthogonal array, non orthogonal experiment, sequential experiment, mixed experiment, random effects model, optimisation of multiple-end variables, Taguchi method, multiple regression analysis, analysis of proportions.

3. *Multivariate Analysis* (1) (7.5 hrs.×4 days): Introduction to multivariate analysis, principal component analysis, variable selection in regression analysis, logistic

regression analysis.

4. *Multivariate Analysis* (2) (7.5 hrs.×4 days): Latent structure analysis of categorical data, graphical modelling, canonical correlation analysis, covariance structure analysis integrating regression analysis and factor analysis, data mining.

5. *Statistical Methods for Clinical Trials Seminar* (1) (6 hrs.×7 days): Introduction to clinical trials, planning, designing, elementary statistical methods including non-parametric method and cross-over design.

6. *Statistical Method for Clinical Trials Seminar* (2) (6 hrs.×24 days, 2 days per month): Introduction to statistical inference, regression analysis, ANOVA, analysis of categorical data, analysis of survival data, dose-response analysis, sample size determination, meta-analysis, statistical guideline for regulation.

7. *Data Management in Clinical Trials Seminar* (camping system, 6 hrs.×5 days): Outline of data management in clinical trial.

3.6. STATISTICAL ANALYSIS SOFTWARE SEMINARS BASED ON JUSE-QCAS OR JUSE-MA

1. *QC Practice Seminar* (6 hrs.×3 days): Process analysis, problem solving, QC seven tools, and regression analysis.

2. *Design of Experiment Seminar* (6 hrs.×3 days): Factorial experiments, orthogonal array, QC game.

3. *Multivariate Analysis Seminar* (6 hrs.×3 days): Principal component analysis, multiple regression analysis, and correspondence analysis.

4. *Reliability Analysis Seminar* (6 hrs.×2 days): Analysis of reliability data and field data.

5. *Seminar for Questionnaire Planning and Its Analysis by Personal Computer* (6 hrs.×2 days): Application of multivariate analysis to the analysis of questionnaires.

3.7. CORRESPONDENCE COURSE (6 MONTHS)

This course is based on two textbooks, one for methods and the other for practice of quality control.

Similarly the Japanese Standards Association (JSA) provides some standard courses, in particular, ISO 9000 and ISO 14000 seminars.

4. IN-COMPANY TQM EDUCATION AND TRAINING

Although these external seminars provide a very good opportunity for TQM education and training the internal education of people in a company is even more important for practising these methods and techniques in their ordinary activities.

Most companies, if not all, arrange education and training courses in TQM for their

employees. Ideally for in-company education a company should be equipped with:

1. A hierarchical education system;
2. Tutors with various achievement levels;
3. Taylor-made software for statistical analysis;
4. Database of company's past projects and case studies;
5. Annual company-wide conference for statistical activities.

In this section we describe two characteristic cases of in-company education system.

4.1. THE CASE OF TAKENAKA CORPORATION

The Takenaka Corporation was the winner of the first Deming Prize in the construction sector and should be regarded as the leader of the sector. Its education schedule has been introduced by Jido (1990-91), from which we reproduce his Table 3.2 (Table 1 here).

We can see from Table 1 that the Takenaka Corporation is giving in-company seminars by in-company instructors and extra professionals for its employees to learn the TQC (TQM) concepts and statistical methods as well as using extra seminars provided by JUSE and JSA. It should be noted that a hierarchical system is taken so that senior instructors who have finished an advanced course can teach the elementary course. It is essential for the staff and foremen to learn statistical methods based on their own problems. A more recent example of this approach is seen at the Toyota Motor Corporation.

4.2 CASE OF TOYOTA MOTOR CORPORATION

According to the highly stable condition of manufacturing processes in Japan a recent tendency of TQM is changing from statistical approaches to a more philosophical (or conceptual) approach with slogans such as customer's satisfaction, market in (rather than product out), source control and so on. It is, however, obvious that the philosophy of TQM can only be carried out with the scientific approach. Furthermore the recent development of statistical methods has enabled us to handle new types of problems and data coming out of off-line as well as on-line processes. It is therefore very inappropriate to adhere to the classical SQC (Statistical Quality Control) approach and it is strongly recommended to go beyond it. Under these circumstances Toyota's approach is remarkable in that it is convinced of the necessity of the new scientific SQC method and it is practising it. We will briefly introduce the system here and refer to Amasaka and Osaki (1999) as well as to Amasaka et al. (1999) for details.

First, Toyota has developed its own methodology called 'SQC Technical Methods' integrating statistical methods such as Seven New Tools and other basic SQC methods, multivariate analysis and design of experiments with engineering technology, which can be used efficiently and appropriately at each step of problem solving in the course of research, development, manufacturing and marketing. This is carried out by assessing a one shot analysed with a ready made statistical method. They call it mountain climbing for problem solving by use of 'SQC Technical Methods'.

To support the efficient utilisation of the 'SQC Technical Methods' the integrated SQC network TTIS (Toyota SQC Technical Intelligence System) has also been developed. It is composed of TSIS (Toyota SQC Intelligence System), TPOS (Toyota TQM Promotional SQC Original Soft), TSML (Toyota SQC Manual Library) and TIRS

(Toyota Information Retrieval System).

Table 1. QC Education Schedule in Takenaka Corporation (Table 3.2 of Jido, 1990-1991)

HIERARCHY	PURPOSE	IMPLEMENTATION PROCEDURE	
		Seminar	Follow-up
Directors	To acquire knowledge to evaluate TQC activities as top management	Director Special Course (JUSE)	To enhance knowledge through attending President Diagnoses & Consultations.
General managers	To acquire fundamental knowledge and concept of TQC, as “upper” middle management.	Executive Course (JUSE)	To hold Branch General Manager’s QC Diagnoses and Consultation To participate in various QC Diagnoses and Consultations. To present the outcome of TQC activities at in-house gatherings and conventions.
Senior managers	To acquire principal knowledge and basic statistical methods of TQC as middle management.	Manager Course (JUSE & JSA) In-house TQC Manager Seminar (5 days)	
Managers			
QC Specialists	To acquire the TQC concept, statistical methods and other professional knowledge becoming QC promoter in his department.	Various outside seminars (JUSE & JSA)	
Engineers	To acquire the TQC concept and statistical methods.	In-house TQC Basic Course (B) 15 days	To present QC activities at various gatherings and conventions.
Administrators	To acquire the TQC concept and basic statistical methods.	In-house TQC Basic Course (A) 10 days	
Staff members	To acquire the TQC concept and often-used QC techniques	In-house TQC Elementary Course 3 days	
Clerical workers	To acquire the TQC concept and knowledge required for QC circle activities	Seminars and lectures conducted by in-house instructors	QC circle gatherings and conventions.
New recruits	To acquire basic TQC concept	In-house TQC Orientation Course 1 day	

JUSE: Union of Japanese Scientist & Engineers, JSA: Japanese Standards Association

TPOS is the friendly tailor-made software of Toyota and it is composed of TPOS-PM (Multivariate Analysis), TPOS-PS (General SQC Methods), TPOS-PO (design of experiment), TPOS-PK (sensitivity analysis) and TPOS-PR (reliability analysis). Multivariate analysis, for example, contains discriminant analysis, multiple regression analysis (1), (2) and principal component analysis. One can refer to various successful applications in real business through TSIS and also find past successful examples of problem solving in Toyota by TIRS. To sum up TTIS is, as stated in Amasaka and Osaki (1999), the intelligent system for SQC applications consisting of four main systems synthesised to grow while supplementing one another. TTIS has been very efficiently used in in-company education and training of SQC in Toyota.

Toyota also employs the hierarchical system of education and training. It is intended, in addition to educate beginners, to train the in-company SQC special staff and advisors who can act as SQC promotion leaders of workshops of 200 departments and also to be engaged in the SQC seminars as trainers.

Now the Toyota education system is planned and implemented in six ranks: Beginner (100%), business (100%), intermediate (60%), lower advanced (15%), upper advanced (5%) SQC classes and SQC special advisor class (2%). The ratios of participants to the total of twelve thousand employees are given in the parentheses so that 100% of employees are, for example, expected to attend the beginners and business classes.

The beginners and business classes are designed to cover the daily works while at the two middle class courses participants will learn and practice the new SQC methods. The two highest classes are aimed at training the trainers and leaders of respective workshops and for extra professional purposes advanced lectures are also given. Qualifications for SQC special staff and SQC special advisors are determined and the respective titles are given to successful candidates. According to Amasaka and Osaki (1999) eight hundred special staff and advisors who have successfully completed the six steps are now actively engaged in their respective works.

Three courses for the beginner personnel are prepared in more detail: technician, sales and clerical courses. Typical curricula of the technician course (3 days, 21 hrs.), which are composed of twelve lectures, are given in Amasaka and Osaki. It should be noted that in the second lecture they learn how to integrate various statistical methods to solve real problems using the Toyota Technical Methods. The TPOS is fully utilised throughout the twelve lectures so that each trainee can take the TPOS back to his or her own workshop for practical use.

5. TQM SYMPOSIUMS AND CONFERENCES

JUSE and JSA have been promoting many conferences and symposiums on various topics and at a variety of levels. It is important to attend those meetings to present their own activities and to learn of achievement by others. An annual Conference on Science SQC is being held within the Toyota Group inviting top management and external professionals to attend and it is a very good incentive for employees to present their achievement to the heads of the company.

Of course the Annual Conference and Symposium of the Japanese Association of Quality Control are also giving a very good opportunity for researchers of TQM to present their achievements as well as to learn from others.

6. CONCLUDING REMARKS

As stressed in the text the most important thing for training researchers in the company is that trainees themselves have their own motivations. Then it is essential to teach statistical methods based on the real problems they are confronted with. When they have their own motivations and related data, it is very easy to teach them statistical ideas. It does not depend on the particular field where they are working. It inevitably suggests to them not to work alone when analysing their data, but to be aware of the phase of R & D activities he or she is, and to include manufacturing, marketing and after market research.

I also suggested that statistics training is most efficiently done by in-company trainers with some appropriate software and database of the company's past achievements. Then an in-company hierarchical education system of special SQC advisors and staff is essential for discovering skills and also for maintaining the system itself. However, if the in-company education system is not matured enough the courses outside the company may also be efficiently utilised.

In Japan, JUSE, JSA and other Institutions are providing a sufficient variety of courses, philosophical as well as technical, for TQM training. Researchers can also consult with the experts in the universities. Those experts have usually some connection with JUSE or JSA and they can introduce appropriate tutors for the companies. It should be noted that even the most prominent companies such as Toyota Motor Corporation and Takenaka Corporation are utilising the courses of JUSE introduced in § 3.

REFERENCES

Amasaka, K., & Osaki, S. (1999). The promotion of the new statistical quality control internal education at Toyota Motor: A proposal of 'science statistical quality control' for improving the principle of total quality management. *European Journal of Engineering Education*, 24, 259-276.

Amasaka, K., Kosugi, K, & Maki, K. (1999). A proposal of the new SQC internal education for management (in Japanese). *Quality, Journal of the Japanese Society for Quality Control*, 29, 292-299.

Hirotsu, C. (1992). QC technology (in Japanese). *Quality, Journal of the Japanese Society for Quality Control*, 22, 238-258.

Ishikawa, K. (1969). Education and training of quality control in Japanese industry. *Reports of Statistical Application and Research, JUSE* 16, 85-104.

Jido, J. (1990-1991). TQC in Takenaka Corporation. *Reports of Statistical Application and Research, JUSE* 37, 29-44.

Mizuno, S., & Kume, M. (1978). Developments of education and training in quality control. *Reports of Statistical Application and Research, JUSE* 25, 78-102.

Okuno, T. (Ed.) (1990-1991). Special issue: The Deming Prize. *Reports of Statistical Application and Research, JUSE* 37.

Chihiro Hirotsu
Faculty of Science & Technology, Meisei University
2-1-1 Hodokubo, Hino-City, Tokyo 191-8506, Japan
E-mail: hirotsu@ge.meisei-u.ac.jp

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 65-68.

Printed in Granada, Spain

S. P. MUKHERJEE

DISCUSSION

The growth of statistics (as a scientific method) has been stimulated by its applications in a wide array of scientific investigations. Such applications present an interesting panorama, ranging from cases where an imaginative application of appropriate statistical tools has considerably enriched the substantive content of a study to situations where statistics has been misused or even abused.

To enable researchers in various disciplines to derive due benefits from proper use of statistical methods and techniques, a strong collaborative effort by statisticians and users of statistics has to be undertaken. Training researchers in the use of statistics has assumed great significance not merely to avoid abuses and misuses of statistics, but – more importantly- to enable researchers to make efficient use of appropriate statistical methods and techniques.

Training is different from teaching; we teach a subject, we train an individual (or a group). Researchers to be trained are distinct from regular students in classroom teaching. Training of researchers should equip them to handle their research problems in general and not just one particular problem or project. To tackle a particular problem or project, a researcher may need to involve a statistician as a consultant. It must be admitted that training researchers in the use of statistics is a more onerous task than teaching statistics or even doing research in statistics by oneself (or jointly with some other statistician(s)).

The three papers slated for this session focus attention on several issues involved in training researchers in three important areas viz. medical research, total quality management and behavioural science. Of course, the paper by Estepa and Sánchez-Cobo is not restricted to a particular area of application and deals with the understanding of association in general. This paper as well as the one by Svensson contains findings of some investigations to support certain points raised by the authors.

Svensson's paper highlights the need to train both statisticians and applied researchers in order to produce good quality research, where right applications of the right statistical tools add value to the substantive content of the research output. In this context, the statistician consulted or otherwise involved should not only possess adequate knowledge of relevant statistical methods, but should also be able to properly interact with the researcher, have interest in problem-solving and be able to communicate abstract ideas in understandable language.

Svensson lays great stress, -and quite rightly so-, on proper planning of an experiment and on a proper comprehension of the measurement process and hence on the peculiarities, if any, of the data generated to derive inferences. In fact, one should carefully examine the data including the inherent uncertainty in measurements to decide on the appropriate methods of statistical analysis.

In this connection, one is reminded of what a great German philosopher had once said about the role of measurements in science and technology:

- If I can define it, I can measure it;

- If I can measure it, I can analyse it;
- If I can analyse it, I can control it;
- If I can control it, I can improve it.

Rating scales and questionnaires are quite often used in behavioural and medical research. Responses to an item in such a questionnaire are converted to scores using some procedure like Likert's scaling, assuming a normal distribution for the underlying trait. In some situations, like in attitudinal surveys, this assumption implies that most of the respondents are neutral or indifferent, which in turn would imply that the basic purpose of the opinion or attitude survey is foiled. The suggestion to use a distribution generated in terms of taking out a normal distribution from a rectangular one for this purpose has not found many takers.

Svensson identifies several factors which account for the somewhat unimaginative, computer-driven routine application of hackneyed and relatively simple statistical tools most often encountered in practice, sometimes reflected in published papers. Situations involving ordered categorical data, or variously censored data, or meta-data culled from various sources, etc. have to be, first of all, identified to have these characteristics and the need for dealing with them in their proper perspectives has to be appreciated. Even when statistical packages are available for some of these situations, a properly trained statistician should be involved right from the beginning i.e. the stage when objectives of an experiment or investigation are being developed.

Estepa and Sánchez- Cobo present the interesting findings of an empirical research on the understanding of association and its implications for the training of researchers. As is well-known, confusion of association or correlation with causation, of zero (linear) correlation with independence, lack of clear comprehension about positive and negative as well as linear and non-linear as also perfect (mathematical) and average relations, and similar other problems have plagued results and theories or explanations of different phenomena in several branches of human knowledge. Estepa and Sánchez-Cobo have dwelt on the motivation to steer clear of such confusions, misconceptions and consequent misuses. It is pretty difficult to provide adequate direction or guidance to resource persons involved in training researchers in this connection and the authors have not come up with any. Some possible strategies could be:

1. Stress on appropriate real-life situations to motivate participants, like unilateral dependence of crop-yield on rainfall as against bilateral interdependence between scores in mathematics and in physics in a school examination, say;
2. Introduction of topics systematically in a given order or sequence and not all at a time, like correlation first and regression, next;
3. Stress on rigour as a characteristic of any approach to a problem in science;
4. Indications about situations in which certain techniques should be used and their possible limitations, including the underlying assumptions and the extent to which these are critical;
5. Hints about possible enrichment of statistical methods through proper recognition of different special features of an experiment and the resulting data.

Or, in other words,

1. Start with a data-set and a knowledge of the background experiment /investigation;

2. Raise relevant questions to be answered through data-analysis;
3. Use known statistical methods to derive answers;
4. Note assumptions made and, if possible, examine if they are valid;
5. Seek other methods, with the help of a competent statistician, if needed;
6. Generalise the ambit of investigations and data-sets where these methods can apply, with care and caution;
7. Provide inputs to development of new statistical methods to deal with other problems where existing ones may not yield good answers to questions likely to arise.

Chihiro Hirotsu delineates a hierarchical training system in statistical methods for implementation of TQC or TQM. Obviously, training for this purpose has to be dove-tailed into the over-all company-wide effort towards TQM, and most of it should be practice-oriented and meant for people who are self-motivated to pick up statistical methods for enhancing their problem-solving abilities. Hirotsu refers to five courses in the context of TQM viz. elementary statistics, design of experiments, multivariate analysis, advanced (beyond ANOVA) methods and applications. He also provides a whole profile of courses meant for different categories of people in different methodologies and bearing on different themes. He goes on to provide illustrations of his points from several reputed Japanese organisations.

Hirotsu includes ‘reliability analysis’ within the course on DOE. Firstly, he should spell out whether only design reliability considerations are meant to be covered or even other aspects of reliability analysis. Within the latter also, there are varied topics like stress-strength analysis, fault tree analysis, failure mode and effect analysis, redundancy allocation, structural and reliability importance of components, etc. One has to restrict oneself to those which are of direct relevance to given situations. Probably, an omnibus course content will be too heavy and is unwarranted. There are many multivariate problems in reliability analysis, which would go along better with multivariate analysis.

Courses mentioned as D, E and F are really useful , but to conduct such structured and tightly-timed courses the faculty has to do a lot of homework for making a prepared, focused, rigorous yet easily comprehensible presentation.. In the list of courses there is no mention of calibration procedures and estimation of uncertainty in measurement, areas which are of great relevance to TQM and which derive strength from statistical methods. Response surface methodology has not been emphasised. No reference has been made about Bayesian methods, even of elementary ideas to take care of prior information, which almost always exists in some form, or another. Optimal process adjustments, associated optimisation methods, process and machine capability analysis, Trend analysis, and several other recent topics do not find a place in the list of courses. A bit of mathematical modelling of customer satisfaction on the lines of ACSI or ECSI may also be added to the list.

All this would mean that a TQM man has to know a lot of statistics -not all simple or unsophisticated- irrespective of his background or his responsibility profile. This does not sound as feasible or desirable, particularly in view of the fact that ideas like those of interaction between factors, interpretations of entities like principal components or factors (in factor analysis) and methods for choosing appropriate probability models are not easy tasks even for well-equipped trainees.

While one can agree with Hirotsu that the courses on advanced methods and applications should be taught within the company, this enjoins on the company to hire or retain very competent statisticians. What is badly needed is a mode of training that draws heavily on actual cases, allows sufficient scope to trainees to absorb new ideas

and methods, does not leave everything to appear on the computer screen for an understanding of the results but takes full advantage of computers for making useful interactive presentations, and is guided by someone with imagination and integrity.

REFERENCES

Estepa, A., & Cobo, F. T. (2001). Empirical research on the understanding of association and implications on the training of researchers. In C. Batanero (Ed.), *Training researchers in the use of statistics*(pp. 35-51). Granada: International Association for Statistical Education and International Statistical Institute.

Hirotsu, C. (2001). Statistical training of researchers for the total quality management. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 53-63). Granada: International Association for Statistical Education and International Statistical Institute and International Statistical Institute.

Svensson, E. (2001). Important considerations for the optimal communication between the statistician and medical researchers in consulting, teaching and collaborative research with a focus on analysis of ordered categorical data. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 23-35). Granada: International Association for Statistical Education and International Statistical Institute.

S. P. Mukherjee
University of Calcutta, Department of Statistics,
35 Ballugunge Circular Road,
Calcutta 700019, India.
E-mail: csri@satyam.net.in

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 69-72.

Printed in Granada, Spain

MARTHA B. ALIAGA

DISCUSSION

I would like to express my appreciation to the Round Table Committee for inviting me to participate in this discussion. I will discuss below each of the papers.

1. EMPIRICAL RESEARCH ON THE UNDERSTANDING OF ASSOCIATION AND IMPLICATIONS FOR THE TRAINING OF RESEARCHERS

Antonio Estepa and Francisco Sánchez Cobo present a coherent set of findings about the way students perceive association. These results came from an assessment study on how university students understand correlation and regression.

The paper touches upon a number of issues, which I found to be very important in the training of researchers. I am very impressed with their findings and found little to take issue with. My comments will therefore focus on questioning how to remedy the misconceptions on the part of the students.

The learning of statistics must move from passive to active. We face many challenges when teaching an algebra-based introductory statistics course. Among the numerous obstacles is the perception of many students that statistics is boring and pointless. One of our principal goals as teachers is to engage students in the subject and to teach them that statistics is full of ideas and methods that will make them more informed users of the information they encounter every day.

To reduce "statistical anxiety" and nurture an appreciation that statistical thinking and methods can be helpful in solving real problems, we can bring real problems to solve in the classroom. Could it be possible that we use more formulas than needed? With the availability of an ever-increasing technology, like the graphing calculator, we can move to the discovery approach.

With more emphasis on teaching "statistical thinking" students can unlearn the "law of small numbers" and understand that variation exists in all processes.

I always found it surprising how difficult it is for students to distinguish the dependent from the independent variable. Is IQ the independent and income the dependent variable? Could it be possible to consider income the independent and IQ the dependent? Income may permit private tutors and IQ could be improved.

Finding a relationship between variables is an important step toward modelling phenomenon. However, how do relationships arise? Both variables may change with a third variable. One variable may be the cause of the other. The relationship may be the result of chance variation alone. The relationship may be the result of aggregation of several sub-populations. The relationship may be the result of the operational definition of the variables.

I found it very helpful to teach with counterexamples to accelerate the learning process. In the case of association, using the graphing calculator in class permits the students to enter, manipulate, and plot data quickly and conveniently. The graphing calculator minimises hand calculations and eases data plotting situation. It also helps

one verify the derived hand calculations, making a formidable formula much more manageable.

2. IMPORTANT CONSIDERATIONS FOR OPTIMAL COMMUNICATION BETWEEN STATISTICIANS AND MEDICAL RESEARCHERS IN CONSULTING, TEACHING AND COLLABORATIVE RESEARCH- WITH A FOCUS ON THE ANALYSIS OF ORDERED CATEGORICAL DATA

Elizabeth Svensson's paper primarily reflects on the lack of understanding about the relationship between the measurement properties of data and the choice of statistical method of analysis among researchers and statisticians. Beverige, quoting Schiller, suggests that:

> "The slowness and difficulty with which the human race makes discoveries and its blindness to the most obvious facts, if it happens to be unprepared or unwilling to see them, should suffice to show there is something gravely wrong about the logistician's account of discovery" (Beverige, 1961, p. 112).

Tukey (1962) asked himself how can new data analysis be initiated. And his answer was that we should seek out wholly new questions to be answered. Statisticians might assume that they play a natural role in teaching the statistical methods, but researchers from other disciplines have the technologies at hand to collect and store data and have assumed a role in the analysis. Much of the development of new methods is going on outside the traditional statistical literature, and it is no longer sufficient to scan statistical journals to be aware of what is going on.

The easy availability of manuscripts and software on the World Wide Web has already had a major effect. What are some of the risks? The software may not have been tested. An article electronically published may not have been peer-reviewed, and it might contain errors, etc, etc.

I am a strong advocate of changing the basic curriculum in the first courses offered in statistics. Technological change, and both innovation and improvement is often so rapid that if we are trained to run the machinery of today, the value of what we learn will be quite limited. The life cycle of technology is short, but the principles, which permit learning, are long-lived. The focus of teaching has to be on describing the principles of efficient learning. The learning process, however, is not a common sense activity. We need to teach students how questions should be asked as well as the scientific method. We need to teach students how to think within statistics. I propose a following list of topics as we teach them in our Stat. 170

Anne Hawkins, director of the Centre for Statistical Education, at the University of Nottingham, UK once suggested that as statisticians, we have important specialist skills, knowledge and understanding, not all of which may be shared by everyone, and that is what makes us specialists.

However, it is in our interests, as well as being our professional duty, to learn how to communicate the general principles of what we have to offer to as wide an audience as possible. This requires not only a continuing commitment to the study of statistical teaching and learning processes, but also the willingness to implement indicated changes into classroom practice. Specialists in the field of statistical education are beginning to make progress in a variety of ways, but the task that faces them is far from trivial. The problem of how to educate the educators is almost as pressing as the need to produce and research innovations for them to adopt. This is something for which all

members of the statistical profession have the responsibility, not just those who are identified as specialists in statistical education. A professor of statistics who cannot be persuaded that participation in a conference on statistical education has relevance to him or her has indeed failed to grasp the basics!

3. STATISTICAL TRAINING OF RESEARCHERS IN TOTAL QUALITY MANAGEMENT

I enjoyed reading Hirotsu 's article, which provides an overview of a training system for statistical methods in Total Quality Control in Japan. I will not argue that one finds motivated students in the in-company educational system. Instead, my comments will be about the "content" of those courses offered in the educational system.

When teaching at universities, many instructors may not give careful thought to course objectives nor consider their major goal to be that of exposing students to statistical metaphors. This might be because a strict curriculum, perhaps built around a specific traditional text, has been pre-ordained by the department. Unfortunately, for the vast majority of students, the experience of taking a statistics class is a negative one.

In an effort to overcome such problems, I am wondering whether industry should play a more aggressive role in offering assistance to local universities in order to help add a more practical flair to the introductory classes.

The University of Michigan attempts to be a leader in Interdisciplinary teaching approach and offers research opportunities for undergraduates. We offer a class, Stat. 170, where we tie the statistical ideas into an overall approach of scientific inquiry. Much of the teaching in quality is based on system thinking. Maximising a function of several variables, for example, cannot be accomplished variable by variable, so the best team is NOT necessarily the one with the best players. An organisation with excellence in each department may produce an inferior product or service. Some points to consider:

- Descartes admonition: "Sense perception can be sense deception";
- Examples that should cause us to think carefully about comparisons;
- Probability: Interpretation;
- When are surprises not uncommon?
- Paradoxes;
- Concepts of conditioning;
- Relationships;
- Cause and effect versus alternatives;
- The regression effect;
- Systemic biases;
- Size and other selection biases;
- Observer and experimenter bias.

I advocate more explicit discussion of goals, the degree of change I think will require Content, Pedagogy, and Technology.

REFERENCES

Beveridge, W. I. B. (1961). *The art of scientific investigation*. New York: Vintage Books.

Estepa, A., & Sánchez Cobo, F. T. (2001). Empirical research on the understanding of association and implications on the training of researchers. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 37-51). Granada: International Association for Statistical Education and International Statistical Institute.

Hirotsu, C. (2001). Statistical training of researchers for the Total Quality Management. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 53-63). Granada: International Association for Statistical Education and International Statistical Institute.

Tukey (1962). The future of data analysis. *Annals of Mathematical Statistics*, 33, 1-67.

Svensson, E. (2001). Important considerations for the optimal communication between the statistician and medical researchers in consulting, teaching and collaborative research with focus on analysis of ordered categorical data. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 23-35). Granada: International Association for Statistical Education and International Statistical Institute.

Martha Bilotti-Aliaga
The University of Michigan,
Department of Statistics,
4077 Frieze Building, Ann Arbor Michigan, 48109, USA.
E-mail: aliaga@umich.edu

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 73-85.

Printed in Granada, Spain.

ANN-LEE WANG

HOW MUCH CAN BE TAUGHT ABOUT STOCHASTIC PROCESSES AND TO WHOM?

Researchers quite often need to model and analyse real-world random phenomena using stochastic processes. Learning stochastic processes requires a good knowledge of the probability theory, calculus, matrix algebra and a general level of mathematical maturity. However, not all researchers have a good foundation in probability and mathematics. In this paper, we discuss the different approaches to the teaching of a first course in stochastic processes to researchers. Difficulties in the understanding of stochastic processes and the various mathematical techniques used in stochastic processes are discussed. Proposal for the core topics of such a course and ways of teaching them are put forward.

1. INTRODUCTION

The abundance of books on stochastic processes shows the perceived usefulness and applicability of stochastic processes in modelling random phenomena. These books range from the more elementary books such as Bhat (1984) and Taylor and Karlin (1994), intermediate books such as Cox and Miller (1965), Karlin and Taylor (1975) to more specialised books such as Gardiner (1997) and Tijms (1994).

Researchers in science, engineering, computing, business studies and economics quite often need to model real-world situations using stochastic models in order to understand, analyse, and make inferences about real-world random phenomena. Finding a model usually begins with fitting some existing simple stochastic process to the observed data to see if this process is an adequate approximation to the real-world situation. If these simple models are found to be inadequate, ways of extending these models may then be developed.

Learning stochastic processes requires a good knowledge of the probability theory, advanced calculus, matrix algebra and a general level of mathematical maturity. Nowadays, however, less probability theory, calculus, matrix algebra and differential equations are taught in the undergraduate courses. This makes it a little bit difficult to teach stochastic processes to researchers.

Faced with a group of researchers, one has to decide on the amount of mathematical techniques that the students are able to understand in a reasonable time span. The researchers' prior mathematical knowledge puts a limit on the amount and the level of stochastic processes that one can teach. It is necessary to think about what and how to teach a first course in stochastic processes to researchers.

In this paper, we shall first consider the different types of stochastic models and some reasons why researchers use stochastic models. A short account of the different approaches to the teaching of stochastic processes is then given, and the core topics for a first course is put forward. The obstacles and the difficulties inherent in the

understanding of stochastic processes are discussed, and some suggestions for the teaching of these topics are proposed. Some comments about the misuse of stochastic models are stated.

2. STOCHASTIC MODELS

A first approximation for a model is usually a deterministic one. Sometimes adding some variation or some stochastic elements to a deterministic model may give a better representation of the random phenomenon under study, so that there is a need to use stochastic models. Stochastic models are constructed with certain assumptions in mind. If the model proved to be badly inadequate, the assumptions are most probably wrong. In case the model seems to fit the data collected, it may provide a better understanding of the random phenomenon.

Stochastic models may be broadly classified into two groups: the purely empirical models which would only give a good representation of the data and models that attempt to incorporate some links between the estimated parameters and the underlying physical process. Examples of a purely empirical model are the models of daily rainfall observations of Stern and Coe (1984), who used non-stationary Markov chains to fit the occurrence of rain and gamma distributions with varying parameters to fit the rainfall amounts. Examples of the second group are the stochastic models for rainfall considered by Rodriguez-Iturbe, Cox and Isham (1987) where the parameters of the models have a physical interpretation. A more detail grouping of stochastic models into four different types is given by Cox (1997).

Stochastic models are used in several fields of research. Some models used in the engineering sciences are models of traffic flow, queueing models, reliability models, spatial and spatial-temporal models. In the computer sciences, the queueing theory is used in performance models to compare the performance of different computer systems. Applications of random walk models are to be found in electric networks (see Doyle and Snell, 1984) and in physics, chemistry and biology (see Weiss, 1983 for a non-technical description). Various diffusion processes models are used as stochastic financial models. Stochastic models are also used to construct epidemic models. More examples of the use of stochastic processes can be found in Gardiner's book (1997). For a short and succinct introduction to stochastic modelling, please refer to Isham (1991). The mathematical techniques and the numerical computation used in stochastic models are not very simple.

In an introductory course, the hope is to teach researchers a small number of stochastic models effectively to enable them to start thinking about the applications of stochastic processes in their area of research. A researcher could use one of these small numbers of stochastic models as a starting point for modelling the random phenomenon that he/she is investigating. This simple model might be adequate for the researcher's interest, and if it is not, later modification could be made.

These small numbers of stochastic models are the core topics to be taught in an introductory course on stochastic processes directed to researchers in the physical sciences, engineering, operational research and computing science. These researchers have a stronger background in mathematics and probability than researchers in the biological sciences. There also seems to be a low usage of stochastic processes in recent biological and health sciences publications as mentioned by Harraway et al. (2001).

Table 3 of their paper indicates that only about 1.71% of the papers surveyed used some stochastic processes, and the topics from stochastic processes that appear in these

papers are transition matrices, random walk, Markov processes, time series, periodogram, cross correlation and random intervention analysis. Spatial analysis merits a classification by itself and about 4.47% of the papers surveyed used this method of analysis. Notwithstanding this evidence, some very highly complex stochastic models could be used in the biological sciences. Such models are outside the scope of this paper and will not be discussed here.

I consider the core topics to be discrete time Markov chains with stationary transition probabilities, the Poisson process and birth and death processes. The type of researchers attending the course will determine what other specialised topics such as renewal theory and queueing theory need to be taught.

The needs of different research areas are different, since some stochastic models are only used in a particular research area. It is not reasonable to require all researchers to learn models that they will not need in their research. In some area, such as physics, different notations and nomenclature are used. These have to be explained to the researchers and perhaps some other adjustments to the lecture notes need to be carried out for certain applications of stochastic processes. Researchers, while learning the three core topics, would have picked up a fair amount of probability reasoning and mathematical techniques that are very useful in the learning and understanding of more specialised topics.

Some textbook writers have different views on what should be taught first. For example, Tijms (1984) and Kao (1997) start off with renewal processes, the Poisson process, proceed to Markov chains and then other more specialised topics. However, understanding renewal theory may require more mathematical skills than the average researcher has. This approach may be suitable for the more mathematical minded researchers. In general, it is better to begin with Markov chains.

Even if one chooses to start off with Markov chains, there are two ways in which the topic is laid out in textbooks. Books such as Cox and Miller (1965) and Bhattacharya and Waymie (1990) begin with random walk. Other books such as Karlin and Taylor (1975), Taylor and Karlin (1994) and Kulkarni (1995) begin with Markov chains and random walks are introduced as examples of Markov chains.

Starting a stochastic process course with random walk has its advantages. Feller (1957) remarked in Chapter III that “exceedingly simple methods may lead to far reaching and important results” in random walk. It is possible to discuss many results that are sometimes counter-intuitive using simple methods. Unfortunately, such elementary methods may not be so easily understood by some students. My experience is that not all students are comfortable with combinatorial methods, and I think that it is better to teach random walks as examples of Markov chains.

3. OBSTACLES IN THE LEARNING OF STOCHASTIC PROCESSES

There appears to be a lack of papers on the teaching of stochastic processes. Writers of textbooks sometimes will suggest what selection of the chapters in their books can be used to make up a course in stochastic processes in the prefaces to their books. A few writers do comment on how students/researchers should learn stochastic processes. Some comments are:

> *"Although the basic concepts of useful stochastic models are at the core simple and intuitive, in fact many students find it difficult to translate a specific applied probability problem into an appropriate stochastic model. The student can only acquire the skills of modelling new situations by a considerable amount of practice in solving problems on his*

own " (Tijms, 1986, p. XI).

"The best way to learn the material in this book is by solving problems given in exercises " (Kulkarni, 1995, p. X).

Therefore, the advice is practice, practice and more practice. The only paper commenting on the teaching of stochastic processes that I have come across is the one by Cox and Davison (1994). In it, they remarked that:

"A genuine understanding of the key concepts and principles of the probability theory is essential for any work in stochastic processes. These concepts are simple, but treacherously so. There is probably no branch of mathematics in which it is so easy to advance plausible sounding arguments that are in fact wrong! " (Cox & Davison, 1994, p. 25).

Not many researchers have "a genuine understanding of the key concepts and principles of probability". All of them have some knowledge of the probability theory, and there is not enough time to teach both probability and stochastic processes in an introductory stochastic processes course to researchers. What can be done is to selectively teach the most relevant topics and their applications in selected stochastic processes. It is necessary to teach with as little mathematics as possible. Each new concept has to be taught in small doses with graphical displays and diagrams where possible. Successive new concepts are built on previously taught concepts. Teaching using such small building blocks progressively may make it easier for the researchers to understand and to learn the required concepts. Researchers need to be reminded every now and then that they have to put in a fair amount of effort into learning too. Some reinforcement in the form of requiring researchers to attempt tutorial problems frequently has to be carried out. In this section, some obstacles in the learning of stochastic processes and some ways of overcoming them will be discussed.

3.1. INTUITION

Researchers' intuitions get in the way of their understanding of stochastic processes. Since some results in stochastic processes are counter-intuitive, these would add to the frustration of researchers in their attempts to understand stochastic processes.

A demonstration, via simulation, of a Bernoulli process at the beginning of the course might serve as a warning to researchers about intuition. The simulation of ten thousand tosses of an unbiased coin can be used as an example. The result of the simulation can be used in various ways. The proportion of heads is most probably very close to 0.5. Moore (1990) on pages 97-98 recounted that G. L. Buffon, Karl Pearson and John Kerrich had actually tossed coins for respectively 4040, 24,000 and 10,000 times. The observed proportions of heads were respectively 0.5069, 0.5005 and 0.5067.

Given these facts, if one then displays the sequence of heads and tails, the researchers may or may not be surprised to see the runs of heads and tails. Another graphical display showing the lead of head over tail (represented on the vertical axis) as the number of tosses increases (represented on the horizontal axis) will perhaps lead the researchers to think about counter-intuitive happenings in stochastic processes. Most researchers would expect to see the points oscillating around the horizontal axis. What they actually see is a graph where for a large proportion of the time, the points are above (or below) the horizontal axis. This seems to run counter to their intuitive feelings of how a random process should behave. The graphical display might just keep them in the straight and narrow path of thinking in terms of probability and stochastic processes.

Readers who may want to read more about intuition and probability and ways to resolve the conflicts and to build on it for undergraduates and school children may read Borovcnik and Peard (1996) and the extensive (ninety-one references) bibliography therein.

3.2. TRANSLATING A WORD PROBLEM INTO PROBABILITY STATEMENTS

Anyone who has any experience in teaching mathematics would realise that students have difficulties in translating a word problem into mathematical statements. The same thing holds true in stochastic processes as has been remarked by Tijms (1986). One way to overcome this is to use examples that are as close to the subject matter of the researchers' interest as possible while teaching. These examples have to be preceded by simple unambiguous examples that will help researchers to understand the various concepts in stochastic processes.

As an illustration of translating a word problem into probability statements in stochastic processes, consider the case of a Markov chain. The definition of the various states of a Markov chain in an example has to be taught carefully and thoroughly. States have to be defined such that they conform to the Markov chain assumptions. Emphasis has to be placed on what the answers that are being sought are. For the same random phenomenon, different state spaces may be defined for different objectives. We can use the coin tossing example to illustrate this.

In the coin tossing example, we can consider at least three different stochastic processes as shown below:

1. The lead of head (H) over tail (T): in other words a simple random walk. The states may be defined as $X_0=0$ and for $n \geq 1, X_n = Z_1 + ... + Z_n$ where the Z_i are independently and identically distributed random variables with equal probability of taking on the values of 1 or –1 depending on whether H or T occurs at the n-th toss.
2. Success run: Suppose we consider the occurrence of H as a success. The states of a Markov chain, W_n, may be defined as the length of a success run at the n-th toss of a coin.
3. Result of two consecutive tosses of a coin: Define a Markov chain where the states X_n, for $n > 1$ are defined according to the (n-1)-th and the n-th tosses.

These examples may be followed by a few simple examples of coloured balls in boxes, i.e. examples involving the withdrawing, replacing and inter-changing of the balls in various manners. An illustration of this type of example is the following:

> *Example 1* (Karlin & Taylor, 1975). N black balls and N white balls are placed in two urns so that each urn contains N balls. At each step, one ball is selected at random from each urn and the two balls interchange. The state of the system is the number of white balls in the first urn. Determine the transition probability matrix of the Markov chain.

These coin tossing and coloured balls in boxes examples may serve as an introduction to the state space and the time parameter space of a Markov chain. These examples are usually devoid of ambiguity and are on the whole non-emotional, non-political and non-cultural. Researchers would be able to understand the various concepts in stochastic processes with less interference. Examples from the subject matter of the researchers are then used to show how the concept of Markov chain can be used in the area of interest to the researchers.

In contrast to state space, researchers usually do not have much difficulty in defining the time parameter space of a Markov chain.

3.3. TRANSITION PROBABILITY

Most undergraduates find great difficulties in determining the transition probability matrix, ***P,*** of a Markov chain. This is in part due to the fact that they have great difficulties in defining the state space and the time parameter space of a Markov chain. This usually arises because of students' failure to fully understand the nature of the random process of a given problem. A Markov chain is a random process that evolves with time. It moves from state to state in accordance with the transition probabilities of the Markov chain. A future state is dependent only on the present state and not on the past history of the Markov chain. Given the state of the process at time n, X_n, students have to be able to write down all the possible states of the process at time (n+1), X_{n+1}. They would then be in a position to evaluate the transition probability $P[X_{n+1}|X_n]$. Without a full understanding of the random process under study, students may not be able to write down all the possible mutually exclusive and exhaustive events for X_{n+1} and so be unable to find the transition probability.

Falk (1986) discussed three issues concerning the learning and understanding of conditional probability. Let us consider these issues in the context of the transition probability of a Markov chain.

Consider two events A and B. The first issue raised by Falk (1986) is that if students perceived that it is reasonable to ask about $P[A|B]$, some of them would hesitate at evaluating $P[B|A]$ as this seems "unnatural" to them. In a Markov chain, however, other than the transition probability $P[X_{n+1}|X_n]$, it is often quite meaningful to consider $P[X_n|X_{n+1}]$.

The second issue raised by Falk (1986) is that in evaluating $P[A|B]$, has the sample space changed with the extra information provided by the event B? For transition probability, a tree diagram or similar diagrams showing the possible progression with time of the Markov chain is helpful in concentrating the minds of researchers. Hopefully they will be able to evaluate $P[X_{n+1}|X_n]$.

The third issue raised by Falk (1986) may be termed as translating a word problem into probability statements. Students have to decide whether the word problem is in the form of $P[A|B]$ or $P[B|A]$. In a Markov chain, this issue may not arise if the students have understood the random process fully. Researchers may also be reminded to concentrate on $P[X_{n+1}|X_n]$.

In a course for researchers, demonstrations of how to obtain transition probabilities can be carried out initially with some coin tossing examples and simple examples of coloured balls in boxes. A review of conditional probability will be helpful. These examples are then to be followed by examples from the researchers' subject area. These examples have to be carefully explained in detail to show how probability and conditional probability are used in practice. Researchers with a better grasp of probability theory can be required to work out the transition probabilities themselves.

3.4. MATHEMATICAL TECHNIQUES

In order to understand and use stochastic processes, researchers need to know a fair amount of mathematical techniques. The mathematical techniques used range from the sum of geometric series, difference equations, recursive relations, generating functions, convergence, matrix operation, and conditional expectation to differential equation.

These techniques need to be either reviewed or taught at length in context as the needs arise. Textbooks usually compiled the mathematical techniques in an appendix or as a group within each chapter. This is good for a smooth uninterrupted flow to the discussion of the various topics. In contrast, to be effective, the mathematical techniques have to be developed in front of the researchers as the teaching of each topic progresses. The effectiveness will be enhanced when the examples used to illustrate the techniques are in the subject matter of the researchers.

Among the various techniques, the technique of first-step (or last-step) analysis is very useful in stochastic processes. In this method the probability of an event is obtained by considering what would happen after the first step has been taken (or the last step has been arrived at). The theorem of total probability is then applied to obtain the desired probability. As an example, consider the derivation of the Chapman-Kolmogorov's equation for $P_{ij}^{(n+1)}$, i.e., the probability of arriving at state *j* in the *(n+1)*th step given that the process was initially at state *i*. Let the state at time *n* be X_n.

$$P_{ij}^{(n+1)} = P[X_{n+1} = j | X_0 = i] = \sum_{k \varepsilon S} P[X_{n+1} = j, X_1 = k | X_0 = i]$$
$$= \sum_{k \varepsilon S} P[X_{n+1} = j | X_0 = i, X_1 = k] P[X_1 = k | X_0 = i] = \sum_{k \varepsilon S} P_{ik} P_{kj}^{(n)}$$

In the above, in the second equation we consider all possible transitions that can occur in the first step and then we use the law of total probability. In the third equation, we use the formula for $P[A \cap B | C]$ for any three events *A, B* and *C.* In the fourth equation, we use the definition of transition probability and apply the Markov property to obtain the Chapman-Kolmogorov's equation for $P_{ij}^{(n+1)}$.

Using this first step analysis technique, the probability of ruin in the gambler's ruin problem can be shown to satisfy a set of difference equations. The use of the first-step analysis technique is not as straightforward here as in the case of $P_{ij}^{(n+1)}$.

Consider a particle moving along the line interval [*0, c*]. It moves one step to the right (one positive unit) with probability p and one step to the left with probability q where $p > 0, q > 0$ and $p + q = 1$. Suppose the particle is initially at point a or $X_0 = a$. Let u_i be the probability that the particle reaches *0* before it reaches *c*, with $a \geq 1$ and $c > a$. Let E_i denote the event that the particle reaches *0* before it reaches *c* when it is initially at *i*. Let F_i denote the event that the particle take one negative step when it is at *i*. Then using the first-step analysis and the theorem of total probability we have,

$$P[E_i] = P[E_i \cap F_i] + P[E_i \cap F_i^c]$$

This can then be written as

$$P[E_i] = P[F_i]P[E_i|F_i] + P[F_i^c]P[E_i|F_i^c] = qu_{i-1} + pu_{i+1}$$

Students usually have difficulties in ascertaining $P[E_i|F_i]$ and $P[E_i|F_i^c]$. Let us consider $P[E_i|F_i]$ first. An extra effort seems to be needed by the students to understand that $P[E_i|F_i] = P[E_{i-1}]$. When the particle reaches $(i-1)$, the movement of the particle will start afresh from a new starting point. In this case the new starting point or the new initial point is $(i-1)$. And so we have $P[E_i|F_i] = P[E_{i-1}]$. Similarly, one can obtain $P[E_i|F_i^c] = P[E_{i+1}]$.

The first-step analysis technique may be generalised and used to prove the Chapman-Kolmogorov equation, $P_{ij}^{(m+n)}$, and then to prove the first entrance decomposition theorem, to establish the mean first passage time and the mean time to absorption. As an illustration, we can prove the first entrance decomposition theorem as follows. Define $T_j = \min[n \geq 1 | X_n = j]$. Then

$$\begin{aligned} P_{ij}^{(n)} &= P[X_n = j | X_0 = i] = \sum_{v=1}^{n} P[X_n = j, T_j = v | X_0 = i] \\ &= \sum_{v=1}^{n} P[T_j = v | X_0 = i] P[X_n = j | X_0 = i, T_j = v] \\ &= \sum_{v=1}^{n} f_{ij}^{(v)} P[X_n = j | T_j = v] = \sum_{v=1}^{n} f_{ij}^{(v)} P_{jj}^{(n-v)} \end{aligned}$$

where $f_{ij}^{(v)} = P[T_j = v | X_0 = i]$, and is called the first passage time.

Before considering the first passage time, the time to absorption, limiting distribution and stationary distribution, evaluation of the powers of $\boldsymbol{P}$ and the classification of states have to be taught. Powers of $\boldsymbol{P}$ can be obtained by several methods. A review of these methods may be considered. Some $P_{ij}^{(n)}$ may be obtained by straightforward probability argument. For example, $P_{00}^{(n)}$ of an unrestricted simple random walk may be obtained in this manner. Tree diagrams are also very useful in the determination of certain $P_{ij}^{(n)}$.

The partitioning of the state space can be carried out by using a transition graph. Drawing a transition graph poses no difficulty to the students. The classification of the state space into equivalent classes and the identification of transient and recurrent states also do not seem too difficult. Of course it is still possible to set problems that will tie some students into knots while they try to draw a transition graph and partition the state space.

What appear to be difficult for the students are the understanding of first passage times and the evaluation of first passage probabilities. The stationary distribution of a finite Markov chain may be obtained by solving a set of equations. Researchers most probably know enough matrix calculus to be able to find the stationary distribution. Sometimes the set of equations can be readily solved as the solution involved only a geometric sum. Matrix algebra is also needed to obtain the probability of absorption and the mean time to absorption.

Computer packages for matrix operation may be used if they are available and

researchers are familiar with them or are willing to learn to use them. Some textbooks provide computer program codes (for example Kao, 1997, who uses MATLAB) or offer for sale software packages (for example Tijms, 1994) for analysing stochastic models. One software package for generating large Markov chain models and analysing the models is MARCA obtainable from North Carolina State University.

3.5. BIRTH AND DEATH PROCESSES

The Poisson process is a good and simple introduction to continuous time discrete state space Markov chains in terms of concepts and the mathematical techniques used. In deriving the probability that n events occur in the time interval (0, t), the researchers are introduced to the use of the differential equation in stochastic processes. The technique of generating function may also be introduced here to obtain the required probabilities. As exponential distribution features prominently in continuous time discrete state space Markov chain, a thorough review of its properties and the relationship with the Poisson distribution has to be carried out. Generalisations of the Poisson process can also be taught.

In a demonstration of the derivation of the relevant probabilities in the birth and death process, it may be better to start off with just one individual in the process. The relevant differential equation is then written out. This approach gives the researchers an intuitive feel for the mathematical techniques used. As most researchers are familiar with differential equations, it will not be too difficult for them to understand the derivation. A generating function approach is then used to solve the differential equation. A variation of the technique of first-step (last-step) analysis used here will yield the backward (forward) equations.

One would begin the discussion using a pure birth process with initially only one organism in the colony. A graphical display like the one in Cox and Miller (1965, p. 156) is a good visual starting point. This can then be followed by the demonstration of the derivation of $P_{1n}(t)$, i.e., the probability that there are $n \geq 1$ organisms present at time t, when initially one organism was present. The derivation of the relevant probabilities for a birth and death process will then be easier for the researchers to accept and understand. For example, the starting point of the derivation of the relevant probabilities of a birth and death process may look like this:

$$\begin{aligned} P_{1n}(t) &= [\lambda\delta t + o(\delta t)]P_{2n}(t-\delta t) \\ &+ [1-\lambda\delta t - \mu\delta t + o(\delta t)]P_{1n}(t-\delta t) + o(\delta t) \end{aligned}$$

which gives

$$\frac{dP_{1n}(t)}{dt} = -(\lambda+\mu)P_{1n}(t) + \lambda P_{2n}(t), \qquad n \geq 1.$$

Using generating function one might arrive at

$$\frac{\delta G}{\delta t} = (\lambda G - \mu)(G-1)$$

$$\text{where } G(\theta, t) = \sum_{n=0}^{\infty} \theta^n P_{1n}(t), \quad |\theta| < 1.$$

And the solution for $\mu > \lambda$ is

$$G(\theta, t) = \frac{\mu(1-\theta) - (\mu - \lambda\theta)\exp[(\mu - \lambda)t]}{\lambda(1-\theta) - (\mu - \lambda\theta)\exp[(\mu - \lambda)t]}$$

These equations look intimidating, but would not be quite so if one has understood the derivation of a Poisson process and a pure birth process. In the derivation of the Poisson process, it would have been shown that for $n \geq 1$,

$$P_n(t+h) = P_n(t)P_0(h) + P_{n-1}(t)P_1(h) + \sum_{i=2}^{\infty} P_{n-i}(t)P_i(h)$$

where $P_n(t) = P[X(t) = n]$. From the above equation, one would obtain the following:

$$\frac{dP_n(t)}{dt} = -\lambda P_n(t) + \lambda P_{n-1}(t) .$$

The solution is

$$P_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \quad n = 0,1,2,...$$

As for the pure birth process, one may start with

$$P_{1n}(t) = P[X(t) = n | X(0) = 1] = [1 - \lambda\delta t + o(\delta t)]P_{1n}(t - \delta t) + \lambda\delta t P_{2n}(t - \delta t) + o(\delta t) .$$

For $n \geq 2$, $\frac{dP_{1n}(t)}{dt} = -\lambda P_{1n}(t) + \lambda P_{2n}(t)$. Let $G_1(z,t) = \sum_{n=0}^{\infty} z^n P_{1n}(t)$, $|z| < 1$.

The differential equation can then be written as

$$\frac{\delta G_1}{\delta t}\left\{\frac{1}{G_1(G_1 - 1)}\right\} = \lambda$$

The solution is $G_1(z,t) = \frac{ze^{-\lambda t}}{1 - z(1 - e^{-\lambda t})}$, which gives $P_{1n}(t) = e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}$, n=1,2,3,…

Here the introduction of differential equations and their solutions are carried out in stages and in increasing complexity. This slow introduction will help to ease the way for the researchers to use these techniques in stochastic processes. It is more effective to start teaching a topic from the beginning.

Various other topics in birth and death processes may be explored depending on the interest of the researchers. A discussion of an embedded process in the birth and death process may be carried out. This embedded process turns out to be a simple random walk with one absorbing state. Other results may then be introduced without proof. The researchers will just have to take many things on faith.

4. SOME WORDS OF CAUTION

Some researchers may become over zealous in the application of stochastic processes. They may assume that the model they have set up explained all there is to be explained about the random phenomenon under study. This is usually not the case. Klemes (1994) had written forcefully, in colourful language and with examples on this point for the case of models in hydrology. Klemes warned against attempts "to create knowledge about nature by misguided manipulation of mathematical formulae". His other comments and the comments he quoted in his paper (forty-three references) were for the main part concerned with the building of hydrologic models. These comments are, however, also relevant when one tries to construct stochastic models.

If a deterministic model is adequate, there is no necessity to consider a stochastic model.

5. IMPLICATIONS FOR RESEARCHERS IN STATISTICAL EDUCATION

More research needs to be carried out on how to teach stochastic processes to researchers. One area is the writing up of suitable illustrative examples in the various subject areas. Cox and Davison (1994) gave some suggestions as to what types of examples are suitable for some fields of engineering sciences. The actual examples still need to be written up. One way to obtain these examples is to look at information in as many textbooks as possible. Another way is to seek help from experienced researchers. It will be nice if it is possible to come up with materials that are like Saville's (2001) workshops for agricultural researchers. In the mean time, we have to make do with textbooks and the researchers who want to learn stochastics processes as our guides.

Another area is the research into the teaching of mathematical techniques in stochastic processes. Some researchers in mathematical education might have done some work on some of these techniques. Perhaps there are things that we can learn from them. At the present time, more graphical displays, line graphs, diagrams and simpler examples, such as those covered in this paper, can be used to illustrate the mathematical techniques.

6. IMPLICATIONS FOR THE TRAINING OF RESEARCHERS

Researchers have to be taught with as little mathematics as possible. Where possible some computer simulation of the stochastic processes under study may be demonstrated to the researchers. Basic concepts and mathematical techniques have to be carefully explained with, where possible, examples drawn from the researchers' field of interest. Software packages, if available, may be used.

7. CONCLUSION

Stochastic processes are new topics for most researchers. Most researchers begin the study of stochastic processes with a vague understanding of the essential concepts and principles of probability and usually with a lack of mathematical skills. The task of teaching them stochastic processes is not an easy one. The material for the course has to be taught in small quantities and at the same time alternate with tutorial problems.

Graphs and diagrams need to be used to help researchers to understand and learn the concepts. Using these graphs and diagrams in tutorial problems will help them to better understand the tutorial problems and to solve them too. Researchers need to be reminded to think and to practise, practise and do more practice in solving problems on his/her own.

Mathematics features prominently in stochastic processes. One cannot avoid using mathematics. What can be done is to use as little mathematics in teaching as possible.

New concepts may be introduced by using coin tossing and coloured balls in boxes examples. These examples are mostly devoid of ambiguity. When the researchers are comfortable with the new concepts then examples from the researchers' area of interest can be used to illustrate the applications of the new concepts.

The concepts and techniques learned in Markov chain, Poisson process and birth and death processes will enable the researchers to understand other specialised stochastic models with more ease.

APPENDIX

Websites and a paper related to software for stochastic processes:

1. Tijms, H. C. (1994) at Vrije Universiteit, Department of Econometrics, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Package name: STOCHAST.
2. Website: http://www.econ.vu.nl/medewerkers/tijms/default.htm. Reviewed in OR/MS Today, February 1993, 60-62.
3. MARCA: http://www.csc.ncsu.edu/faculty/WStewart/MARCA/marca.html
4. Goel, P. K., Peruggia, M, & An, B. (1997) Computer-aided teaching of probabilistic modeling for biological phenomena, *The American Statisticians*, 51, 164-169.

REFERENCES

Bhat, U. N. (1984). *Elements of applied stochastic processes*, Second Edition. New York: Wiley.

Bhattacharya, R. N., & Waymire, E. C. (1990*). Stochastic processes with applications.* New York: Wiley.

Borovcnik, M., & Peard, R. (1996). Probability. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (Vol. 1, pp. 239-287). Dordrecht: Kluwer.

Cox, D. R. (1997). The current position of statistics: A personal view (with discussion). *International Statistical Review*, 65, 261-290.

Cox, D. R., & Davison, A. C. (1994). Some comments on the teaching of stochastic processes to engineers. *International Journal of Continuing Engineering Education*, 4, 24-30.

Cox, D. R., & Miller, H. D. (1965). *The theory of stochastic processes*. London: Chapman.

Doyle, P. G., & Snell, J. L. (1984) *Random walks and electric networks*, Mathematical Association of America. http://front.math.ucdavis.edu/math.PR/0001057

Falk, R. (1986) Conditional probabilities: Insights and difficulties. In R. Davidson, & J. Swift (Eds), *Proceedings of the Second International Conference on Teaching Statistics* (pp. 292-297). Victoria, Australia: International Association for Statistical Education.

Feller, W. (1957) *An introduction to probability theory and its applications*. Volume I. New York: Wiley.

Gardiner, C. W. (1997). *Handbook of stochastic methods for physics, chemistry and the natural science*. Berlín: Springer-Verlag.

Harraway, J., Manly, B., Sutherland, H., & McRae, A. (2001) Meeting the statistical needs of researchers in the biological and health sciences. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 177-195). Granada: International Association for Statistical Education and International Statistical Institute.

Isham, V. (1991). Modelling stochastic phenomena. In D. V. Hinkley, N. Reid, & E. J. Snell (Eds.), *Statistical theory and modelling* (pp. 177-203). London: Chapman.

Kao, E. P. C. (1997). *An introduction to stochastic processes*. Chicago: Duxbury Press.

Karlin, S., & Taylor, H. M. (1975) *A first course in stochastic processes*. Second Edition. San Diego: Academic Press.

Klemes, V. (1994) Statistics and probability: Wrong remedies for a confused hydrologic modeller, in V. Barnett, & K. F. Turkman (Eds.) *Statistics for the Environment 2: Water Related Issues* (pp. 345-366). Chichester: Wiley.

Kulkarni, V. G. (1995) *Modeling and analysis of stochastic systems*. London: Chapman.

Moore, D. (1990) Uncertainty. In L. Steen (Ed.), *On the shoulders of giants: New approach to numeracy*, (pp. 95-137). Washington: National Research Council,

Rodriguez-Iturbe, I., Cox, D. R., & Isham, V. (1987) Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society of London*, Series A, 410, 269-288.

Saville, D. J. (2001) A hands-on, interactive method of teaching statistics to agricultural researchers. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 197-213). Granada: International Association for Statistical Education and International Statistical Institute.

Stern, R. D., & Coe, R. (1984) A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society*, Series A, 147, 1-34.

Taylor, H. M., & Karlin, S. (1994) *An introduction to stochastic modeling*. Boston: Academic Press.

Tijms, H. C. (1986) *Stochastic modelling and analysis: A computational approach*. Chichester: Wiley.

Tijms, H. C. (1994) *Stochastic models: An algorithmic approach*. Chichester: Wiley.

Weiss, G. H. (1983) Random walks and their applications, *American Scientist*, 71, 65-71.

Ann-Lee Wang
University of Malaya,
Institute of Mathematical Sciences,
50603 Kuala Lumpur Malaysia.
E-mail: j2wang@ns1.umcsd.um.edu.my

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 87-101.

Printed in Granada, Spain

ALAN MCLEAN

STATISTICS ON THE CATWALK: THE IMPORTANCE OF MODELS IN TRAINING RESEARCHERS IN STATISTICS

This paper emphasises the pervasive role of probability models in statistics, and the importance of the role of prediction in statistics. I argue that all thinking, including everyday decision making, is based on the use of models: theories, stereotypes, metaphors, stories, myths, equations, diagrams, blueprints. Scientific thinking, including statistical thinking, is not different from 'everyday' thinking, but is a formalisation of it. Just as we 'learn from experience', a scientific theory is tested against observed data.

Statistics is a body of techniques for developing and assessing models, particularly those involving uncertainty. This modelling process takes place at all levels of a statistical analysis, not only in 'model selection'. Examples are given to illustrate these processes. Particular attention is given to the role of hypothesis testing, showing how it is a form of model selection between two models, one of which is 'privileged'. Researchers, if they are to understand the role of statistics in scientific research, must understand the role of models in science generally and in statistics in particular.

1. INTRODUCTION

Practising statisticians are very familiar with the concept and use of statistical models, but models appear very little, if at all, in elementary courses or in the texts on which those are based. This can reasonably be considered a serious deficiency, because all statistical work involves the use of models, just as all scientific theorising does.

While one can argue for a very elementary introductory course which is data driven, which barely mentions probability, and is oriented toward everyday life applications, researchers have a strong need for a more thorough and sophisticated understanding of the role of statistical methods in their disciplines, and the role of models in statistics. The statistical training of researchers in many fields, however, is frequently little more than a traditional trek through elementary techniques, so they are likely to be only tentatively aware of the pervasiveness of models.

This paper represents an extension of my thinking. The core of the predictive or forecasting approach to statistics (McLean, 1998) is that of the pervasive role of probability models in statistics, and the importance of the role of prediction in statistics. In the paper presented at the 52nd ISI Session (McLean, 1999b) I considered some aspects of hypothesis testing. In this paper I extend this thinking in two ways: first, a more complete consideration of the nature of hypothesis testing, and second, consideration of the relationship between everyday thought processes and those involved in statistics, and science in general.

My reading connected with the well-known hypothesis testing controversy (including Morrison & Henkel, 1970; Henkel, 1976; Gingerenzer, 1993; Harlow, Mulaik, & Steiger, 1997; Batanero, 1999; and Ito, 1999) suggests that there is

considerable confusion among researchers as to the nature of testing. Discussions in email mailing lists reinforce this conclusion. The recognition of the role of probabilistic models in statistics enables a simpler, more consistent view of the nature of statistics, and of the nature and role of hypothesis testing. In this paper I present some thoughts in that direction.

Statistical and scientific thinking do not differ from everyday thinking to the extent that may commonly be thought. In this paper I discuss the ways in which all these modes of thinking involve the use of models, and consider in a very general way some of the characteristics of models. The major part of the paper deals with the way models apply in statistics at every level, from basic descriptive statistics on. As mentioned above, a substantial part of this section deals with hypothesis testing.

2. MODELS

Models constitute our knowledge of the universe

There are a great many philosophical arguments about the nature of ‘reality’ and our perceptions of it, and the nature of the ‘scientific method’, but it appears to me that the most useful simple approach is as follows.

Science, as a body of knowledge, comprises two types of information structure: observed *data*, and *theories, which* relate these data to each other. Measuring the data may involve passive observation – for example, recording the numbers of birds nesting in a sanctuary each season – or may involve experimentation – for example, recording the time taken for a sphere to travel down an inclined plane. With some philosophical qualifications, the data represent facts, the observable face of the real universe.

It is tempting to think of a theory such as the theory of relativity, of evolution, of the electron, or of covalent bonding as describing how the universe really works – the hidden mechanism which produces the results - but it is more accurate to think of each theory as a *model* of ‘reality’, not the real thing. For example, Newton's laws constitute one model of motion; special relativity a more general model. Both work very well. Further, special relativity can be thought of as a refinement of Newtonian mechanics.

Evolution, in the sense that living things change over generations, is an observable fact, as any animal or plant breeder can attest. So is the existence of different but clearly related species. In the terms I used above, these are data. Darwin’s Theory of Evolution provides an explanation for the development of species through the mechanism of natural selection. This theory has itself developed over the decades since Darwin.

Models are not restricted to science. A religious belief or a philosophy is a model of the universe. Freud and Jung developed models of how the human mind works. According to the context, a model may variously be described as a metaphor, a story, a mathematical structure, a stereotype or a myth. In all of these, some aspect of ‘reality’ is described, well or badly.

Indeed, we use models unconsciously in all decision making, from deciding when to change lane when driving to selecting one’s partner: *'Is that car likely to accelerate now? "Is that girl likely to say yes if I ask her out for a drink?*

We think of other people in terms of stereotypes. A stereotype is a model - the real person is replaced by a generic dummy, based on age, sex, race, etc, the details of the model depending on one's personal experience, upbringing and education: *'He is dark and different and looks dangerous, so I'd better be careful. "She is friendly and knows my parents, so what she is selling must be a good buy.'*

By their nature, stereotypes are simplistic, often wrong, and therefore prone to

offend some people when expressed. But we all use them, whether or not we ‘should’. What is your first image of someone who is ‘female, blonde and good looking’? ‘male, dark and different’? ‘middle aged, male, well-dressed in a dark suit’? ‘young, bearded, hairy, in jeans and thongs’? My first picture of a person – the model on which I initially operate – is based on first impressions. As I get to know him or her better, my picture improves – I work with a better model.

Properties of models

The purpose in creating any model is to achieve some understanding of the working of the universe, and hence to gain control of some sort. This control is expressed primarily through the ability to predict what will happen under given conditions. In short, the prime characteristic of a model is that it be *predictive.*

A model may be *deterministic*: it says what will happen under specified circumstances. Newton's laws and relativity are deterministic. A model may be *probabilistic*: it predicts in the sense that it specifies what can happen, and assigns a probability to each possible outcome. A *causal model* provides predictive ability through a theoretical framework relating the variables involved such that if one or more variables are changed, the results can be predicted. The results may be deterministic or probabilistic. A scientific theory should be a causal model. Such a model is superior to a *purely predictive model*, in that it gives a description of the way the universe works, and provides a measure of control over the results of actions.

Note that the personal models described above (for example, stereotypes) are generally predictive probabilistic models. Experience (or brain washing!) perhaps has led me to believe that 'short, fat, bald men driving Mercedes tend to be aggressive, careless drivers'. There is no theory that connects 'short/fat/bald/Mercedes' with 'aggressive/careless'. Apart from an expectation that men driving Mercedes are probably wealthy, and to get that way they have probably been aggressive, the prediction is based only on an association that I have formed between the two sets of characteristics.

Models and truth

A model is only ‘true’ internally. *Within* a model certain things are taken to be true, in the same sense as in mathematics; assumptions are made, and ‘truths’ deduced from these. Externally, a model is a *good* model if it can be used to predict with adequate confidence what will happen in certain circumstances.

3. THE SCIENTIFIC APPROACH

There has been much discussion over the years about what the Scientific Method is, and whether it exists. My impression is that this argument springs primarily from an overly prescriptive definition, in an attempt to identify it as something quite different from other modes of acquiring understanding of the universe. The scientific approach grows naturally out of everyday ways of acquiring knowledge. I observe something interesting, conjecture something about it, and test my conjecture to see if it is correct. For example, I wake early, see that it is light, conjecture that it is day, and look at the clock to check this. Or I arrive at a class, find low attendance, think of possible reasons – and ask those students who are there if there is an assignment due in another subject soon. Or I check my bank balance, find it lower than I expect, think of possible reasons again – and look for a statement of transactions from the bank. Or I observe politicians

telling lies, theorise that politicians usually lie, and check this by observing politicians speaking. These examples are trivial, but they illustrate that in everyday life we develop (usually ill-defined) theories and test them using information.

In scientific research this process is formalised. Conjectures are structured as well defined models and, ideally, tested by *experiment*. The word ‘experiment’ has a range of meanings, but usually implies some measure of control over one or more factors. In a statistical experiment this control may be obtained by using randomisation, either through stratified sampling or random allocation of treatments. In many cases, however, this control is very limited; at worst, perhaps, the experiment consists of using whatever relevant data is available.

Desirably, the experiment should be as objective as possible and the results reproducible. Again, in many discipline areas, these properties may be limited. For example, in research in education and psychology there are likely to be severe problems of definition, Some examples are given below:

1. A model of planetary rotation may be tested by observing sunrise and sunset times.
2. A model of electromagnetic radiation may be tested by setting up a laboratory, experimenting to generate radiation under specific conditions, and measuring magnetic fields.
3. A model of bird behaviour may be tested by identifying birds living under particular conditions (or by generating those conditions) and observing their behaviour.
4. A model of bird behaviour may be tested (as with Skinner's pigeons) by endeavouring to train them and recording the effects of this training.
5. A model of variation in communication skills with gender may be tested by taking random samples of males and females and testing the people chosen.
6. A model of the effectiveness of a new drug may be tested by taking a random sample of people, randomly selecting some to be treated with the drug, the remainder to receive a placebo, and the effects measured.

Of these, the first two models are deterministic (with some measurement error to be dealt with). The third and fourth may be probabilistic. The last two are certainly probabilistic models, and the experiments described are standard statistical experiments. In 5, the model will say something like: 'A randomly chosen male is likely to have better communication skills than a randomly chosen female.' The precise interpretation of 'communication skills' will depend on the test carried out. In 5, the model will say something like: 'If a patient is treated with this drug, he or she will probably be cured.' Again, the precise meanings of 'treated' and 'cured' depend on the details of the experiment.

In research then, a model is developed and experimentally tested against observed data. If the data do not support the model, the model is rejected, or modified and retested. The researcher is not identifying the ‘truth’ about the universe, but simply establishing an explanation of how observable data are generated. This explanation is tentative, in that in the future further data may cause it to be rejected or modified. Models develop over time, becoming established as they survive the impact of further tests against data. A new model is more readily accepted if it is consistent with established models. This amounts to saying that the new theory is tested against the same data as was used for the earlier theories.

The scientific method is apparently very distinctive in that it concentrates on testable models and requires them to be tested. For example, religious models, and models which may be considered similar, such as 'New Age' ideas, are generally not tested, and

in many ways not testable. However, although some of these models appear to continue to be accepted despite the observable data, some do become unacceptable. For example, few people these days accept a model that says that the earth (globally, rather than locally) is flat or that it rests on a tortoise. A similar comment applies for the models that everyone uses to make decisions in daily life. People do learn by experience - although most people are in some ways 'slow learners'!

4. MODELS IN STATISTICS

Statistics is a body of techniques for developing and assessing models, particularly those involving uncertainty or ‘noise’. This modelling process takes place at all levels of a statistical analysis. Note that in statistics we can talk of generic models, such as 'the linear regression model', the 'general linear model', etc. In this paper when I refer to a model, I mean a particular model; for example, a model describing the relationship between communication skills and gender. This model may be a fixed mean model, or a linear model, or a non-linear model of some sort.

One can identify three general ways in which statistics works with models. First, some form of averaging can be used to smooth out measurement error. This can be considered as part of the measurement process, so is hardly ‘statistical analysis’. The measurement noise can also be considered part of a probabilistic model, and its smoothing as part of the estimation of that model.

Second, the use of causal models with a probabilistic component is one of the primary applications of statistics in many disciplines. Statistical methods are used in the development phase of such a model, in estimating its parameters on the basis of data. They are also used to verify the model, again on the basis of data; that is to ‘test’ the model. Some statistical methods are of course developed to deal with specific types of model.

Statistical methods are particularly applicable in environments where there is a large amount of uncertainty, such as economics, biology, psychology and sociology. In such cases, testing a theoretical structure involves techniques of identifying possible contributions to that uncertainty, and hence ways of controlling and minimising it.

Third, statistics frequently involves development of *purely predictive* models, that is, those which are not causal. These are opportunistic, with little or no underlying theory to account for the relationship between the variables, which is no more than an *association* between them. In practice, it is hoped that a predictive model is in some sense causal. Often there may be some understanding of why the variables are related, but the details are not understood.

According to Occam’s Razor the best model to use is the simplest which satisfies the requirements of the situation. If the purpose of the analysis is purely predictive, a purely predictive model is adequate. If the purpose is to identify a control mechanism, a causal model is required. If there is a choice between a purely predictive model and a causal model of equivalent complexity, the latter would normally be preferred. The unconscious models that individuals use in decision making are frequently purely predictive, and based on experience – ‘sample data’. As long as they enable good decisions to be made, this is adequate.

5. HOW MODELS ARE USED IN STATISTICS

The aim here is to show that *throughout* statistics we are working with models. This includes the rationale for what we do when we do descriptive statistics.

5.1. AN EXAMPLE OF MODELLING SAMPLE DATA

I have a set of incomes measured on a sample (Table 1). I could simply store the data, but to use them I must process them - I form a relative frequency distribution (Table 2), probably grouping the data, graph the distribution, calculate summary statistics. I do this because I am not really interested in the sample at all, except in that it enables me to describe the population from which the sample came. I have only a sample, so I cannot obtain a precise description. Instead, I aim for a model of the way incomes are distributed in the population.

Consequently, I smooth the data by grouping them. This filters out some of the statistical noise, to show any underlying pattern. With luck I can then expect the population to show this same pattern, so it provides the basis for the model of the population.

Table 1. Income Data ($000)

22	23	25	25
17	21	18	19
21	31	27	27
19	25	17	36
26	18	29	23

Table 2. Frequency of Incomes

Income	Frequency
10-14	0
15-19	6
20-24	8
25-29	4
30-34	1
35-39	1
Mean	23.45
SD	5.0312

Calculating summary statistics does the same thing. The distribution is commonly modelled by calculating measures of location and spread, and skewness and kurtosis if these are appropriate. An alternative is to obtain the five-figure summary (median, quartiles, maximum and minimum), often as a box plot. So a model, with the intention of applying that model to the population, approximates the sample data.

5.2. AN EXAMPLE OF MODELLING EMPIRICAL POPULATION DATA

I am interested perhaps in establishing a new business in a particular locality, and I need to know how large an area I should need to consider as my catchment area for customers, so I am interested in the income distribution for Melbourne, and how it varies across localities.

It may be possible for me to obtain this information from the most recent Census, but this information is by no means up to date; in fact it would be out of date on the day after the Census was taken, let alone by the time it is released. Further, even if I view the aggregated data that would be made available to me as 'raw data', I would almost certainly want to smooth it and approximate it further, simply for reasons of efficiency. I do not need the detail available. In fact for this example I would probably use a simpler model: one which gave only the number of households in each locality and their mean income, or one which gave the number of households with income greater than $*x* in each locality.

Further, it is very commonly the case, and arguably always so, that the real population of interest is not clearly defined, so that it has to be modelled using the population data that are available.

5.3. ESTIMATING MODEL PARAMETERS: A MEAN

I take a random sample in order to estimate the mean income of households in a locality in Melbourne, and will obtain a confidence interval in the standard way. It is common for people to speak of this as estimating the 'true population mean' but it is more accurate to say that what is being estimated is the mean of the model used. This model is:

$$Y_i = \mu + \varepsilon_i; \quad \varepsilon_i \sim N(0, \sigma_Y^2); \quad iid \tag{1}$$

where Y_i is the income of the i-th household. This is, if you like, an incompletely specified model, but it is certainly a model. Note that it says that the only parameters of interest are the mean and variance, and the variance does not change as successive observations are taken.

Most important, the requirement for 'independent identically distributed' observations is in many practical applications a very strong assumption! When the population is not clearly defined, mistakes may be made in the selection, and some respondents do not answer or drop out, it is a very strong assumption indeed.

It is of course true that an estimate of the model mean is in turn an estimate of the population mean - but this is likely to change with time, and if the population is ill defined it is not at all clear what the 'true mean' means. So it is easy enough to estimate the model mean, with a confidence interval, but if the model is not a good reflection of the population, the confidence interval is meaningless.

5.4. ESTIMATING MODEL PARAMETERS: REGRESSION

As a real estate agent I want to identify the factors that affect the price of a house for sale. I have taken a sample of recent houses sold by my firm and recorded a number of variables, which are presented in Table 3. (This example excludes the three P's - Position, Position, Position, which refers to an old Real Estate adage in Australia - that the characteristic of a property which overwhelmingly determines its price is its location and assume that the houses are comparable in terms of this variable.) This is a standard example to demonstrate multiple regression.

As soon as we move into this topic it is clear to all that we are talking in terms of models. However, the fixed mean model is rarely mentioned and people still talk in terms of 'true' - 'the 'true' value of the slope of the regression line'. This must be

confusing to students - we talk of 'assuming a linear relationship', then talk of the true values of the parameters of this assumed relationship.

The situation is properly as follows. The basis from which we start is the fixed mean model (1) where Y_i is now the price of the i-th house in the sample. This is effectively the model for house prices in which none of the variability in prices is explained by other factors. If a model involving other factors is to be used instead of this, it must perform better - otherwise we would be using a more complex model for no benefit.

Table 3. Example

House	Price ($'000)	House size (squares)	Age (years)	Block size (000 sqft)	Heating
1	89.5	20.0	5	4.1	1
2	79.9	14.8	10	6.8	1
3	83.1	20.5	8	6.3	2
4	56.9	12.5	7	5.1	2
5	66.6	18.0	8	4.2	2
6	82.5	14.3	12	8.6	3
7	126.3	27.5	1	4.9	1
8	79.3	16.5	10	6.2	1
9	119.9	24.3	2	7.5	3
10	87.6	20.2	8	5.1	2
11	112.6	22.0	7	6.3	3
12	120.8	19.0	11	12.9	3
13	78.5	12.3	16	9.6	3
14	74.3	14.0	12	5.7	1
15	74.8	16.7	13	4.8	2
Mean	88.84	18.17	8.67	6.54	

For the house price data, we may consider a model in which the price of a house is related linearly to its size:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i; \quad \varepsilon_i \sim N(0, \sigma^2_{Y|x}); \quad iid \tag{2}$$

where x_i is the size of the i-th house. The question is whether this model is likely to perform better than the fixed mean model. This is clearly a model selection process, but it is rarely presented in this light. Teachers refer to the 'true line', which is clearly meaningless, hence confusing to students. The decision is usually carried out using hypothesis testing.

Similarly, with a multiple regression model the standard F test compares the model with at least one nonzero coefficient with the fixed mean model. Again, this is a model selection process, but carried out using hypothesis testing. Again, as pointed out in 5.2, the population from which the sample was taken is likely to be ill defined. We would certainly like to apply the results to a wider population; that is, use the estimated model (carefully!) more widely.

5.5. A CROSS TABULATION EXAMPLE

I am interested in the relationship between a person's attitude to drugs and his or her age. I develop a set of statements and look for level of agreement on a standard five-point scale with each. With age, I am essentially concerned with broad age bands: 'young', 'middle' and 'old', suitably defined. I take a stratified sample and administer my

questionnaire.

The sample members are then cross-tabulated for each statement. For each of the age groups the sample frequency distribution provides an estimate of an empirical model for the pattern of agreement. There is no relationship between the variables if the three models are identical; there is a relationship if they are not identical. The choice then is between two 'super models' - one in which the three strata models are identical, with this common model estimated by pooling the data, and one in which at least two of the three differ. Again this is an exercise in model selection.

In this example there are three real subpopulations which are being modelled. In many applications this is not the case. For example, in researching the side effects of a drug, one might wish to compare the effect of the drug in two or three combinations plus a placebo. In this case random selection is used to create the effect of random selection from several *notional* subpopulations. Again, these notional subpopulations are modelled by the data, and the question is whether the models are better taken to be identical, or different.

5.6. AN ANALYSIS OF VARIANCE EXAMPLE

In one way analysis of variance the modelling process is the same as in the cross tabulation examples briefly described above. The nominal explanatory variable is used to divide the population into strata, real or notional. The question is again whether the models for the strata are identical or not. Because the dependent variable is numeric, the strata models can be expressed algebraically, and different levels of difference considered. The simplest set of models is one that differs only in mean. The choice then can be expressed in familiar terms as between the (overall) fixed mean model (1) and:

$$Y_i = \mu + \alpha_j + \varepsilon_i; \quad \varepsilon_i \sim N(0, \sigma_Y^2); \quad iid \tag{3}$$

where α_i is the stratum effect.

6. HYPOTHESIS TESTING

We construct stories of how the universe operates - we call these stories 'theories' or 'models'. Significance testing is one way in which we choose between stories as to which is (probably) more useful in a specified context.

First, hypothesis testing is not restricted to statistics or academic research. If you are told some piece of news or gossip, you automatically check it for plausibility against your knowledge and experience. If you are at a seminar, you listen to the presenter in the same way. If what you hear is consistent with your knowledge and experience you accept that it is probably true. If it is very consistent, you may accept that it IS true. If it is not consistent, you will question it, perhaps conclude that it is probably not true.

IF the news is something that requires some action on your part, you will act according to your assessment of the information. If the news is important to you, and you cannot decide what to do on the basis of prior knowledge, you will presumably go and get corroborative information, hopefully in some sense objective information.

This describes hypothesis testing almost exactly; the difference is a matter of formalism. As indicated in the examples above, a statistical hypothesis test compares two probability models of 'reality', so it is a technique for model selection. It does

however have two special characteristics.

First, one of these models is *embedded* within the other: that is, one model is a particular case of the other. Neither of these models is 'true' - but either or both may be good descriptions of the two populations, in the sense that if you do start to randomly select individuals, the results agree acceptably well with what the model predicts. The role of hypothesis testing is to help you decide which of these is (probably) the better model - or if neither is.

Second, one of these models is 'privileged' in that it is assumed 'true' - that is, if neither model is better, then you will use the privileged model. In most cases, this means the simpler model.

More accurately, if you decide that the models are equally good (or bad) you are saying that you cannot distinguish between them on the basis of the information and the statistical technique used! To decide between them you will need either to use a different technique, or, more realistically, some non statistical criterion. For example, in a court case, if you cannot decide between the models 'Guilty' and 'Innocent', you may always choose 'Innocent'. In more typical statistical applications, the choice is usually (following Occam) the simpler model, commonly the embedded model.

There is no necessary statistical reason why one model is thus privileged. In my earlier paper (McLean 99a) I stressed my belief that this approach reflects our (and Fisher's) cultural heritage rather than any inherent need for it to be that way.

Given that the null model is privileged, a test is only carried out if the sample data suggest that it should be rejected; that is, the alternative model appears to be better. The test provides a measure, the *p*-value of the test statistic, of how much better. If this measure is sufficiently ‘significant’ we decide that the alternative model should be used.

A commonly expressed view is that for a continuous variable ‘the point null must be false’. This objection misses the point completely. It springs from the idea that a test identifies something true (or false) about the universe, and the probability that a parameter equals a particular value is effectively zero. But testing is not about ‘truth’ but about ‘usefulness’; the null model is only a model. It is certainly usable and may be better.

6.1. TESTING IS ABOUT DECISIONS

Fisher’s approach to significance testing is often described as not involving decision making: That a test is used to assess the evidence in favour of the alternative, enabling a statement about the *significance* of this result in the continuing development of scientific knowledge. In fact a significance test does entail a contingent decision, in the sense that the result tentatively establishes current knowledge. Further, this result will be used to determine the future direction of the research.

Neyman and Pearson introduced the concept of explicitly deciding which of two hypotheses is true, and consequently the concepts of type I and II errors.

Recognition that a test is a choice between two models helps us to see that the two approaches differ more in terms of their areas of application than in substantive terms. Fisher’s approach emphasises the idea of *conditional rejection of the null*, so is appropriate in scientific research. The Neyman-Pearson approach applies in areas where a decision clearly must be made, such as in quality control. For example, with a bottle filling machine, which is periodically tested as to the mean contents, the null is that the machine is filling the bottles correctly. Rejecting the null entails stopping the machine; accepting it means the machine will not be stopped.

The role of decision making in testing is not confined to the test itself. In a piece of

research it is often easy enough to identify a characteristic of interest – the problem is how to measure it. If I am interested in the relationship between *ability in statistics* and *ethnic background*, for example, I measure the statistics ability using an examination of some sort; I measure ethnic background by defining a set of ethnicities. There are literally an infinite number of combinations that I can use – infinitely many different exams, all purporting to measure 'statistics ability' (even if I change only one word in an exam, I cannot be absolutely certain of its effect, so it is a different exam!) and a very large number of definitions of 'ethnicity'.

I now apply the test to a group of people of varying ethnicity, score them on the exam and analyse the results, including a hypothesis test, to decide if statistics ability is related to ethnicity. This test might be a simple ANOVA, a Kruskal-Wallis or a chi square test, depending on how I score the exam.

The point here is that the definition of the models being compared in the test includes the definition of the variables used. If I reject the null model I am NOT saying that there is a relationship between statistics ability and ethnicity, only that there is a relationship between the two variables I used.

Please note that the test is not saying this – I am. The test merely gives me a measure of the strength of the evidence provided by the data ('significant at 1%' or '*p*-value of .0135'). This measure is only relevant if the models I have used are appropriate. I can use other evidence to decide if this is so. So in a research project there are three levels at which judgement is used to make decisions:

1. Deciding what variables are to be used to measure the characteristics of interest, and how any relationship between them relates to the characteristics;
2. Deciding on the model to be used, and how to test it;
3. Deciding the conclusion for the model.

In each of these there is evidence we use to help us make the decision. Accepted theory and practice, and general experience, help with the first and the second. With the second there are also related hypothesis tests such as tests for normality. For the third, the hypothesis test provides the evidence.

It is essential to recognise that in research a theory is presented, together with observational evidence to support it. The evidence does not 'prove' the theory to be 'true'. If the evidence is accepted as sufficiently strong, the theory will (tentatively) be accepted as the current model. Further observational evidence may continue to support the theory, or may weaken its support. In any case, the judgement of the scientific community as to what is 'sufficiently strong' evidence is simply that – judgement. It is as objective as possible, but no more.

Statistical hypothesis testing is a tool, which helps with the assessment of the evidence within certain constrained and generally well-defined bounds. Within those bounds it can be extremely useful. Unfortunately those bounds are too often not understood or are ignored. The way hypothesis testing is commonly presented in the textbooks and the classrooms does not help.

6.2. A MULTIPLE REGRESSION EXAMPLE

Selection of the best model in a multiple regression application is a good example of the role of hypothesis testing for students. For the house price data (Section 5.4) the Excel output for a regression against all the variables is shown in Table 4.

The variables *oil* and *electric* are dummy variables introduced to account for type of heating.

Table 4. Excel Output for a Regression Analysis

Regression statistics	
Multiple R	0.983021
R Square	0.966331
Adj. R Square	0.957149
Standard Error	4.366877
Observations	15

ANOVA	df	SS	MS	F	Significance F
Regression	3	6020.470208	2006.823403	105.236689	0.000000
Residual	11	209.765792	19.069617		
Total	14	6230.236000			

	Coefficient	Std. Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-5.328977	7.451269	-0.715177	0.489401	-21.729118	11.071164
House size	4.086875	0.276215	14.796016	0.000000	3.478930	4.694819
oil	-11.12633	2.731337	-4.073584	0.001840	-17.137968	-5.114696
Block size	3.609421	0.572917	6.300074	0.000058	2.348438	4.870404

The *F* test compares the linear model with no restrictions on coefficients:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \varepsilon_i \,;\, \varepsilon_i \sim N(0,\sigma^2) \tag{4}$$

with the constant mean model (1). Each *t* test compares the above linear model with no restrictions (4) with the same model with one restriction; for the first variable:

$$Y_i = \beta_0 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \varepsilon_i \,;\, \varepsilon_i \sim N(0,\sigma^2) \tag{5}$$

The point I wish to make here is that each of these tests provides evidence for a choice. The overall decision as to the best model – within the limits of the data and the type of model considered – is based on a number of tests. And again, it is a matter of judgement, external to the statistical analysis, although using statistical experience.

6.3. A NULL FREE APPROACH

The standard approach to hypothesis testing, recast in terms of comparing models, is to use as the measure of strength of evidence the *p*-value, that is the probability of the sample result (or 'worse') occurring *if the null model is used*. This can equivalently be expressed in terms of critical values. I suggested above that the null need not be so privileged, that the reasons why it is are cultural rather than statistical. It is true that if we formulate the models in the usual way, the null model is fully specified, so a *p*-value can be computed, while with the alternative this is not so. For example, to test

$$\begin{aligned} H_1 &: \mu = \mu_0 \\ H_2 &: \mu \neq \mu_0 \end{aligned} \tag{6a}$$

one is forced to start by using H_1.

If the test is carried out in the usual way, and the null rejected, the question must then be: 'If $\mu \neq \mu_0$, what is the best estimate of it?' Under normal circumstances the answer is 'The observed value of the sample mean.' In a very real sense, then, the true hypotheses of interest are:

$$\begin{aligned} H_1 &: \mu = \mu_0 \\ H_2 &: \mu = \mu_1 \end{aligned} \tag{6b}$$

where μ_1 is the observed value of the sample mean. That is, the choice is between two models that are identical except for the value of the mean. The first model is the best available under one theory; the second is the best available on the basis of the sample evidence.

From this point of view, neither model is privileged. A *p*-value can be computed based on either. In this example, it is reasonable to compute one sided *p*-values, and because the *t* distribution is symmetric, these will be identical:

$$P\left(\bar{x} < \mu_0 \middle| \mu = \mu_1\right) = P\left(\bar{x} > \mu_1 \middle| \mu = \mu_0\right)$$

supposing that $\mu_1 > \mu_0$. This approach appears to be more natural for students.

6.4. BAYESIAN METHODS

The Bayesian approach has not been mentioned, but it certainly forms part of the background. The predictive approach (McLean, 1998) is not unrelated to the Bayesian. For the present topic, my aim has been to clarify the nature and role of classical hypothesis testing, so these matters have been left in the background. The same comment applies to decision theoretic methods. To the best of my knowledge, nothing that I have said is in conflict with these approaches.

7. MODELS AND RESEARCHERS

Researchers using statistical techniques must understand clearly that they are working with models of the real world, not with the real world itself. They are not discovering truths, but creating a better description of the world. This description is primarily predictive. Statistical analysis is concerned with assessing the predictability of results provided by a model; whether the model is also a causal description is outside its scope.

Second, a choice between statistical models based on hypothesis testing is made within the context of a general model. The choice is between different versions of that general model. Consequently, the test is only valid if the assumptions of that general model are valid.

Third, they must understand that a statistical analysis does not, even within the context of the general model, *prove* one version is superior to the other; it simply indicates that one is likely to perform better than the other. All conclusions are tentative, although actions may be based on them.

Lastly, any statistical analysis involves considerable judgement. At the simplest level, what is an appropriate level of significance in a hypothesis test is a matter of judgement. Questions of definition of variables, wording of questions in a questionnaire,

the effect of non-response, are all involved in the definition of the models concerned, and all involve some personal judgement on the part of the researcher.

Whatever methods are adopted in teaching, these levels of understanding must be conveyed in teaching statistics to researchers – as indeed to anyone using statistics.

8. FINAL NOTES

A participant in the discussion suggested that Galileo is an example of how hard it is to defend scientific theories. I totally agree that it is important to make students realise that statistics can help them to decide between competing models.

Another question raised was what reality does a model reflect . The question of the nature of reality, or whether it even exists, is an interesting philosophical question. One can argue that not even observed data are real since they are perceived through the mind. My opinion is that we should in practice act as if there is some underlying 'reality', our observed data, if obtained carefully enough, are more or less 'real'. The important thing is that our theories and beliefs that account for the data, are models.

It was also suggested that reality is subject to change with time. In order to avoid misunderstanding the reality, we do have to look at it either as a dynamical system, or, by fixing time, to model a section of the reality. When we 'look at it as a dynamical system' we are using a model, usually more complex, since the evolution of the system has to be modelled.

Models are important, in the sense that a model is a useful fiction that reflects some aspects of the real world, that a model is our invention, which makes the problems easy to handle. However, we should keep in mind that we are sometimes liable to persist with the models themselves, disregarding the real world. Models make problems *possible* to handle. And we certainly tend to persist with some models when the 'real world 'evidence is that they should be rejected. This happens in both daily life and scientific research.

Another point debated was if I distinguish between subjective and objective probability. I do not believe 'objective probability' exists; if it does, it is inaccessible to us. In general terms, the causes of a particular result (for example, getting heads when a coin is tossed, or selecting an outcome through 'random' selection) are so complex that they have to be replaced by the notion of probability. So probability always occurs as a model. If by 'objective' probability is meant the frequentist approach: this is a widely used method of estimating the parameters of a probability model. Note that it is still subjective, in the sense that judgement is used in this process. The frequentist interpretation of a probability as literally being a (long run) proportion is simply incorrect. The probability of getting a head on the next toss is in no way the same thing as the expected proportion of heads in the next 10,000 tosses.

Models might also be seen as machines that serve to produce knowledge, rather than as images from reality, although I certainly do not see it as a static image. However, the model does not produce knowledge – it *is* the knowledge.

Another important point is students' models and how they sometime make incorrect generalisations. Learning is a process of developing in the mind models of the world, and learning to manipulate those models. Students often establish incorrect models, and must correct those through further learning. This same comment, of course, applies for all people in daily life. We all have mental models that could be improved. The process of improving models, as students or in daily life, can be painful.

REFERENCES

Batanero, C. (2000). Controversies around the role of statistical tests in experimental research. *Journal of Mathematics Thinking and Learning*, 2(1-2), 75-98.

Cook, T. D., & Campbell, D. T.(1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.

DeGroot, M. H. (1986). A conversation with Persi Diaconis. *Statistical Science*, 1(3), 319-334.

Fisher, R. A. (1948). *Statistical methods for research workers*, 10th edition. Edimburgh: Oliver and Boyd.

Gingerenzer, G. (1993). The superego, the ego and the id in statistical reasoning. In G. Keren, & C. Lewis (Eds.), *A handbook for data analysis in the behavioural sciences: Methodological issues* (pp. 311-339). Hillsdale, NJ: Erlbaum.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.

Henkel, R. E. (1976). *Tests of significance*. Beverly Hills, CA: Sage.

Howson, C., & Urbach, P. (1989). *Scientific reasoning: The Bayesian approach*, La Sale, Ill: Open Court.

Ito, P. K. (1999). Reaction to the invited papers at the meeting on statistical education and the significance tests controversy. *Proceedings of The 52nd Session of the International Statistical Institute*. *Bulletin of the International Statistical Institute* (Tome LVIII, Book 3, pp. 167-168). Helsinki: International Statistical Institute

McLean, A. L. (1998a). The forecasting voice: A unified approach to teaching statistics, In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics* (pp.1193-1199). Singapore: International Association for Statistical Education and International Statistical Institute.

McLean, A. L. (1999b). The predictive approach to teaching statistics. *Department of Econometrics and Business Statistics Working Paper* 4/99. Melbourne: Monash University.

McLean, A. L. (1999c). Hypothesis testing and the Westminster system. *Proceedings of The 52nd Session of the International Statistical Institute, Contributed Papers* (Book 3, pp. 287-288). Helsinki, Finland: International Statistical Institute.

Morrison, D. E., & Henkel, R. E. (Eds.) (1970). *The significance tests controversy. A reader* Chicago: Aldine.

O'Conner, M. P., & Spotilla, J. R. (1992). Consider a spherical lizard: Animals, models and approximations. *American Zoologist*, 32, 179-193.

Popper, K. R. (1977), *The logic of scientific discovery*, 14th ed. New York: Routledge.

Alan McLean
Monash University, Department of Econometrics and Business Statistics,
900 Dandenong Road, Caulfield East,
Victoria 3150, Australia.
E-mail: alan.mclean@buseco.monash.edu.au

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 103-112.

Printed in Granada, Spain.

GUDMUND R. IVERSEN

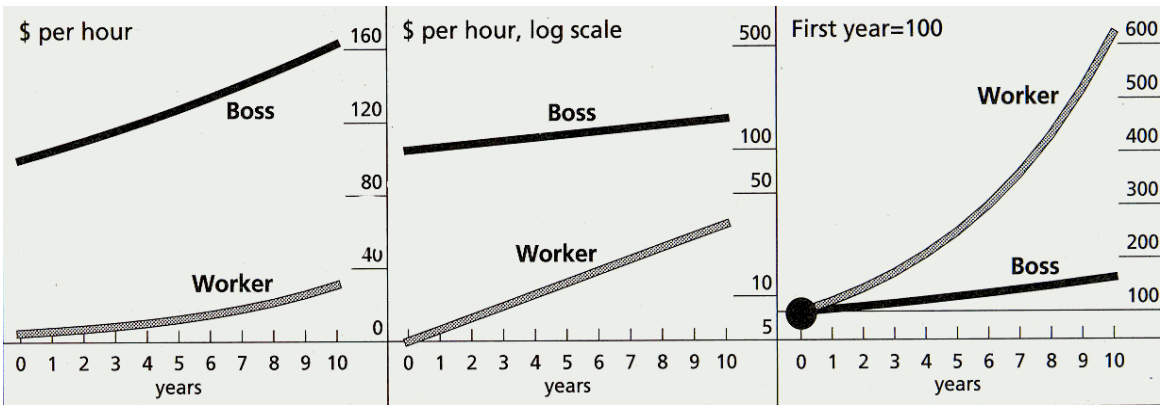
BAYESIAN MODELS AND WORLD CONTRUCTS

Statistical methods have an impact on the results of any statistical study. We do not always realise that the statistical methods act in such a way as to create a construction of the world. We should therefore be more aware of the role of statistics in research, and the question is not so much about what we teach researchers but that we train them to be aware of the impact of the methods they use. This becomes particularly important in statistical inference where we have the choice between the classical, frequentist approach and the Bayesian approach. The two approaches create very different views of the world. The paper explores the relationship between a model chosen before the analysis and the construction of the world after the analysis. Bayesian statistics may ease the conflict between model and construct.

1. WHO IS BETTER OFF?

The English weekly newsmagazine *The Economist* once showed Figure 1 in an article written as part of a series on statistics (Source: *The Economist*, May 16, 1998, p. 79).

Figure 1. A Comparison of Wages for Bosses and Workers



The purpose of the graph was to make a comparison between the wages of bosses and workers. The comparison was made with time series data over a ten-year span, and the graphs plot three aspects of wages against time.

2. COMPARING GROUPS

Statisticians are very good at comparing groups. Typically, the comparison of two

groups is made into a comparison of two means or, perhaps, a comparison of two percentages. If necessary, it is even possible to compare quantities such as two variances or two regression coefficients. Statisticians are good at computing the proper test statistic and finding the resulting *p*-value that will help decide whether the difference between the groups is statistically significant or not.

But are statisticians doing the "right" thing by making such comparisons? What does it mean to say that two groups are different with respect to some variable? A researcher brings a statistician data on two interval/ratio variables for observations in two groups and asks for help to find out whether the groups are different or not. The researcher would typically have had a statistics course or two, particularly if she is doing biological research. If she is a social scientist, her training in statistics may have been rather minimal. The statistician's immediate response is to do a *t*-test for the difference between two means, assuming no wild departures from normality and that the underlying variances are not strikingly different. The statistician enters the data in a statistical software package of some kind and asks for the *t*-test. Based on the *p*-value returned by the software package, the statistician will tell the researchers whether there is a statistically significant difference between the two groups or not.

Statisticians are conditioned to do this from a long career in statistics, and hardly ever do they stop and consider whether they have done the right thing. Does it make any sense to compare the means? The statistician, forcing the comparison of two groups into a comparison of two means, implies that the statistician has constructed a reality of the world for the researcher, whether she wanted it or not. Maybe the comparison of means forces the re-searchers down paths they have not intended to take. After all, there are many ways in which things can be different.

Two groups being different can mean that *all* the observed values in one group are larger than the observed values in the other group. "Different" can mean that *some* of the values in one group are larger than *some* of the values in the second group. Different can mean that some type of an *average* is larger in one group than the other group, be it the mean, the median, or whatever the favourite average may turn out to be. So, what is the meaning of "different" when it comes to a comparison between Workers and bosses? What can we conclude about what that world out there *really* is like for the two groups?

The graphs in Figure 1 provide quite an eye-opener for students when they are exposed to them. First, it is possible to probe the students to see what their thoughts are on the meaning of the statement that two groups are different. The end of the class discussion usually consists of an agreement that two groups are different on some variable if the two corresponding means are different. Students sometimes go as far as saying that it could be that the two medians are different, when they remember something about skewed distributions. When the question is made more concrete, and students are asked to think of a comparison between wages for blue-collar worker and white-collar workers, they usually respond that the white-collar workers would be expected to have higher wages, and so the groups are different.

When students are pushed, they propose that there is a list somewhere containing wages, and they base their answer on the existence of such data. The implication of their answer is that they base their thinking on the existence of a true fact out there, in the "real world." There are wages out there, and different groups have different wages. That is a fact about the world.

The students think they know what they mean by the study of the difference between two groups, but do they really? Are there *facts* out there, waiting for statistics to be discovered? The same discussion in a group of statisticians would not have been very

different; perhaps more sophisticated, but in the end not very different.

However, when faced with the graphs about wages for workers and bosses, students are no longer as certain about the factual world as they were in the beginning. The figure contains three different graphs displaying the same factual world, but the conclusions from the three graphs are very different. According to the first graph, the wages in dollars per hour are plotted as a dependent variable on the vertical axis against time as an independent variable on the horizontal axis. For ease of comparisons, the points for each group have been connected by curves, which results in two curves, one for the bosses and one for the workers. The top hourly pay is $160 for the bosses in the last year.

The first graph shows the curve for the bosses to be considerably higher than the curve for the workers across the ten years. From that it may be possible to conclude that the bosses are better off than the workers are. Is that is the way the world really is? Is that a fact that has been uncovered about the world? Or is this maybe simply a construction of the world we have created and are now, as statisticians, forcing on the researcher and thereby on those who read the research report? Does having more money even mean being better off? And does the researcher recognise that we, the statisticians, have added something to the research finding? The result is not just the data speaking, it is also a particular way of displaying the data that is speaking.

Turning to the middle graph, the dependent variable has been changed. Instead of actual wages per hour, the wages have been transformed into logarithms. This bothers students right away. They say they do not live on logarithms of money, they spend real dollars and cents. Students are not as familiar with logarithms as they used to be, now that cheap calculators are readily available for multiplication and divisions. After some discussion, however, it is possible to get students to understand that the use logarithms of wages makes it possible to see percentage increases over time, and that is what the middle chart shows.

In the middle chart, the points for the years are again connected to give us two curves, one for the workers and one for the bosses. The curve for the bosses still lies above the curve for the workers, but now the curve for the workers is rising faster than the curve for the bosses. Somehow, the workers are gaining on the bosses. Maybe they will even pass the bosses some day. The workers may not be as badly off, after all. Well, this is a different reality we use statistics for to construct and paint for the researcher. What is the researcher to do? All of a sudden, maybe statistics is not as helpful as she thought it would be.

The third graph, on the right, again shows two curves. But this time the curve for the workers lies above the curve for the bosses. Earlier, we thought we had shown that the bosses are better off than the workers are! The graph shows annual wages, and setting both wages equal to 100 at the beginning of the time period compares them over time from a common base. The curve for the workers now goes up much more rapidly than the one for the bosses, meaning that the workers have gained more than the bosses have.

3. OUR CONSTRUCTION OF THE WORLD

Each one of these three graphs paints a different picture of the world. The same data are used in the three situations. Still, through statistics, we have constructed three different realities. So, which is the "right picture"? The obvious answer is that none of them is the right picture; it all depends upon how we look at it. This may not be what

the researcher wants to learn. And this may be where we have not taught our students, turning into researchers, that *the results all depend.*

Lawyers are used to thinking this way. A case before a judge and a jury is not so much about what is true and false, and whether there is a real world out there? Instead, it is getting a client acquitted, in case of the defence, and getting the accused to be found guilty, in the case of the prosecutor. From time to time both sides even invite statisticians to appear as expert witnesses. It is always amazing to see how two statisticians can use the very same data to arrive at very different conclusions. Statisticians on the two sides of the case construct two very different realities of what the world is like and hope that the jury will accept their constructs.

In answer to the question of what we should teach researchers, maybe we should include in our courses a visit to a courtroom and watch statisticians in action. That will very quickly make anyone realise that there is not one, factual world, for us to discover. It is not that researchers should learn certain, specific statistical methods, and after they learn those methods, all is well. Instead, it all depends. Maybe that is the one thing researchers should learn from us: The results heavily depends on the statistical method used, not just the data, and we should not worry about whether they know time series analysis or incomplete two-way analysis of variance or whatever else we teach. Maybe we should not even teach them statistics; they should come to statisticians for their statistical analyses.

4. LIBERAL ARTS EDUCATION

At this point, let us take a small detour into the American system of higher education. Perhaps the greatest American contribution to higher learning consists of the system of liberal arts education. For four years, after secondary school, students attend a liberal arts curriculum, are not expected to learn a profession, but instead are expected to immerse themselves in a liberal arts way of approaching life. Those who teach at such a liberal arts college, their task is not so much to teach specific statistical methods as it is to convey a way of statistical thinking to the students, based on randomness, variation and statistical regularities. The hope is that such an approach will make the students better citizens. If they need technical training in statistics for a career, they will get that through their graduate studies after college.

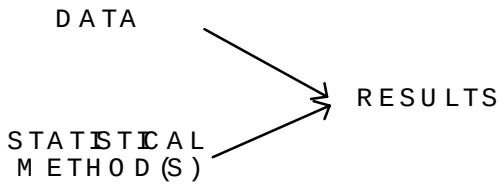
5. EFFECTS OF DATA AND OF METHOD

Schematically, the discussion above can be illustrated as shown in Figure 2. Thus, to consider the training of researchers in statistics, we should consider the schematic view of the research process as it is shown in the graph in Figure 2.

The graph picks up on how the results obtained from a particular research project come from two sources. One source is the data, obviously, and the other source is the statistical method. This little figure always surprises people. We like to think that results from a research process somehow are “The Truth” about the topic being studied. The purpose of teaching students statistics in a liberal arts program is not to make them into researchers or amateur statisticians, being able to do empirical research themselves. Instead, the purpose is for them to be able to understand the role played by statistics in today's society. Hopefully, this graph helps them better to understand the role of statistics in research. Having seen the three sets of curves from *The Economist*, they

begin to appreciate the role played by statistical methods, beyond the data themselves.

Figure 2. Schematic View of the Research Process



We are used to thinking that the data affects the results, and we teach about the presence of sampling variation. This aspect of the research makes sense to the students. What they find discouraging and surprising is that the statistical methods themselves somehow can have an effect on the results. When they learn that *the* correlation between two variables equals 0.87, they like to take this as a fact about the relationship between the two variables in the same way as, in physics, a metal has a specific gravity constant. The students get disappointed when they learn that this correlation is the number we get when we base the analysis on least squares. For example, had we used absolute values instead to fit a line, then the result would have been different. Any other measure of the strength of the relationship between two variables is similarly dependent on how it is defined. Again, the impact of the method shows its ugly presence.

So, how are we to look at the end result of a research process? What should we, as statisticians, impress upon researchers who make use of our methods? We have the responsibility to stress that any statistical result from a research process represents a construction of the world created jointly by our data and by our methods. Just as with the three graphs from *The Economist*, there is no Truth out in the world with a capital T. The results obtained from the empirical world consist of a *construct of the world* the researchers create. It seems as if statisticians often forget that. We gather to discuss what researchers should know about statistics when they go about their tasks. We can make up a wish list of statistical methods that it would be nice if researchers knew how to use in their work. But our work is not done by simply producing such a list. Statisticians have not executed their responsibilities if that is all they do.

6. RESULTS OF THE RESEARCH PROCESS

Statisticians need to do more. They need to make researchers aware of the fact that the result of a research process is a particular construction of the world. This construction comes from the combination of our data and our statistical methods. One implication of this is that it is not as important *what* we teach researchers as it is important that they recognise the full implication of using statistics.

It is tempting to think as a parent and ask what tools do we let our children work with as they grow up. We hesitate to let a young child play with a chain saw, and perhaps we should hesitate letting researchers have access to certain statistical methods unless they are fully prepared and ready for such usage.

One good thing about the pre-computer world of statistics was that statistical methods were not as easily accessible as they are today. It used to be, that to do a ten-variable regression analysis, one put in a good bit of thought about whether it was worth doing before employing several graduate students to compute sums of squares and sums

of cross products and invert matrices.

Now, this question does not even come up any more about whether a particular analysis is worth doing or not. With a few clicks of a computer mouse, the results are there for all to see in a matter of a very short time. Maybe this is not necessarily such a good thing. Maybe statisticians should require people to have a license before they are allowed to use a method such as multiple regression. It is scary to think of all the many misuses that have taken place with such analyses, because the tool is so readily available. In the wrong hands, multiple regression software may be as dangerous as the chain saw in the wrong hands.

Clearly, researchers should know multiple regression, and clearly we should encourage researchers to use multiple regression. But, perhaps we forget to let people know what they are getting into. It has already been alluded to the fact that an analysis based on least squares would give results different from an analysis based on absolute values. Just because historically it was computationally more appealing to use squares than absolute values, and the derivative of a square is easier to work with than the derivative of the absolute value function, we should not necessarily continue to use squares. But more than that, it is important that we tell researchers what they are getting into by using something like regression analysis.

Statisticians do not necessarily tell researchers that their results are partly a consequence of the methods they have been taught. Statisticians let people think that the sums of squares they get are *the* measures of the impacts of different variables. We also let people think that *the p*-value they get for a coefficient is a measure of the degree to which we can reject a null hypothesis and assess the impact of a particular variable.

7. DESCRIPTIVE AND INFERENTIAL STATISTICS

All empirical fields need some way of simplifying their data, and let us distinguish for a moment between descriptive and inferential statistics. Descriptive statistics/exploratory data analysis is not as controversial as inferential statistics, even though both impose something on the research process, and this is not always fully acknowledged. The impact of descriptive statistics is less controversial, and below the discussion is limited to the impact of inferential statistics. Here, of course, the major choice is between classical and Bayesian statistics. More specifically, through the uses of inferential statistics, statisticians impose a major construction of the world based on the results of empirical research.

Statistics is a funny field. Statisticians teach it, but mathematicians, economists, biologists and many other folks also teach it. Statisticians have completely lost control of their field, if they ever had such control, and they have let other people step into the vacuum left by statisticians. One can wonder what a sociology department would say if a department of mathematics and statistics started teaching sociology just because some of the faculty had a couple of courses in sociology in years past.

The goal of most statistics training seems to be to make the students into amateur statisticians who can tell when to use a *t*-test and when to use a chi-square test, but not have any understanding of why they are doing what they are doing. It would seem that the danger of doing this is so great that perhaps we should not teach statistics to other than statisticians. Through our inferential methods we are imposing a worldview that the world perhaps is not ready for.

8. STUDENT REACTION TO INFERENTIAL STATISTICS

The two competing views, classical versus Bayesian, are very different in their approaches to statistical inference. Most discussions of which of the two to use have centred on preferences expressed by statisticians. Statisticians have not listened to their customers and heard anything about what view of the world the customers think should be imposed. Do statisticians realise the heavy hand imposed by statistics, and even more importantly, do the users of statistics realise those heavy hands? And how is it that statisticians have the right to tell the customers what is best for them? If statistics is to be taught, perhaps statisticians should listen more to their customers, the users of statistics.

At ICOTS V in Singapore in 1998, I gave a talk on some of these issues (Iversen, 1988). I said, here I am, a most excellent teacher of statistics who can explain the most obscure points with great clarity. I tell my students there will be a question about the interpretation of a specific confidence interval on the next test, and then I get these kinds of answers:

"95% of the intervals would fall between the two values of the parameter."
"95% of the intervals will lie in this interval."
"95% chance that the actual value will be contained within the confidence interval."
"95% sure that μ is between the two numbers."
"95% of the data will fall within this interval."

These, and other answers, probably show that the students have not studied the right part of the book, and the instructor has not reached them in the treatment of the material in class, particularly if they did not attend class that crucial day when confidence intervals were introduced. But more than that, I see these answers as cries in the wilderness about how the world view we try to construct for our customers is not a world view our customers are comfortable with.

9. THE MANY MEANINGS OF THE WORD *PROBABILITY*.

Maybe one reason why statisticians have such difficulties teaching frequentist statistical inference, aside from the fact that it goes about drawing conclusions in strange ways, is the insistence on the use of the word *probability*. That word carries with it the notion of uncertainty, and that is the reason why many students have difficulties with the concept of a *p*-value. When the *p*-value is introduced for the first time and defined as the probability of rejecting a true null hypothesis, such a definition is received by a sea of blank stares. What we are uncertain about is whether the null hypothesis is true or not, and so here seems to be a way of deciding about that. As we know, this is not so.

One way around this difficulty and help students create the worldview statisticians have in mind, is to limit the usage of the word probability and ask "how often?" instead. When the *p*-value is described as a proportion which tells us how often we get the observed or more extreme data from a population where the null hypothesis is true, then students can construct the world view statisticians have in mind. Statisticians cannot blame researchers for constructing a random view of the world since the probability word is so misunderstood that way.

Students have no difficulties asking what is the probability that a population mean μ

is larger than 100, say. They ask this question right after we start in on statistical inference and long before any mention has been made of Bayesian statistics. Our job is to teach them that the question makes no sense. Instead, what if we ask *how often* is μ larger than 100? Then students answer always or never, depending on the value of μ. Maybe statisticians should banish the use of the world *probability* and substitute *how often*, instead, if we stay with the frequentist approach. Then, perhaps we can stay frequentists and still be honest with ourselves.

10. ROLE OF THE STATISTICAL MODEL

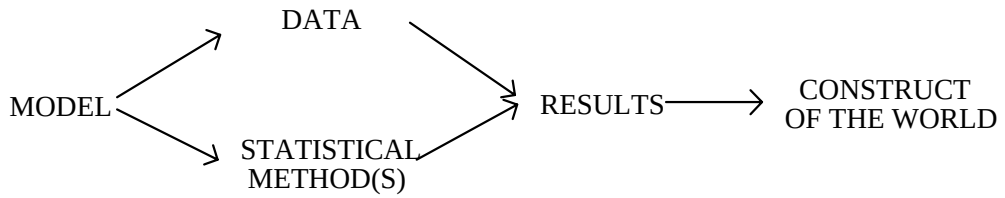
What has been discussed so far is obviously the role of the statistical model in the research process, even though it has not been explicitly said so yet. We were all taught that the model should be stated in terms of properties of the world we wanted to study. From the model we would then come up with the proper statistical method and guidance on what types of data should be gathered. Finally, we would examine in what way the data fit the model we had prescribed. As a particularly beautiful example, we can make a few assumptions of how telephone calls could be expected to arrive, and from that it is possible to derive the Poisson distribution.

How does the issue of a model relate to the construction of the world we come up with after the data have been collected and analysed? I will argue that for many users of statistics, there is a big difference between the model being used and the construction of the world that arises as a consequence of the statistical analysis by the researcher. We may think that the two are the same, and we may want the two to be the same, but often they are not. It is in the difference between the two we may find clues to how we should teach statistics to researchers, particularly how to teach statistical inference, in such a way that they understand what those ideas are all about.

Figure 3 shows an expanded view of the earlier Figure 2 on the research process. On the left side the term "Model" has been added, and on the right side the term "Construct of the world" have been added. The arrows indicate how we start with a model, which then dictates the choice of data and statistical method. From those come the results, and finally the results of the analysis are used to construct a worldview.

How the model and the construct relate to each other? Are they the same or are they different? Ideally, the model arises from the substantive issue at hand, and the construct of the world is the model in its updated form. If the model is one of the standard statistical models we use and it is not fully understand by the researchers, then there is a risk that the construct and model are in conflict with each other.

Figure 3. Schematic View of the Research Process with Model and World Construct



In its most elementary form, a model is something that simply stands for something else. It can be a physical item, such as a model toy car, or it can be a model expressed in some mathematical way. We like to think that the model represents the phenomenon we are studying, and by gathering and analysing data, we will learn more about the model

and thereby the world that it stands for. Statisticians are used to thinking that way, and because they understand the model, their construction of the world will resemble the model. We understand that finding a correlation coefficient of 0.87 implies a model of linear, least square analysis, with its strengths and weaknesses. But not all researchers using statistics think this way. The users often construct their worldview in such a way that the 0.87 becomes a property of the world, as they see it.

How do non-statisticians select and use statistical models? This raises the question of the origins of our models in relation to the uses of our statistical methods. It seems as if many researchers use statistical methods with scant thought to why or why not they should use a certain method. Methods are most often used for their convenience and not because the underlying model in some way fits the problem at hand. The choice of model may not be the correct one, and the data may not even satisfy all the requisite assumptions needed for the use of the model.

All this is particularly true in statistical inference. The two major competitors represent very different models, and they construct their own views of reality. The frequentist and Bayesian constructs are very different, and how does a researcher make a choice between the two? How does a researcher chose between the frequentist and the Bayesian construction of reality? Backing away from that question for a minute, how do researchers view and understand the reality constructed by each of these two approaches?

The attempts of the students to express the idea of a confidence interval in ways that make sense to them, shows a major display of the types of difficulties we are facing in our teaching of statistics to users of statistics. It seems that, for better or for worse, if we teach the students statistics, we should teach them methods such that their construct of the world is as close to the model as possible.

This argument takes us directly to Bayesian statistics. The quotes above on confidence intervals show that many students unknowingly introduce Bayesian interpretations even when they try to do classical inference. In the Bayesian realm, the model views the world as a random world, full of uncertainties. Students intuitively conclude that the best way to model this uncertainty is to use probabilities for what we are uncertain about, thereby taking them into the Bayesian model with prior and posterior distributions.

11. BRINGING MODEL AND WORLD CONSTRUCT CLOSER

So, where does all this leave us? I have argued that researchers construct their view of reality as a function of the data and the statistical method they use. They often overlook or misunderstand the role of the statistical method, and the statistical model that gave rise to the particular method does not necessarily agree with the researchers' construction of the world. This is particularly so in statistical inference. There, to bring the model and construct together, we owe it to the world to use Bayesian statistics. That way, we are permitted to deal with the uncertainty we have about population parameters, and this is the way many researchers construct their worldview.

This is particularly so for people with only a weak background in statistics. But even well trained researchers, who use frequentist methods, very often interpret their results in a Bayesian probabilistic way. Even for many experienced researchers, when they explain what a researcher has learned after rejecting a null hypothesis, we would find traces of Bayesianism. As long as we are uncertain about values of parameters, we will

fall into the Bayesian camp. As statisticians, we owe it to researchers using statistics in their research to make clear the impact statistics has on their work and enable them to choose Bayesian methods. We should train researchers well enough to make it possible for them to understand the role Bayesian statistics can play in their work. That way, the worldview they construct may actually be a reflection of their models.

ACKNOWLEDGMENTS

I am very grateful to Mary Gergen, Roberta Rehner Iversen, George Cobb, Thomas Moore, Rosemary Roberts, Milo Schield and Jeffrey Witmer for comments on earlier drafts of this paper. I also want to acknowledge my intellectual debt to L. Jimmie Savage. Working for him as a research assistant was a unique and very rewarding experience.

REFERENCES

Iversen, G. R. (1998). Student perceptions of Bayesian statistics. In L. Pereira-Mendoza, L. Seu Kea, , T. Wee Kee, & W. K. Wong (Eds.) (1998), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 231-238). Singapore: International Association for Statistical Education and International Statistics Institute.

Savage, L. J. (1981). *The Writings of Leonard Jimmie Savage. A Memorial Selection*, Washington D.C.: The American Statistical Association and The Institute of Mathematical Statistics.

Gudmund Iversen
Department of Mathematics and Statistics,
Swarthmore College, 500 College Avenue
Swarthmore, PA 19081, USA.
E-mail: iversen@swarthmore.edu

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 113-116.

Printed in Granada, Spain.

THEODORE CHADJIPADELIS

DISCUSSION

Wang's interesting paper about teaching and using stochastic processes shows that estimating and using transition probabilities is essential, especially in political and social sciences. Predicting the behaviour of a structure (e.g. an educational system, the World Wide Web, a traffic network, an academic organisation) by using stochastic processes and operational research methods gives the opportunity to evaluate the effectiveness of the structure and to control its future evolution.

The papers by McLean and Iversen deal with the need of understanding and explaining the real world. All of us (statisticians and researchers) try to understand and explain the world using theories and data. Statistics is just one of several methods dealing with the above issue. As statisticians, we use data to estimate and predict, and we assign probabilities to express a complex non-deterministic world.

The Greek philosopher Aristotle said that there are three kinds of events:

- Events for which we know the outcome (that is, events with probability one);
- Events for which we can predict the outcome (that means that we can compute a frequency distribution without using data, working with probabilities);
- Events, about which we can say nothing (we only can estimate a frequency distribution using data, working with statistics).

In any case before reaching a conclusion you need a reasonable explanation for every model or outcome in the frame of the original problem. As Kruskal (1979) said,

> *"A scientist confronted with empirical observations goes from them to some sort of inference, decision, action, or conclusion. The end point of this process may be the confirmation or denial of some complicated theory; it may be a decision about the next experiment to carry out. An end point is typically accompanied by a statement, or at least by a feeling, of how sure the scientist is of his new ground. These inferential leaps are, of course, never made only in the light of the immediate observations. There is always a body of background knowledge and intuition, in part explicit and in part tacit."* (Kruskal, 1979, p. 84).

These three papers deal with the need to use "statistics" and "probability" to understand and describe the "real world". All of us agree that, statistics is not a set of rules and recipes for the analysis of data. It is not reduced to the use of complicated computer programs and nice graphs. It requires a good knowledge of the observed phenomenon, the planning of observation, good knowledge of the data gathered, and description and examination of the hypotheses about the parameters of the analysed phenomenon.

There is an old Chinese proverb: "If you give a man a fish, he will have a single meal; if you teach him how to fish, he will eat all his life". The latter concept means that we design courses that encourage students to think creatively and imaginatively

about their scientific research problems and the role of modern theories of statistical data analysis and modelling. Compare this with courses that usually describe the mathematical solutions of various routine statistical analysis problems without much proof, and are based on probability models for observed data whose validity is not usually checked.

But should we make (or try to make) a "statistician" out of a researcher? Or should we make researchers understand that they need a statistician as a helper, a colleague or even as a leader? Potential information can only be useful if it is generated. One reason why some Japanese industries achieve high levels of productivity is that employees are provided with statistical tools for generating, analysing, and acting on their own information.

As a member of the board of the Greek Statistical Institute I have been involved in many training staff seminars for researchers and university lecturers. Although we try to "educate" them in statistics, talking about assumptions, requirements, statistical and stochastical thinking, they always ask for the analysis of their specific data set, and look for examples that are similar to their own problems. We should educate the statisticians in a more practical sense:

> *"The brilliant minds of mathematical statistics would do well to leave the construction of abstract admissible decision functions, cease to ride martingales into the teeth of zero-one laws and join the few of us who are attempting to stem the tide of confusion"* (Hunter, 1981, p. 113).

Hunter (1981) also criticised statistics education in Mathematics departments saying:

> *"The statisticians' training, narrow and technical, is the orderly climb up a staircase of mathematical problems that each have only one right answer. Later steps rest on earlier ones. Progress is always up. Teachers watch the climbing techniques of the fledgling statisticians, and help them master the steps, one at a time. Statisticians' work, for which this training is supposed to equip them, is the disorderly climbing of rugged hills, outdoors, in fair weather and foul. The path is anything but clear. A promising path can get lost in tangled undergrowth or a patch of dense forest. Or else: a path branches in several directions and there is not enough time or money to explore all of them to determine which is the best to follow"* (Hunter, 1981, pp. 113-114).

Statistical literacy and thinking is another issue. It is useful to the general public in understanding and criticising what is written in the press, what is seen on TV, what is presented by the authorities. We should educate the general public in order to become critical citizens. But this is the story of statistical education in compulsory education.

Let me tell you some conclusions I have drawn from the papers at the conference:

- Learn as much as you reasonably can about the general subject matter field and the specific environment in which the data were collected;
- As part of this effort, statisticians need to probe, be curious, and ask good "non-statistical" questions;
- Correlation measured from an observational study does not imply causation; Confusing correlation and causation is particularly troublesome in the social sciences;
- The real problem is often different from the one initially posed;
- An empirical approach is sometimes better than a theoretical one;
- Scientific logic is our business. Statisticians can often be most helpful by getting

perspective on all aspects of a particular problem and then contributing ideas related to scientific method (Hooke, 1980);
- Try to understand what is really going on;
- Valuable data are sometimes non-numerical;
- Scientific inference is broader than statistical inference.

And finally, let me tell you an old story about the practising statistician by Salsburg (1973). David Salsburg asked himself the question, what is it really like to be a practising statistician. Below we reproduce his answer:

> *"The statistician is first called into consultation during the design of a scientific experiment. At this point, the texts tell us the statistician is supposed to estimate minimal sample sizes and prepare a BIBD that produces all kind of clever contrasts for testing.*
> *I do not do this. Instead, I spend my time asking stupid questions. I know that when the experiment is finished I will have to analyse the data. I must protect myself from impending chaos. With such a fear behind me, I ask such questions as whether it is possible to observe something every 15 minutes or whether this thing they have given a name can, in fact, be observed at all.*
> *I ask them what can be wrong. Frequently, I am the only one at the conference with a non-deterministic outlook. The others conceive only three or four clear-cut outcomes. I think about the in-between outcomes, the two correlated variables that happen to go different ways, the test tube someone is bound to drop, the patient who revives from death's door on placebo.*
> *I know that when it is all over with, the man who must make some kind of decision about the results will ask me to compare two means or to show him a linear regression. I try to make sure that the design will produce two comparable means regardless of how many test tubes are dropped and that, somewhere, there will be a somewhat controllable variable manipulating a somewhat responsible variable.*
> *The bulk of my time, however, is spent in trying to make sense out of data… When I see data it is frequently because the results have not made sense to the client… I feel very uneasy with a client who nods blandly and takes back my numbers for his report. I feel better if he argues with me. After all, I do not know an isatine derivative from an isonitroso-acetlyamine, but I hope to God he does…*
> *I suspect that at least 50 per cent of all the data accumulated today never gets more than a cursory look, and I doubt if 5 per cent of it gets examined effectively. All that money, all that anguish, all that pain will have gone for nothing and even be spent again and again in unknowing duplication unless these floods of data are converted to usable information. This is, par excellence, the place for the well-trained statistician…"* (Salsburg, 1973, p. 152).

REFERENCES

Hooke, R. (1980). Getting people to use statistics properly. *The American Statistician*, 34, 39-42.

Hunter, W. G. (1981). The practice of statistics: The real world is an idea whose time has come. *The American Statistician*, 35, 72-76.

Hunter, W. G. (1981). Six statistical tales. *The Statistician, 30,* 107-117.

Kruskal, W. H. (1979). Comments on Dr Braham's article. *Journal of the American Statistical Association*, 74, 84-86.

Iversen, G. (2001). Should we even teach statistics? A Bayesian perspective. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 103-112). Granada: International Association for Statistical Education and International Statistical Institute.

McLean, A. (2001). Statistics on the catwalk. The importance of models in training researchers

in statistics. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 87-101). Granada: International Association for Statistical Education and International Statistical Institute.

Salsburg, D. S. (1973). Sufficiency and the waste of information. *American Statistician*, 27, 152-154.

Wang, A. L. (2001). How much stochastics processes do we teach and to whom? In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 73-85). Granada: International Association for Statistical Education and International Statistical Institute.

Theodore Chadjipadelis
Aristotle University of Thessaloniki,
Department of Education,
54006 Thessaloniki, Greece.
E-mail: chadji@eled.auth.gr

PART 3

HOW TECHNOLOGY AFFECTS THE TRAINING OF RESEARCHERS

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 119-128.

Printed in Granada, Spain.

SANDRA MCDONALD

PRACTICAL AND EDUCATIONAL PROBLEMS IN SHARING OFFICIAL MICRO DATA WITH RESEARCHERS

Many commentators have noted the need for reform in statistical education. They tend to focus on the analytical techniques that are critical to understanding and producing good quality statistical outputs. This paper adds to these discussions and looks at some of the main analytical issues that transpire from researchers accessing the particular form of statistical data sets in a national statistical office. However the conclusions are much more widely applicable to other data sets. It also considers the more practical, but very important skills and knowledge, applicable to all types of data, that a researcher needs, such as fashioning the data set into a format that is most useful to them, and ensuring they obtain access to data that will allow them to fulfil their research objectives.

1. INTRODUCTION

In this paper I focus on researchers' experiences with using micro data collected for the purpose of generating official statistics, basing that on the experiences of Statistics New Zealand which has recently provided researchers with access to micro data. It provides insights into the specific pitfalls and problems that can be experienced both by the researcher and a national statistical office where official statistical data sets are being used. There are recommendations that both parties can implement to facilitate smoother working relationships.

However the ideas and conclusions do not have to be limited to just the use of official statistical data sets. Many of the experiences can be applied to other data sets, especially where researchers are using data that are collected by somebody else. These are increasingly being made available for wider use by institutions such as universities and data archives like the UK Data Archive. While the data appears to be ready to use, researchers often needs to invest almost as much time and energy in familiarising themselves with the data as they would if they were collecting their own.

2. STATISTICS NEW ZEALAND'S ENVIRONMENT

The collection of data required to produce official statistics is expensive and a significant burden on individual and business respondents. To ensure society gets sufficient return for the cost it is important to make the best use of the data.

In common with national statistics offices (NSO's) around the world, Statistics New Zealand (SNZ), for reasons such as resource constraints, competing priorities and arguably tradition, does not undertake as much in-depth analytical work as it would like.

Besides, it is not always desirable for all analyses of official statistical data collections to be carried out by an NSO. As Biggieri and Zuliani (1999) note NSO's have staff with good quantitative skills but who aren't always so in touch with the sorts of policy issues or decisions that their data sets could be applied to.

To facilitate contributions by other researchers an NSO traditionally makes a wide range of aggregated output available. Increasingly though it is becoming more common for an agency to make unit record data more accessible to researchers through micro data sets. Benefits of access to micro data for researchers include the ability to recode subset and sort the data, derive new variables, and deal appropriately with outliers, but most importantly micro data is essential for using multivariate techniques.

SNZ's legislation does not allow it to publicly release confidentialised micro data but it does permit provision of access to unidentified micro data under certain conditions. This has resulted in the particular form of statistical micro data access available in New Zealand.

The Data Laboratory was established in 1997 to provide a transparent process for managing access to micro data to approved researchers. The objective is to assist SNZ to increase the value obtained from its data sets. For the three years it has been operating the Data Laboratory has successfully resulted in a significant increase in the amount of research undertaken. Prior to the Data Laboratory, SNZ micro data was used infrequently, once or twice a year at most. More recently access has been provided to micro data for up to 15 projects each year.

Overall the Data Laboratory has greatly increased the opportunities for the government sector and the research sector to collaborate for the benefit of public research. It has, however, also highlighted some aspects that researchers using official statistical data sets have found difficult.

These are discussed in detail in the following sections and lead to the conclusions and recommendations for training of social science researchers undertaking research using the statistical data sets collected by other agencies. In its turn, SNZ has identified areas where it can instigate improvements to make the work of researchers' a lot easier.

Table 1: Use of the Data Laboratory (since July 1997)

Types of researchers	Number of projects	Data used
Government departments (nine of the projects were contracted to academic researchers)	23	Population census (4) Household economic survey (4) Household labour force survey (4) Other household surveys (7) Business data (5)
Academic researchers (including post-graduate students)	11	Population census (4) Household economic survey (3) Other household surveys (4)
Research institutes/ independent researchers	2	Population census (1) Household economic survey (1)
Local government	1	Population census (1)

3. UNDERSTANDING ISSUES THAT ARE CRITICAL TO A NATIONAL STATISTICS OFFICE

a. Respondent trust

While the Statistics Act provides SNZ with the power to compulsorily collect data, it

is preferable for the department for respondents to willingly provide information without the need to resort to the compulsory provisions. To maintain high response rates, and therefore provide data collections that are widely regarded as providing high quality outputs, a NSO needs to be trusted by the public, and in particular by its respondents. Any diminution of that trust would quickly affect the quality of the data it collects.

SNZ achieves good response rates. Apart from the obvious benefits to data quality it reduces the need for substantial work that would be required to compensate for low response rates, so reduces potential costs.

It is important, therefore, for researchers to recognise the constraints on policies and operational activities where statistical agencies rely heavily on public trust and good will, and how quickly that trust can be undermined if an individual respondent's privacy is breached. There are a number of occasions where, through loss of trust, a national agency's reputation has been impacted overnight and has only been retrieved over a long period of time and with a great deal of money.

A NSO offering access to micro data needs to be clear about its position on data access within the context of the wider research environment. SNZ's approach is to provide access to micro data in a way that ensures there is little risk of disclosure at an individual level. Where there is the potential for conflict between protecting respondents' information and a research objective SNZ will act conservatively and choose the protection of the respondent. However this may mean that there are constraints on what can be made available to a researcher that have more to do with public perception than with any specific concerns about confidentiality.

Researchers who are exclusively focused on their research objectives may not understand or accept the need for such a stance. While they are unlikely to want the NSO to act in any way that would be detrimental to the quality of what it collects, researchers may not always agree that their proposal will have an adverse impact. A researcher, therefore, needs to be aware of where their research fits within the wider context of the NSO's operations.

b. Reducing disclosure risk in a unit record data set

When a researcher requests access to micro data, SNZ asks them to carefully consider the data they require and to justify the inclusion of variables in the subset of the data they get access to. By providing access only to the data researchers need for their research and not the full data set it assists with minimising disclosure risk. Researchers will not get approval for access to micro data for a project that appears to be a "fishing expedition", where they want all available variables and do not have a well-defined research problem.

Glencross and Mji (2001) and Bishop and Talbot (2001) both identify the importance of this stage. Glencross and Mji describe it as formulating the research problem, involving two key tasks of identifying the 'what' and the 'why'. Where a research project involves the use of data that has been collected by some else, whether by an NSO or some other agency, as opposed to collecting their own there are issues that a researcher needs to consider.

To obtain approval for access to SNZ's micro data researchers need to be very clear about the outcomes they are trying to achieve and to show that the data they are requesting will assist them to attain those outcomes. They should ask appropriate questions to find out what they need to know about the data. This can be a "chicken and egg" situation if they don't have the data in front of them.

The agency holding the data set must make sufficient documentation about the data set available and be prepared to spend some time answering queries about the data collection and how the data set is structured.

Another area that involves negotiation with the researcher is the process of modifying a data set to reduce the risk of disclosure. Because SNZ cannot provide public access to micro data we develop a data set specific to the requirements of each researcher. While we have some standard approaches to reducing disclosure, such as limiting regional detail or providing age rather than birth date, other steps will depend on what data the researcher has requested and also the sensitivity of the data itself. This process is undertaken in consultation with the user to reduce potential adverse impacts on their research as much as possible.

c. Confidentialising output

An issue to emerge from SNZ's experiences with researchers' use of micro data is their different perspective on confidentiality. At the beginning of any project involving the use of micro data we have found it important to ensure that the researcher understands the importance placed on confidentiality to preserve respondent trust.

Researchers often assume that if the data set does not contain names and addresses there is no need for any further confidentiality protection. They are not immediately aware of how easily disclosures can occur in output that is not adequately confidentialised (for example, when a table contains a cell with a single entry or how an individual's information can be disclosed by decomposing data across several independently produced tables).

SNZ realises that its own staff do not find confidentiality an easy concept to comprehend so it is not surprising that researchers are not knowledgeable about it. To assist researchers SNZ has prepared documentation that explains the theory behind the need for the rules that are in place and describes the different types of confidentiality techniques that are needed for different types of data sets, for example censuses as opposed to samples and household as opposed to business data.

When they are using SNZ micro data, researchers need to become familiar with how the department applies confidentiality rules to its data, as they are required to apply the same techniques to their output. This is made easier for the researcher as we provide them with programs and macros developed in-house to ensure that what they do is consistent with our practice. It also reduces the workload of SNZ staff checking confidentiality of output if researchers use standard output protection techniques.

While I am not proposing that researchers should be trained in specific techniques they do need to understand the general issue of confidentiality and to be in a position to appreciate the critical importance of confidentiality to a NSO.

Interestingly the use of micro data by external researchers has highlighted possible inconsistencies with SNZ's output practices. When output made available by SNZ was limited to aggregated tables it was difficult for users to challenge the rules that were applied.

Users were not always in a position to determine whether they were sensible and also whether they resulted in data being suppressed unnecessarily. Through needing to explain, and defend, the rules to researchers who quite rightly question aspects that seem inappropriate, SNZ is much more aware of the need to be clear and consistent in its own practices.

4. COMPLEXITY OF OFFICIAL STATISTICAL DATA SETS

The stage of the research process that involves researchers in organising the collection of data – the D (Do) part of Bishop and Talbot's (2001) PPDSA cycle - must, in cases where they are using existing data sets, be replaced with interrogation of collection documentation and discussions with the collection agent, which for official statistics is the NSO.

In SNZ's experience, researchers find official statistical data sets much more complex than they expect. Common feedback is that there is a steep learning curve to becoming sufficiently familiar with a data collection to be in a position to apply appropriate analytical techniques and to obtain meaningful results. Researchers using official statistical micro data also need to accept that, in the short-term, they are likely to have little influence over data collections. However, as the NSO is exposed to the policy issues and decisions, their data sets are used for useful changes and may feed back into decisions about collections in the long-term.

As Glencross and Mji (2001) note for the education of social science researchers to be effective it needs to relate to the context of the research. Unless data sets that are relevant to their course are available researchers' training is can be restricted by "reliance on a small number of rather tired old data sets which have been used extensively on many courses" (Chant & Lievesley, 1997). Where students collect their own data the data set is likely to be very small and simple. It is therefore unlikely to have many of the problems inherent in real data set, such as a large official statistical data set, which will almost certainly contain imputed and edited data. A researcher may be unaware of the time and effort they will need to invest to produce a data set in a format that they is suitable for their research purposes.

It would be useful for a researcher, particularly those in the social science arena who are likely to use official statistical data at some time in the future, to have some exposure to them during their training. If this is combined with staff from the NSO being available to explain the data set and the collection objectives, methodology and editing processes, there is a great deal of scope for interaction between students and official statisticians that will be of interest and value to both parties.

The processes that a NSO applies to its data sets once the data has been collected are important to researchers, as these will have relevance to the analytical questions that researchers are investigating. Researchers need to understand that editing is undertaken to reduce discrepancies in the data but this may not result in a perfect data set. Edits are usually undertaken to meet the NSO's main aims of producing aggregated statistics and are unlikely to address the specific requirements of research, which may, for example, involve detailed modelling.

The data set could contain erroneous data, caused by keying and transposition errors, which don't impact sufficiently on the use by the NSO for them to worry about completely eradicating them. A researcher needs to know that it is possible for a data set to contain a respondent who is recorded as being born in 1937 but who might actually have been born in 1973. An inexperienced researcher is likely to be unaware of the possibility of such problems and uncertain how to deal with them when they become apparent. They will need to get advice on the potential impact of such matters on their research and then decide how material they will be to their research and determine the potential impact on their proposed outcomes.

SNZ has found it is essential for researchers to have detailed discussions with its subject matter experts to ensure they understand data set contents. Researchers normally

require assistance to determine which variables are most appropriate to their needs, and they will certainly need details on the coding schema. Being familiar with the collection instrument, often a survey questionnaire allows the researcher to determine what the variables represent. Researchers must also be aware of the collection design to ensure their proposed use of the data is supported by what is collected. This is discussed in more detail in section 5 below.

Researchers need to be willing to discuss their research and to take advice on an appropriate data set formats from subject matter experts. This requires skills in communication and the capacity to be flexible about their research plans based on advice received. While it takes time and effort to understand what is collected and the quality and limitations of the collections, researchers that SNZ have worked with have found that this process has assisted them to clarify their ideas, as well as compelling them to put clear boundaries on their project. SNZ also benefits from finding out about the researchers proposed research.

There are other benefits for SNZ. In the past documentation practices have been based solely on internal needs, with staff operating in an environment where systems and processes didn't change very quickly. Faced with increasing numbers of external researchers it was apparent we needed to provide them with well documented information about collections that was produced with a different audience in mind. With developments in technology, change happens quickly so a more rigorous approach to documentation is essential for SNZ to operate effectively. It is also of benefit to researcher, and now meta-data (questionnaires, descriptions of collections, variable lists, quality issues, contact people, and classification schemes) is available on SNZ's web site.

5. SURVEY DESIGN AND DATA QUALITY

It is not intended here to discuss training in the more technical areas of statistical analysis, such as sampling. The need for such training is not disputed and is well covered in past discussions on researchers' educational needs (e.g. Chambers & Skinner, 1998; Jolliffe, 1998; Manly & McDonald, 1998) and many efforts are being made to ensure students improve these skills.

What I will focus on is the implications of such issues within an official statistical perspective. Glencross and Mji (2001) mention the importance of validity and the need to ask "Does the instrument measure what it is supposed to measure?". A researcher using official statistical data, or any data set that already exists, needs the skills to ask, "Did the instrument measure what I needed it to measure?"

Occasionally SNZ encounters the perception that because the data has been collected by the NSO it has a very high degree of precision. In reality, there are quality limitations based around the sample design and collection instrument(s), which limit the types of analyses that can be undertaken. Inferences based on results from very small samples may not be possible because of associated large sampling errors, so while a data set appears promising it may transpire that it is not valid for the desired purpose.

As Jolliffe (2001) points out researchers need to beware of what techniques are appropriate. SNZ's surveys are not usually undertaken using simple random samples but invariably employ a complex survey design. The design will have been developed for the purpose of producing the NSO's primary outputs and therefore may not be entirely suitable for the researcher's purpose. This means that the researcher needs to understand

the relationship between their objectives and the data collection objectives and design, and needs to use analytical techniques that take the complex design into account.

Researchers also need an appreciation of non-sampling errors in general, and the issues that are specific to each data set, including both the level and composition of non-responses, and frame issues such as coverage of population and diminishing quality over time. Questionnaire design may also result in unknown biases in the data obtained, so some experience with the possible effects of the way a question is asked would be useful background for a researcher.

The issues themselves are not unique to official statistical collections and by inference can be extended to the use of many other data sets. However the way that NSO's respond to issues is determined by the legislative and political environment within which they operate. The most effective way for researchers to be exposed to these issues is to gain experience with real data sets and, in the case of official statistical data sets, to get input and assistance from NSO staff on how they have dealt with design and collection matters.

6. SIZE AND STRUCTURE OF STATISTICAL DATA SETS

As discussed earlier, one technique for reducing the risk of disclosure in a unit record data set is to provide only the subset of variables that the researcher needs to undertake their research. Additionally limiting the size of the data set is valuable on purely practical grounds. Some of SNZ's data sets are very large and complex; e.g. the population census has several million records and about 200 variables once all the derived variables are taken into account. There are often variables that superficially may appear to be the same (for example, Mäori ethnicity and Mäori ancestry) but are, in fact, quite different and valid for use in different circumstances.

Some of the data sets that SNZ has made available to researchers have been as large as one gigabyte. By the time they derive additional variables and do various sorts the data sets become unwieldy and time-consuming to handle. Jolliffe (2001) contends that problems with today's computing power computational problems are less of a problem than in the past. However our experience has been that the size of statistical data sets is still much larger than many researchers are used to and, by approaching their analyses in an inappropriate way, they can still adversely affect the performance of a large organisations computer system, as SNZ has experienced. The guidelines we develop for our own staff in efficient ways to use system resources have been provided to Data Laboratory researchers, with hints on how to sort efficiently and advice on suitable commands to use.

To be efficient a researcher should also be experienced with their analysis software. Sometimes, however use of software familiar to the researcher is not possible. Some software packages have limitations, for instance on the number of records that can be stored. As well as coming to grips with the data, a researcher may also be faced with using new software that may not have the options that they are familiar with. This adds to the demands on a researchers time, as well as their skills and knowledge. Jolliffe (2001) suggests that courses should ensure that researchers understand the principles of using software rather than the specific commands in whatever package is used for the training.

There are implications for staff in an agency who are called on to provide advice to researchers. If the researcher is using software that staff don't know they will be less

likely to be able to provide useful tips to improve the performance or advice on whether the output is successfully producing expected results.

The way data sets are stored (for example, in a hierarchical structure) or the processes used to manipulate the data sets also needs to be taken into account. These are all issues that the researcher needs to clarify at the start of the project. This means that the researcher cannot jump straight into the number crunching that is likely to be their most interesting aspect of the work. Practical experience with large data sets and the software typically used to manipulate and analyse them would be useful to researchers.

7. QUANTITATIVE SKILLS

While researchers usually have good theoretical research and analytical skills they do not always have strong programming or quantitative skills, i.e. they may know what the problem is and how to interpret the results but may not have the skills to produce the results by manipulating the data directly. In some cases they will employ a research assistant to provide programming services. However there is an opportunity for staff from the data collection agency who are likely to have strong quantitative skills, to collaborate with the researcher.

SNZ has been involved in several projects in this way. Successful collaborative research benefits both parties. A researcher learns more about the data and is more likely to acquire a clearer perspective on issues of data collection and processing. Collection agency employees are exposed to ‘real world’ uses of their data, which will benefit their own work. Indeed, collaborative research between the researcher with the theoretical expertise and the statistician with the data expertise may even result in a better research outcome than the researcher could achieve on their own. Svensson (1998) showed that participants in a training course on biostatistics found a multi-professional approach to be valuable.

SNZ has certainly found that its staff welcome the opportunity of working alongside experienced researchers. The researchers also appreciate the skills and knowledge that the staff member brings. This collaboration helps to strengthen the relationship between the NSO and the research sector and SNZ is actively encouraging such projects where it is appropriate.

8. OTHER SKILLS

Finally there are the skills that a researcher needs that are not specifically related to statistical research but which are an important part of successfully undertaking a research project. Some, such as communication, have been mentioned already and are addressed in other conference papers. For instance Jolliffe (2001) suggests improving researchers skills through requiring written and oral presentations as part of their courses. SNZ includes a mock interview with a ‘client’ as part of its in-house training on sample design for newly employed statisticians. Jolliffe also suggests that a researcher should know how to consult, i.e. ask relevant questions, and we have seen how important that is where the researcher is using a data set that already exists.

Other skills that researchers are likely to find useful are project management skills to ensure their efforts deliver results effectively. Glencross and Mji (2001) identify promotional skills as useful to ensure research results are widely disseminated, however

promotional, writing and negotiation skills also have a place at an early stage of a project if a researcher needs to secure funding to undertake their research.

9. RECOMMENDATIONS FOR TRAINING OF RESEARCHERS

The issues addressed in this paper have implications for the training and skills that social science researchers would find beneficial if they were considering undertaking research using official statistical micro data. However the conclusions are wider than just for data sets collected by NSO's and apply equally to other agencies' data sets.

It is recommended that university courses on quantitative research include a section on the use of existing data sets, in particular official statistics, and the pros and cons of micro data, e.g. using hierarchical data sets, dealing with missing values and non-response, with imputed and edited data, appropriate use of weights, analysis of complex sample design, and confidentiality control.

Practical experience with large and realistic and possibly less sanitised, statistical data sets during their training would be useful for researchers to develop an understanding of possible difficulties. Previous exposure to these issues would prevent them wasting valuable time whilst undertaking their research. It would useful for researchers to have the opportunity either to become familiar with software commonly used with large data sets or are trained in the principles of software applications rather than the specific commands of a particular package.

SNZ would support the benefits to be gained by NSO staff assisting with researcher training. It would expose NSO staff to the next generation of researchers, raise the awareness of the NSO and its data resources, and improve interaction between the research and government sectors.

There is scope for more co-operations between the sectors, for instance, in the form of the co-operative resource centres that Glencross and Mji suggest. However, researchers need to understand and be comfortable working in a co-operative research environment and to work collaboratively researchers need to be willing to recognise and understand interests other than their own.

Their training should expose them to the influences affecting government, business and the wider community and the need to balance research objectives against practical considerations of the rights of individual respondents to have their data appropriately protected and used. The resource difficulties that government agencies experience and the legislative environment that controls and constrains agencies are also relevant to what data can be made available and in what forms. SNZ has found it is particularly important that a researcher understands the particular position that it must take on some occasions, especially where that may be less favourable to the researcher.

Researchers need good communication and negotiating skills to be able to write a clear brief on their research, and to be able to defend their requirements while being willing to compromise. These are similar to skills required for writing proposals to obtain research funding.

As we have seen though the NSO can also learn from their interaction with the researcher. SNZ's practices in documentation and in data management are changing as a result of our relationship with researchers.

REFERENCES

Biggieri, L., & Zuliani, A. (1999). The dissemination of statistical literacy among citizens and public administrators. *Proceedings of 52nd Session of International Statistical Institute. Bulletin of the International Statistical Institute. Bulletin of the International Statistical Institute* (http://www.stat.fi/isi99/proceedings.html). Helsinki: International Statistical Institute.

Bishop, G., & Talbot, M. (2001). Statistical thinking for novice researchers in the biological sciences. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 215-226). Granada: International Association for Statistical Education and International Statistical Institute.

Chambers, R., & Skinner, C. J. (1998) Communicating sampling concepts to social scientists: The CASS experience. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics*. (pp. 259-266). Singapore: International Association for Statistical Education and International Statistical Institute.

Chant, D., & Lievesley, D. (1997). The use of data in teaching statistics. *Proceedings of the 51st Session of ISI. Bulletin of the International Statistical Institute* (Tome LVII, Book 1, pp 429-432). Istambul: International Statistical Institute.

Glencross, M., & Mji, A. (2001). The role of a research resource centre in the training of social science researchers. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 245-257). Granada: International Association for Statistical Education and International Statistical Institute.

Jolliffe, F. (1998) A course on sample surveys for statistics students. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics*. (Vol. 1, pp. 267-272). Singapore: International Association for Statistical Education and International Statistical Institute.

Jolliffe, F. (2001). Learning from experience. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 355-370). Granada: International Association for Statistical Education and International Statistical Institute.

Manly, B. F. J., & McDonald, L. L. (1998) Teaching sampling methods using ecological examples. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics*. (Vol. 1, pp. 273-278). Singapore: International Statistical Institute.

Svensson, E. (1998) Teaching biostatistics to clinical research groups. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics*. (Vol. 1, pp. 299-293). Singapore: International Association for Statistical Education. International Association for Statistical Education and International Statistical Institute.

Sandra McDonald
Statistics New Zealand,
Wellington, New Zealand
E-mail: sandra_mcdonald@stats.govt.nz

Batanero, C (Ed.), Training Researchers in the Use of Statistics, 129-137.

Printed in Granada, Spain.

TOSHIRO SHIMADA

PRECAUTION AGAINST ERRORS IN USING STOCHASTIC SOFTWARE

There are many statistics packages available that make it easy to perform stochastic procedures. Therefore, today's students may think they can handle their data processing needs, and obtain stochastic results simply by clicking a PC button. However, without being aware of it, they can make many mistakes, and treat their data incorrectly.

In this paper we compare generalised logistic curves with simple logistic curves and explain their characteristics to help students and researchers avoid mistakes.

1. INTRODUCTION

Many researchers are using information software which they have created themselves, and therefore they do not make mistakes in analysing their data. Moreover, there are many statistics packages available that help them to perform stochastic procedures although the standard options in these packages are not always the best for each specific research problem.

However, some researchers might believe they can solve their data analysis needs, and obtain statistics results simply by clicking a PC button. Without being aware of it, they can make mistakes, and analyse their data incorrectly, which may have serious consequences for their research.

2. EXAMPLES

Figure 1. TSE: Moving Average of Daily Volume on the Tokyo Stock Exchange (Sep.-Nov., 1956, Millions of Shares)

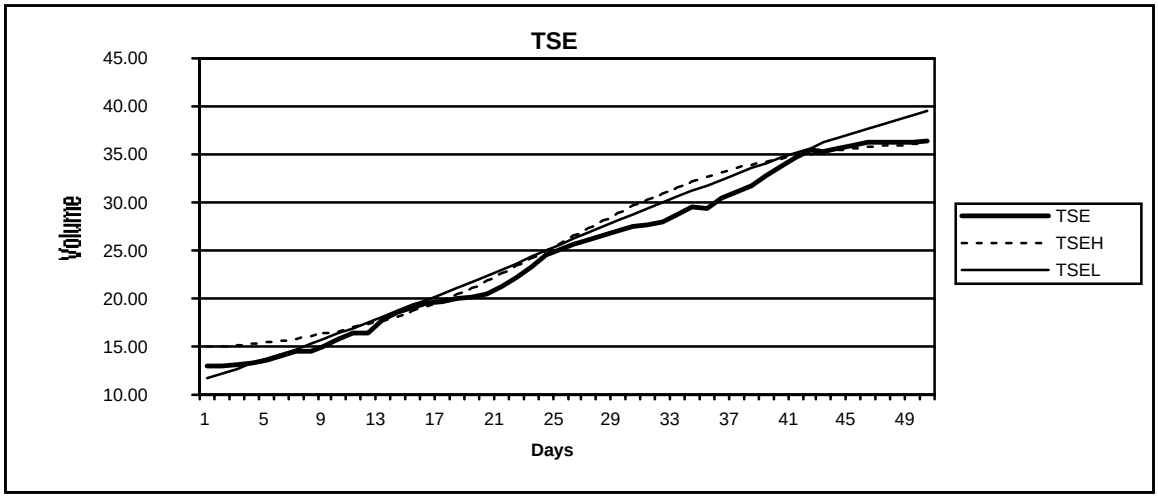
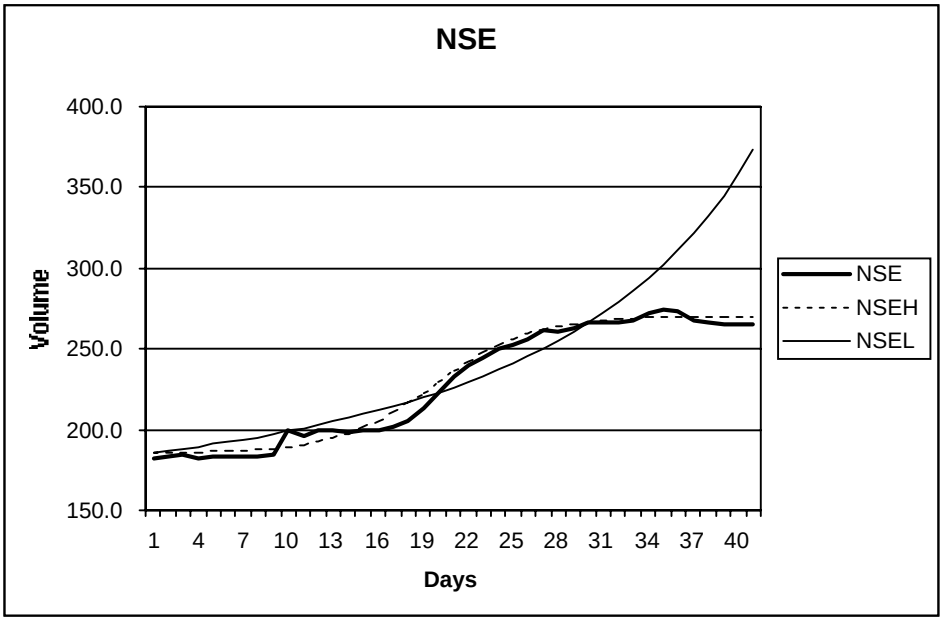


Figure 2: NSE: Moving Average of Daily Volume for the Dow Jones 30 Industrial Stocks of the New York Stock Exchange, (Dec. 26, 1956, Millions of Shares)



In this paper we will be discussing some examples of incorrect use of software in the particular case of logistic curves. Figures 1 and 2 present the two examples we are going to discuss.

3. GENERALIZED LOGISTIC FUNCTION

Graphs of the TSE and NSE data have apparent asymptotes which are different from the x axis. These are called generalised logistic curves. These curves are expressed by the following equation,

$$y = c + \frac{k}{1 + me^{-ax}} \tag{3.1}$$

The common logistic equation is:

$$y1 = \frac{k}{1 + me^{-ax}} \tag{3.2}$$

These equations are very simple, but both are non linear. Many teachers can easily handle *y1* type curves, but the (3.1) type curve is a little different and is rarely treated in stochastic packages. We will determine the four parameters c, m, a and k of (3.1) for TSE and NSE. Since (3.1) is non linear we will first try to linearise this equation. The expression

$$F(x, y; c, m, a, k) = y - c - \frac{k}{1 + me^{-ax}} = 0 \tag{3.3}$$

will represent the estimated curve.

We assume that c_0, m_0, a_0, and k_0 are the approximations of the parameters. (The method of determining them will be discussed later.) The differences between the estimated values and the assumed approximations of the parameters, so-called residuals, are represented as (3.4).

$$C = c_0 - c,\ M = m_0 - m,\ A = a_0 - a,\ K = k_0 - k \tag{3.4}$$

Let Y be the daily volume for n days, and y the estimated daily volume. Then (3.3) may be rewritten as follows:

$$\begin{aligned} &F(x_i, y_i; c, m, a, k) = y_i - c - \frac{k}{1 + me^{-axi}} = 0 \\ &(i = 1, 2, \ldots, n) \end{aligned} \tag{3.5}$$

The residual of y is given by (3.6).

$$Vy = Y_i - y_i \tag{3.6}$$

From (3.4) and (3.6) we obtain:

$$y_i = Y_i - Vy,\ c = c_0 - C,\ m = m_0 - M,\ a = a_0 - A,\ k = k_0 - K \tag{3.7}$$

When (3.7) is substituted into (3.5), the equation (3.5) becomes:

$$\begin{aligned} &F(x_i, Y_i - Vy, c_0 - C, m_0 - M, a_0 - A, k_0 - K) = 0 \\ &(i = 1, 2, \ldots, n) \end{aligned} \tag{3.8}$$

By expanding (3.8) using Taylor's series and by omitting the terms containing powers of the residuals higher than the second degree, we obtain:

$$\begin{aligned} &F_0 - VyFy - FcC - FmM - FaA - FkK = 0 \\ &(i = 1, 2, \ldots, n) \end{aligned} \tag{3.9}$$

where

$$F_0 = F(x_i, Y_i; c_0, m_0, a_0, k_0) \tag{3.10}$$

and Fy, Fc, etc, are the values of partial derivatives of the function F for the values (x_i,Y_i; c_0, m_0, a_0, k_0) (i=1,2,...,n), namely,

$$\begin{aligned} &Fy = 1,\ Fc = -1 \qquad && Fm = \frac{k_0}{(1 + m_0 e^{-a_0 x})^2} \\ &Fa = \frac{k_0 m_0 x e^{-a_0 x}}{1 + me^{-a_0 x}} \qquad && Fk = \frac{-1}{(1 + m_0 e^{-a_0 x})^2} \end{aligned} \tag{3.11}$$

Expressions (3.9) are the linearised equations, so we may solve the problem that the sum of the squares of the residuals is $\sum Vy_i^2$ to be a minimum under auxiliary conditions (3.9). Then we can find the normal equations for fitting a generalised logistic curve (3.1)[2]. They are

$$\left.\begin{aligned} &[cc]C + [cm]M + [ca]A + [ck]K = [co] \\ &[mc]C + [mm]M + [ma]A + [mk]K = [mo] \\ &[ac]C + [am]M + [aa]A + [ak]K = [ao] \\ &[kc]C + [km]m + [ka]A + [kk]K = [ko] \end{aligned}\right\} \tag{3.12}$$

where [cc]=∑FcFc, [cm]=[mc]=∑FcFm, [co]=∑FcFo, etc.

By solving the normal equations (3.12) and determining the values of C, M, A and K, we can estimate the parameters c, m, a, and k using (3.7). Then the problem of the trend line can be solved.

4. THE APPROXIMATIONS OF THE PARAMETERS

After examining Table 1 (given in the Appendix 2) and Fig. 1 we select $c_0 = -12.89$, and write

$$\eta = Y - c_0 = \frac{k_0}{1 + m_0 e^{-a_0 x}} \tag{4.1}$$

Then we obtain a_0, m_0, and k_0, which will satisfy (3.1) for the three values of x (10, 25, and 40) as follows:

$$\left.\begin{aligned} 1 + m_0 e^{-10 a_0} &= \frac{k_0}{\eta_{10}} \\ 1 + m_0 e^{-25 a_0} &= \frac{k_0}{\eta_{25}} \\ 1 + m_0 e^{-40 a_0} &= \frac{k_0}{\eta_{40}} \end{aligned}\right\} \tag{4.2}$$

These equations seem complicated, but if we choose the above mentioned values of x, (4.2) can be solved without difficulty if we write

$$z = e^{-5a_0}$$

then

$$z = \left(\frac{\eta_{10}(\eta_{40} - \eta_{25})}{\eta_{40}(\eta_{25} - \eta_{10})} \right)^{1/3} = 0.5411 \tag{4.3}$$

And hence

$$a_0 = 0.1228$$

Using (4.2), the value of m_0 and k_0 can be easily obtained. Thus,

$$c_0 = 12.89,\ m_0 = 21.10,\ a_0 = 0.1228,\ k_0 = 25.13 \tag{4.4}$$

Using these values, we can find the normal equations (3.12). By solving these we get residuals,

$$\text{C=-1.369, M=-15.24, A=-0.02319, K=2.704} \tag{4.5}$$

From (3.7), (4.4) and (4.5),

$$\text{c=14.26, m=36.34, a=0.1460, k=22.43} \tag{4.6}$$

and hence,

$$TSEH = 14.26 + \frac{22.43}{1 + 36.34e^{-0.1460x}} \tag{4.7}$$

(Sep. 21, 1956; x units: 1 day; y is daily volume in millions of shares.)

This curve (4.7) is shown in Fig. 1 as TSEH.

Many years ago I used FORTRAN and a simultaneous equations package, and easily obtained the above results. My students also understood this procedure without problems.

The generalised logistic curve of NSE shown in Fig. 2 was obtained by the same procedure as that used for TSE. In this case, the approximations of the parameters were

$$c_0 = 183.5,\ m_0 = 362.6,\ a_0 = 0.3085,\ k_0 = 87.96 \tag{4.8}$$

Residuals are

$$\text{C=-1.93, M=-58.7, A=0.00278, K=2.866} \tag{4.9}$$

Hence the estimation of NSE becomes

$$NSEH = 185.4 + \frac{85.09}{1 + 362.6e^{-0.3035x}} \tag{4.10}$$

5. COMMON LOGISTIC CURVES

TSEL In Fig. 1 and NSEL in Fig. 2 are the common logistic curves of TSE and NSE. Their equations are as follows:

$$TSEL = \frac{50.07}{1 + 3.420e^{-0.05095x}} \tag{4.11}$$

$$NSEL = \frac{151.46}{1 - 0.1784 e^{-0.02937x}} \tag{4.12}$$

These parameters are rough values obtained from the procedures as the approximations of the generalised logistic curves. We can see good fitting of TSEH and NSEH to the original data, but TSEL and NSEL are unsatisfactory. In particular, NSE has high values, so after x=30, NSEH goes up very high.

6. CONCLUSION

In this paper we compared generalised logistic curves with simple ones and showed their characteristics to help students to avoid mistakes. I once heard a research report at a yearly meeting of an academic society, given by a researcher from a large Japanese computer maker. He expected the Christmas sales trend to follow a common logistic curve, but the data were clearly of a generalised logistic curve. I asked him about this, and he answered that he did not know the details of the software he was using, but he simply accepted the results.

We must be very careful when using statistical software and teach basic principles to beginners to avoid these kinds of mistakes.

APPENDIX 1.

Table 1. Sample Data Used for Figures 1 & 2

x	TSE	TSEH	TSEL	NSE	NSEH	NSEL
1	13.02	14.95	11.78	182.2	185.6	185.5
2	12.98	15.05	12.24	183.9	185.7	186.8
3	13.07	15.17	12.72	184.3	185.9	188.1
4	13.24	15.31	13.21	182.0	186.0	189.4
5	13.53	15.47	13.71	183.9	186.3	190.9
6	14.02	15.65	14.22	183.0	186.6	192.3
7	14.60	15.85	14.75	183.9	187.0	193.9
8	14.59	16.08	15.28	183.4	187.6	195.5
9	14.94	16.34	15.83	184.3	188.4	197.3
10	15.76	16.63	16.39	199.1	189.4	199.1
11	16.39	16.96	16.95	196.1	190.8	200.9
12	16.37	17.33	17.53	199.3	192.6	202.9
13	17.76	17.73	18.11	199.5	194.9	205.0
14	18.52	18.19	18.71	198.1	197.8	207.2
15	19.12	18.68	19.31	199.5	201.4	209.5
16	19.52	19.22	19.92	199.9	205.8	211.9
17	19.66	19.81	20.53	201.3	210.9	214.5
18	19.94	20.44	21.13	205.8	216.7	217.2
19	20.17	21.12	21.77	213.4	223.0	220.0
20	20.54	21.83	22.40	223.0	229.4	223.0

21	21.20	22.58	23.04	232.8	235.9	226.2
22	22.19	23.36	23.67	239.8	241.9	229.6
23	23.28	24.16	24.31	245.0	247.4	233.2
24	24.46	24.97	24.94	249.8	252.2	237.0
25	25.17	25.79	25.58	252.4	256.2	241.0
26	25.58	26.60	26.22	255.5	259.4	245.4
27	26.08	27.41	26.85	261.3	262.1	250.0
28	26.52	28.19	27.49	261.0	264.1	254.9
29	27.00	28.95	28.12	262.8	265.7	260.2
30	27.47	29.67	28.74	265.8	266.9	265.9
31	27.66	30.35	29.36	266.2	267.8	272.1
32	28.00	30.99	29.98	266.6	268.5	278.7
33	28.69	31.59	30.59	268.0	269.0	285.9
34	29.46	32.14	31.19	272.4	269.4	293.6
35	29.44	32.65	31.79	275.0	269.6	302.1
36	30.49	33.11	32.38	273.5	269.9	311.3
37	31.03	33.53	32.95	267.2	270.0	321.4
38	31.74	33.90	33.52	266.1	270.1	332.6
39	32.80	34.24	34.08	265.2	270.2	344.8
40	33.78	34.54	34.63	264.9	270.3	358.5
41	34.62	34.81	35.17	265.2	270.3	373.7
42	35.40	35.04	35.70			
43	35.35	35.25	36.21			
44	35.58	35.44	36.72			
45	35.91	35.60	37.21			
46	36.21	35.74	37.69			
47	36.30	35.86	38.16			
48	36.31	35.97	38.62			
49	36.25	36.07	39.04			
50	36.33	36.15	39.49			

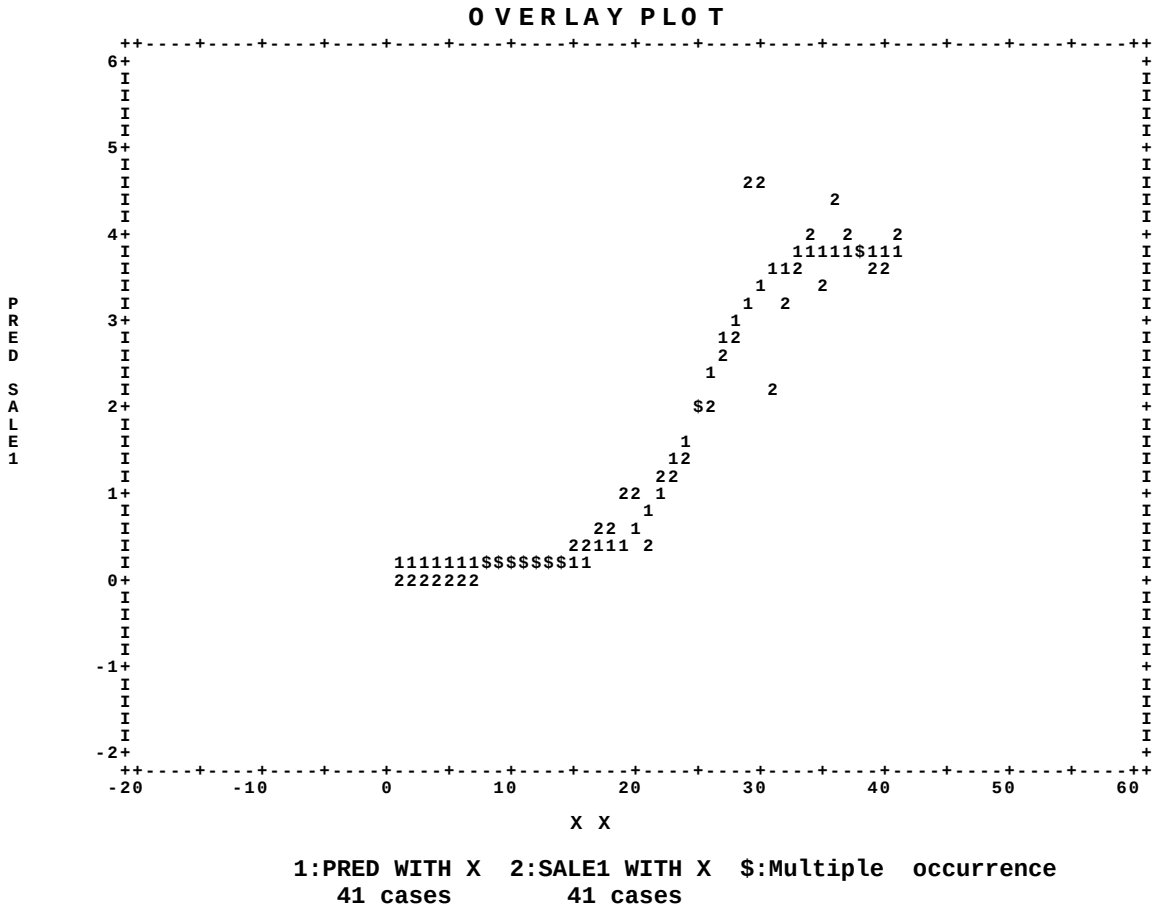
APPENDIX 2.

Another example of the generalised logistic equation uses summer sales of a Japanese department store treated with SPSS version 4. This simple procedure is similar to those of TSE and NSE.

```
//BB00001S JOB (SIMADA,200),'SIMADA',CLASS=S,MSGCLASS=1
// EXEC SPSSX
//SYSIN DD *
RUN NAME       SUMMER SALES (NONLINEAR REGRESSION)
DATA LIST FREE/ X SALE1
BEGIN DATA
 1 .0610  2 .0650  3 .0740  4 .0840  5 .0920
 6 .0690  7 .0270  8 .1740  9 .1690 10 .1950
11 .2190 12 .2510 13 .2640 14 .1450 15 .4860
16 .4960 17 .5480 18 .5500 19 .9570 20 .9620
```

```
21 .3280 22 1.189 23 1.203 24 1.421 25 2.096
26 2.037 27 2.538 28 2.826 29 4.560 30 4.560
31 2.265 32 3.109 33 3.609 34 4.055 35 3.362
36 4.471 37 3.914 38 3.778 39 3.583 40 3.572
41 3.947
END DATA
MODEL PROGRAM  C=0.06 K=113.05 M=681.1 A=0.1960
COMPUTE        PRED=C+K/(1+M*EXP(-A*X))
DERIVATIVES
COMPUTE        D.C=1
COMPUTE        D.M=-K*EXP(-A*X)/(1+M*EXP(-A*X))**2
COMPUTE        D.A=K*M*X*EXP(-A*X)/(1+M*EXP(-A*X))**2
COMPUTE        D.K=1/(1+M*EXP(-A*X))
NLR           SALE1 WITH X/ SAVE PRED
PLOT FORMAT=OVERLAY / PLOT=PRED WITH X; SALE1 WITH X
FINISH
```

Figure 3. Summer Sales (Nonlinear Regression)



The first three steps are Job instructions which will be different for each computer system. The next steps comprise the SPSS data and instructions program for the generalised logistic equation. The variables are x and sales, which are sales indices of daily summer sales for 41 days.

The MODEL PROGRAM line gives the initial parameter values, which I assumed.

The first COMPUTE line shows the generalised logistic equation.

NLR means the NONLINEAR REGRESSION program of SPSS.

The Results of the Summer Sales Regression.

Iteration was stopped after 51 model evaluations. The parameter estimates were c=.2004, k=3.634, m=29800, a=.4127. Fig. 3 shows the original data and the regression results for the summer sales.

REFERENCES

Biehler, R (1997). Software for learning and for doing statistics, *International Statistical Review*, 65, 2, 167-189.

Shimada, T. (1961). Short term trend lines of daily volume of trading in stocks. *Journal of Commercial College of Meiji University*, 44(7), 1-21.

SOFTWARE

(1) SAS, SAS Institute Inc., Cary, NC, USA
(2) S and Splus, MathSoft, Inc.
(3) SPSS, SPSSX User's Guide, Second Edition, 1985, SPSS Inc.,

Toshiro Shimada
Professor Emeritus Meiji University,
8-9-3-314, Kokuryo-machi Chofu-shi,
Tokyo, 182-0022, Japan.
E-mail: simfuji0@mtj.biglobe.ne.jp

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 139-145.

Printed in Granada, Spain.

TAE RIM LEE

TEACHING BIOSTATISTICS TO MEDICAL PERSONNEL WITH COMPUTER BASED SUPPLEMENT

It has become increasingly evident that the interpretation of much of the medical research in health sciences depends to a large extent on statistical methods. For this reason, it is essential that students in these fields be exposed to statistical reasoning, data analysis, and computation early in their careers. With the current advances in the methodology for data analysis we can take advantage of the highly developed computer software. Nowadays the internet provides rapid feedback to the student and dissemination of the latest research information. This paper focuses on the teaching of biostatistics to health professionals with computer-based supplements; courseware, digital library of TV lectures on SAS and interactive communication through the Internet.

1. INTRODUCTION

There is an increasing need for health-care professionals to use statistics in their master's thesis or other research, and nowadays statistical analysis can be easily performed using a statistical package (Binyavanga, 1998). Many medical journals require a high level of statistical sophistication from their authors (Altman,1991, 1994; Hand, 1994, 1996). It has, furthermore, become evident that the analysis of much of the research in the health sciences depends on highly advanced statistical methods.

These facts increase the requirement for biostatistical understanding among health professionals in order to choose the appropriate statistical method and to interpret the results produced by the computer (Korea Biometric Society, 1994; Svensson, 1998). Use of computers is encouraged in teaching to allows the student to concentrate on the interpretation of the analysis rather than on arithmetic calculations. (Phillips, Francis, & Hutcheson 1998; Beur, Richards, & Lancaster,1987). Teachers can pay less attention to students' mathematical calculations by using statistical packages. Algebraic notations in lectures and textbooks has also been simplified to reduce the mathematical skill needed from high school graduates in a statistics course.

On the other hand, most major medical research projects involve a tremendous investment in time and money and result in a large body of data, that need to be analysed with computers. For this reason, in my courses, I allocate 50% of the lecture time for computer practice with SAS and Excel, and students are introduced at the same time to SAS with television lectures that serve to demonstrate computer work and the interpretation of output (Johnson, Johnson, & Stanne, 1985).

In Korea, statistics courses using more flexible modes and newer technologies such as the World-wide web, CD and cable television programs are being offered and these recent developments in technology can potentially lead to great improvement in teaching statistics. With the boom of the Internet, a computer-based learning of statistics has attracted particular attention from both students and researchers.

This paper outlines a flexible method of teaching biostatistics based on various supplementary materials such as courseware diskettes, a CD introduction to SAS and a set of web-based lectures.

2.TEACHING OUTLINES

I have taught biostatistics to undergraduate medical students as an optional topic and to graduate students as a compulsory requirement in their curriculum. In addition I have tried to broaden the scope of this course to attract students who have enough time to work on computer skills.

The main strategy of this course is to focus on statistical applications rather than on statistics itself. Therefore the students taking this biostatistics course are encouraged to apply the appropriate statistical concept to their own papers and to have frequent presentation and discussion during the semester. During these sessions all the students are brainstorming.

Course material

The text of this course is on video tape and delivered by cable television. There are two diskettes and a CD ROM which contains the contents of digital libraries on the w.w.w. as well.. With this courseware students can participate interactively as communication and feedback are an essential part of the learning process for many students. Students submit reports through e-mail and receive individual comments on the evaluation forms followed by a presentation in the next class.

Computer exercise

All chapters covered by this course have computer exercises using SAS or simple statistical software attached to the courseware diskette. All the results are presented and discussed in the class.

Presentation of proposals

In the graduate course by the 3rd quarter, all the students have prepared a proposal for their master doctorate thesis, covered research problems, outlined the purpose of their papers, identified their data collection methods and reviewed suitable statistical analysis methods. All the students have had the opportunity to apply statistical methods to various data sets in the health field.

Evaluation of the courses

This flexible course is offered in the nursing department at the master's level and at the medical college for undergraduates. With the aim of evaluating the course, a survey was taken from the University academic staff. Based on that data, we compared the attitude and performance of the students from ordinary courses and students following this flexible method.

The course syllabus

The following subjects were included in the courses schedule. The two final subjects were excluded in the undergraduate course and are as follows:

1. What is biostatistics;
2. Graphical and numerical summaries of data;

3. Point estimation and interval estimation;
4. Introduction of testing hypothesis;
5. One sample test and two sample test;
6. Categorical data analysis;
7. Midterm exam;
8. ANOVA;
9. Correlation and regression;
10. Presentation of proposals;
11. Non parametric statistics;
12. Multivariate analysis;
13. Survival analysis;
14. Final exam.

3. COMPUTER BASED SUPPLEMENTS

Nursing students working on their masters or doctorate degrees are strongly motivated to use statistics but do not have a strong mathematics or computer background. In contrast, the undergraduate students in their second year of medical college have little motivation and need for biostatistics but a strong motivation to use the computer. These student skills and attitudes suggested us to use the computer based supplementary materials described below.

3.1. COURSEWARE DISKETTE

In recent years there has been widespread expectation that the new generation of instructional computers will solve educational problems and help us to better achieve educational goals. For many educators, computer based education has become an instrument for modern education. Based on programmed instruction, courseware of statistics proves to be an effective media for learning concept principles and techniques. They were developed according to programmed instruction theory and compensated for the disadvantages of older computer devices and books.

Students received instructions from the computer using courseware and then responded to the computer individually. Using their individual pace of learning, the students achieved the fundamental statistical concepts. Computer Based Learning (CBL) resulted in high achievement with less time spent in school, thus increasing the interest in subjects like mathematics and statistics. Statistics particularly covers the data assessment methods, therefore high achievements in learning can be expected.

Statistics seems to be particularly suitable for illustrating the benefits of multimedia-based teaching. On the other hand, statistics connects quite different fields of application. This interdisciplinary character of the science can be well demonstrated by suitable videos and motivating examples closely related to medical data. Courseware of statistics could present an ideal platform for learning statistical concepts, and for discovering basic statistical principles by self-driven experiments.

The direct handling of computers results in a fundamental knowledge of computers for the students. This statistics courseware also contain a program for statistical package practice. The topics included in the courseware are as follows:

1. Basic descriptive statistics;

2. Probability and distributions;
3. Estimation and testing;
4. Correlation and regression;
5. Experimental design.

Each chapter has a self evaluative module through which students can check their performance and receive feedback for each item according to their results.

There was a final evaluation of CBL, and it was done by comparing the students' improvement in performance (experimental group) with a control group. Before and after the CBL coursework was completed, students in both groups were given a test on each chapter.

The improvement in performance between pre-test and post-test is shown in Table 1. The paired t-test based on Table 1, showed CBL was statistically significantly ($p<0.01$). There was also a statistically significant difference according to age and sex but no statistically significant difference by major ($p>0.05$).

Table 1. Performance for Computer Based Learning Experimental Group in Pre-test and Post-test

Variable	Group	Frequency (%)	Mean score in pre-test	Mean score in post-test
Age	Under 30	24(42.10)	52.14 ± 20.45	69.29 ± 28.14
	31-40	22(38.60)	55.00 ± 7.07	60.00 ± 4.14
	above 41	11(19.30)	40.08 ± 2.04	51.00 ± 6.50
Sex	Male	26(45.61)	46.00 ± 8.97	53.00 ± 7.51
	Female	31(54.39)	58.33 ± 7.22	80.35 ± 9.97
Major	Computer	34(59.65)	51.43 ± 9.95	65.71 ± 8.48
	Administration	23(40.35)	45.00 ± 7.07	49.00 ± 11.27
Total		57(100.0)	51.80 ± 8.80	64.70 ± 7.90

The effective index, derived by transforming the score into a standard score is 1.389. Thus the mean score of the results of the CBL schooling was located at a 138.9% when results without CBL are considered to be at 100%.

The correlation coefficient between pre-test and post-test is 0.783. This means that students who had better initial knowledge got a higher score in the post-test. The fitted regression line of the pre-test score and post-test score is $Y = 4.583 + 1.16X$. The statistical test of $\hat{\beta}_1 = 1$ in which there was no effect of CAI was statistically significant ($p>0.01$).

In the graduate class, the students had different ages and a range of mathematical and statistical background. This meant that they needed an individual learning method based on their prior knowledge, which can be easily provided with this kind of computer based learning program.

3.2. DIGITAL LIBRARY OF TV LECTURE FOR SAS ON THE WEB.

In Korea, the government encouraged collaboration among universities and the private sector, and the sharing of existing resources to provide web-based instruction to university students and adults. With the Korean government's support and funding, all

formal higher education institutions are now connected to the education and research network and have computer laboratories (Jung, Chai, & Chai, 1997). Many digital libraries have been established and linked. In 1998, the government started the Virtual University Trial Project in which 25% of higher education institutions in Korea and several private companies used advanced education for university students and working adults.

The Virtual University Trial Project has inspired about 25 percent of the formal higher education institutions and five private companies in Korea to collaborate in providing virtual courses using advanced technology and to explore the possibility of incorporating distance education into the computer-based system and even of establishing a distance technology in the near future. Interactive technologies seem to provide students with opportunities to receive learning support from the instructor. PC network and internet along with a telephone have been used as a formal channel for students to ask questions to their instructors and to interact with other students on the web, to have small group interaction among learners, and have access to a relevant information library (Jung, Chai, & Chai, 1999).

It is very important to note that KNOU (Korea National Open University) established a multimedia digital library system on the Web in 1997. This new initiative urged KNOU to digitise a TV, radio and audiocassette program. These digitised programs were integrated in a certain teaching and learning platform so that the students studied their course of SAS in a multimedia format on the web.

Now this digital library system is free of change to everybody. It can be said that we began to operate video-on-demand, not only for our students, but for the general public in Korea. It can be located at (http://kcsone.knou.ac.kr/frmstudy.htm). The information technology situation is getting better in Korea. In the near future every home will have access to the information superhighway. Therefore, the digital library system or Internet courseware is going to be more actively utilised by students. Multimedia software for statistics can go beyond closed instructional ms-word by offering properly maintained subject-specific gateways to recent statistical data and supplementary information from the rapidly growing internet. These elements are interactive and help the learner with activating simple moves and clicking in some documents.

There is a TV lecture demonstrating statistical analysis of data using SAS. Each chapter has a self-evaluation corner and Q&A corner. The digital library supplied above is operated through the world-wide web. The learning contents are laid down in HTML format. An internet browser gives access to the usual communication functions of a virtual educational network, for instance, to e-mail communication and course related lecture notes. Students with internet access are free to switch to a predefined point in the online course to download the latest lecture or additional information relating to the online lecture from the world-wide web.

Nowadays due to computer availability, there is much research and development done to provide a virtual class. The virtual class is defined as an institution that provides access to its educational services without the need for students to be physically present to receive them (Hiltz, 1995). A communication technology such as the internet system is needed as a main instruction and communication tool (Beur, Richards, & Lancaster, 1987). The motivation for a virtual statistics class is as follows:

1. Variety of educational demand;
2. Change of paradigm for statistical education;
3. Increase of access to the world wide web;

4. Enlargement of educational space;
5. Various levels of teaching-learning system.

The virtual statistics class started in universities in 1998. In establishing the trial project, the government sought to encourage partnerships among universities and the private sector and the sharing of existing resources (Cheng, Lehman, & Reynolds, 1991).

The limitation of web-based statistics education is as follows. From the instructors' point of view, it is difficult to express statistical equations on the web window, and to explain the process of proofs or extensions of equations. Consequently, a virtual statistics class would not be suitable as a graduate course for a statistics major. Also it was very difficult to quantify time needed for 1 credit of learning materials. On the students' side it is costly to access as an online course, and they feel a lack of interaction and personal touch (Hiltz, 1995).

Another survey was done for the evaluation of the overall effects of web-based instruction and to identify important factors to consider in designing effective web-based courses. The major results of the study were summarised as follows:

1. Compared with conventional courses which use textbook and broadcasting programs, web-based courses showed higher course completion and higher student performance. That is, about 75% of the students who took the web-based courses completed their courses and about 85% scored 80 or above out of 100.
2. It appeared that more than 70% of the students were satisfied with the various aspects of the web-based courses. Students showed a high level of satisfaction with the technical and instructional support from the instruction and the online assistants, course design strategies and active online discussion.
3. It was found that content design strategy was the most important factor affecting student satisfaction ($r=0.509$). In addition, students who studied web-based courses in a LAN environment showed more satisfaction than those in a modem environment ($p<0.05$).
4. In general, physical access to the environment, content design strategies and online activities seemed to be the most important factors affecting the effectiveness of web based instruction. One can also share a student's proposal.

5. REMARKS

All the students in the graduate course were interested in the computer based course material and resulted in good performances. They felt motivated by the course materials and expressed that they had a better understanding of reality, and the feedback through e-mail encouraged the students. Medical personnel enjoyed learning with various kinds of computer-based supplementary materials, and video lecture faced statistical challenges. With regard to the statistical issues related to their master's thesis, students were strongly motivated and worked very hard to successfully master the computer skills.

REFERENCES

Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman.
Altman, D. G.(1994). The scandal of poor medical research. *British Medical Journal*, 308, 283-

284.
Beur, G., Richards, H., & Lancaster, P. (1987). Attitude forward computers: Validation of a computer attitude scale. *Journal of Computing Research*, 31, 207-218.
Binyavanga, K. (1998). Statistics education in South African medical training institutions. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 551-556). Singapore: International Association for Statistical Education and International Statistics Institute.
Cheng, H., Lehman, J., & Reynolds, A. (1991). What do we know about asynchronous group computer based distance learning? *Educational Technology*, 31, 16-19.
Svensson, E. (1998). Teaching biostatistics to clinical research groups. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics*. (pp. 289-294). Singapore: International Association for Statistical Education and International Statistics Institute.
Hand, D. J. (1994). Deconstructing statistical questions. *Journal of the Royal Statistical Society Series A*, 157, 17-56.
Hand, D. J. (1996). Statistics and the theory of measurement. *Journal of the Royal Statistical Society Series A*, 159, 445-410.
Hiltz, S. R. (1995). *The virtual classroom*, Norwood, NJ: Ablex.
Johnson, D. W., Johnson, R. T., & Stanne, M. B. (1985). Effects of cooperative and individualistic goal structure on computer-assisted instruction. *Journal of Educational Psychology*, 77(6), 668-677.
Jung, I. S., Chai. S. W., & Chai. S. H. (1997). *Analysis of educational effectiveness using online network.* Seoul: Technical Report KNOU Institute.
Jung, I. S., Lee J. H., & Chai, J. K. (1999). An *evaluation study on the effects of web based instruction*. Seoul: Technical Report KNOU Institute.
Korea Biometric Society (1994). *Application of statistics in medicine and public health workshop.* Seoul: Korea Biometric Society.
Phillips, B., Francis, G., & Hutcheson, M. (1998). Statistics for workers in social and health sciences by flexible learning modes. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 401-408). Singapore: International Association for Statistical Education and International Statistics Institute.

Tae Rim Lee,
Korea National Open University,
Department of Applied Statistics,
169 Dongsung dong, Jongro Ku,
Seoul 110-791 Korea.
E-mail: trlee@av9500.knou.ac.kr

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 147-157.

Printed in Granada, Spain.

DALENE STANGL

DESIGN OF AN INTERNET COURSE FOR TRAINING MEDICAL RESEARCHERS IN BAYESIAN STATISTICAL METHODS

Access to statistical information is at an all-time high, and the information age is fuelling this access at an extraordinary pace. This access increases the capacity for medical researchers to use statistics to guide decision making, yet few courses teach methods to do so. Rarely does statistics training include methods for incorporating statistical output into decision making. Mass education and educational reform is needed. Technological advances of the past decade make this goal possible, and allow us to dramatically change how we use, teach, and think about statistics. This paper covers the conceptual development of an Internet continuing-education course designed to teach the basics the Bayesian statistics to medical researchers. Two questions are discussed: Why the Internet, and why the Bayesian paradigm?

1. INTRODUCTION

Access to statistical information is at an all-time high, and the information age is fuelling this access at an extraordinary pace. Yet, the statistical literacy of the population, even in medical schools, is surprisingly low. A random selection of persons passing through the doors of a medical library would result in few persons that correctly define a significance level, and even the most respected medical journals have statistical mistakes and misinterpretations. And, although the importance of statistics to guide decision making is well recognised, few courses teach methods for incorporating statistical output into decision making.

These observations signal that mass education and educational reform are needed. Technological advances of the past decade make both goals possible, as the Internet increases accessibility to education while keeping costs low and computational advances make Bayesian decision-theoretic analysis possible. Together these advances can dramatically change how we think about, use, and teach statistics.

This paper covers the conceptual development of an Internet continuing-education course designed to teach the basics of the Bayesian statistical paradigm to medical researchers. Special attention is given to two questions: Why internet-based instruction, and why the Bayesian paradigm?

2. WHY INTERNET-BASED INSTRUCTION?

A reviewer recalls the introduction of several new technologies for teaching: filming great teachers, courses taught on television, VCR tapes, and CD discs. Each was

supposed to revolutionise teaching and maybe even make the teacher unnecessary. Now the Internet has arrived. The reviewer asks: "Is the Internet different from the previous attempts and in what ways may it succeed where other technological attempts have basically failed?"

The chances that the Internet will succeed are greater than other attempts due to four interrelated reasons: Tempo of change, interactivity, access and flexibility.

> *"Not since the mainframe era has technology reached a level of relative stability before other discontinuous architectures entered the Instructional Technology (IT) environment. ..., IT planners barely have time to come to terms … before the next 'paradigm' arrives."* (Gartner Group, 1998 c, p. 3)

The speed with which changes are coming prevents IT designers from developing a design process model by which they strategize, prototype, test, develop, and market a product. Products that cannot be updated and adapted quickly barely appear before they are abandoned. Because products offered via the Internet are adaptable rather than static objects, this technology is likely to survive where others have not.

The Internet also has other features that previous technologies lacked, the two most important being that rather than aiming to replace the teacher and textbook it improves upon both. The Internet can make the teacher more accessible and the textbook interactive.

The Internet allows people to come together across great distances without narrow temporal constraints. Via discussion and chat groups, students and faculty can interact across thousands of miles. Continuing education's traditional barriers of geographic proximity and job/family-compatible scheduling will disappear.

This communication medium will enable researchers to seek continuing education at distant universities with timing restricted to only their own availability. Students can work through courses at their own pace, in a convenient location, on their own time schedule. The Internet can make education more accessible, and more individualised.

The Internet can provide much more than a textbook on-line. (Puranen, 1998; Talbot et al., 1998) and multimedia software can provide interactive instruction (Cumming & Thomason, 1998; DiCiaccio, 1998; Giusti et al, 1998; MacRae 1998):

> *"Interactivity is the degree to which the user creates an individual experience with an application or body of content, ranging from "user as passive TV viewer" to "user as author."*
>
> *The critical differentiator in categorising "how interactive" an application is revolves around the degree to which the application can dynamically adapt to or change the user's needs in real time.*
>
> *Interactivity is often incorrectly correlated with media richness. Interactivity has two drawbacks: 1) The complexity and expense in content creation, delivery and the devises to experience it; and 2) the human-factor aspects of interactivity, i.e., how much a given audience can take, or how much time or inclination they have to embrace it. Interactive advertisements on TV sound great in principle, but with 30 seconds to complete the experience, most people will be overwhelmed"* (Gartner Group, 1998 a, p. 4).

While interactivity has also been available via CD ROM technology (Watson, 1998) and other computer based resources (Morris & Le Voi, 1998), products offered via the Internet can be adapted and updated with much less hassle. Authors have greater flexibility and control over changes to their work. The Internet also provides the opportunity to link with related sites and can potentially make education more interactive and more flexible.

3. CURRENT STATE OF INTERNET BASED STATISTICS EDUCATION

The current state of internet-based education suffers from the 'textbook on a computer' syndrome. This means that sites look like a poorly formatted textbook, rather than being coherently mapped, layered and linked. Sites vary greatly in their use of multimedia, an advantage that the Internet offers over textbooks.

There are many interactive simulation demonstrations (see Galmacci, 2001), but rather few that focus on strong case-style applications. Many sites are overwhelming in detail, making the reader want to turn and run and most are beginning bits and pieces rather than a complete work. Few keep their links updated, and few are far enough along in design to foster conceptual learning. Most of us are still learning the tools, tools that change by the minute and we are still realising that publishing on the Web is very different from traditional publication methods (Korpela, 1997). Our Internet products are not yet of professional quality, but this is changing very quickly as universities allocate resources to instructional technology and educators' team up with graphic artists and professional designers.

4. DESIGNING AN INTERNET COURSE

According to Gartner Group (1998, b) to capitalise on the advantages of the Internet requires processes and technologies that make it possible to highly automate content creation, assembly, and production functions. This requires changes and poses challenges to architecture/technologies, products, and organisational procedures. An instructional architecture must be defined and integrated with the enterprise at large. The cut-and-paste tools on the desktop must be replaced with content-aware intelligent products that can process all information resources, not just document elements.

Designing an Internet course requires several steps:

1. Software decisions;
2. Slide development that simulates teacher/student interaction;
3. Script development;
4. Testing;
5. Implementation;
6. Updating.

Deciding on software includes answering questions such as:

- Which software for slide production?
- What type of audio/video will be used?
- What quality is required?
- Will access be via modem or high-speed connection?
- What plug-ins are required?

In deciding upon software, one must remember that resources must be provided at the lowest common denominator among current users. Compromise between using 'state-of-the-art' and 'state-of-the-masses' is requires. While video clips are now easy to incorporate into presentations, if it takes the student two hours to download, they won't use it. You will only reach students whose computing resources easily support everything you use.

Although the public's resources and skills are changing rapidly, make sure that students have the hardware and software to support your product and include training for how to use it. Use multimedia sparingly and keep files small. Make sure multi-media clips enhance comprehension rather than replace something that is more effective.

Test clips on different computers and use compression. Assume the user has a system one-quarter the size and speed of your own. Make sure things are user friendly. Always have a backup-plan if a student has problems. It is naïve to think glitches won't be common.

In developing slides one must think continually about content, structure, and delivery. It requires a sequential process of *Strategize, Prototype, Test, Develop, Test,* and *Enable,* that will ensure projects meet users' needs and are sustainable over the long run (Gartner Group, 1998 b).

Systematically define the scope of the content and the units in which information will be developed. Keep units short and modular. One advantage of the Internet is that material presentation need not be linear. Yet the structure, the mapping, layering and linking of material will be overwhelming if not meticulously planned.

Develop a structure that determines how units will be mapped, layered and linked for easy navigation. Determine how aggregated information will be disseminated for efficient delivery. Make sure materials can be easily searched, retrieved and discovered. Remember different browsers display differently.

Careful attention to page layout, including sizes, colour, and positioning can make your product look great on your own monitor-browser combination and horrible on someone else's. Keep each slide simple and uncluttered. Keep slide organisation simple, systematic, and easy to navigate.

Be sure to abide by copyright laws. Xeroxed material for teaching in the classroom is not the same as putting copies on the Internet. Make sure you don't put things on the Internet that you do not own. Copyright laws do not allow you to put other's materials on the web without express permission. Also realise that making your own materials available on the web will make them quite easy for others to use, with or without your permission. Use password protection intelligently.

Making your presentation interesting requires a lively script. This is as important as including multimedia clips and interactive exercises. A well written and delivered script is more difficult than the slide content and organisation. You must develop the script in a way that gives the student a feeling you are in the room talking with them. Your speaking performance on the Internet will be compared against public television's best program narrators rather than your university's faculty lecturers. Listen to the pro's and

learn from them.

Once your presentation is ready for students, implementation hurdles must be jumped. These hurdles include making sure students have the software they need and know how to use it. Student access is the first and biggest hurdle:

> *"The availability and speed of network access remain a gating factor. The move to untethered (wireless) computing will allow users to obtain real-time information when and where they need it"* (Gartner Group, 1998 c, p. 7).

It will not be long until this gating factor is not an issue, but until then access issues must be addressed. Other hurdles include keeping links current, keeping up with student communications and setting clear communication boundaries. Throughout the implementation phase it is important to get as much feedback as possible about your product and to respond to students with the same level of receptiveness that you give in the classroom.

5. WHY THE BAYESIAN PARADIGM?

The statistical training of medical researchers usually begins with an introductory undergraduate statistics course, which is perhaps followed by a course or two in research and statistical methods while in nursing, graduate or medical school. Any other statistical education comes from 'on-the-job' training and continuing- education courses.

Laake (1998) and Phillips et al (1998) discuss teaching statistics to professionals in health-related sciences. Typically courses cover a mixture of statistical analysis and research design, while nearly all courses cover hypothesis testing, and few courses address the use of statistics directly in decision making.

The call for a stronger link between statistical analysis and decision making is not a new one, but work answering this call is a rarity. Twenty-five years ago leaders in our field were urging statisticians to take greater leadership in decision-making. Rice (1977) reported that, at the 1975 ASA meeting, Sir Claus Moser made this suggestion:

> *"Foremost responsibility (of the statistician) is to contribute to more enlightened and efficient 'decision making' … through the fullest possible exploitation of our skills in analysing and interpreting the data."* (Rice, 1977, p. 104)

In 1977, Dorothy P. Rice, then director of the National Center for Health Statistics, stated:

> *"As in other areas of social policy, health statisticians and health data are increasingly expected to provide keys to rational decision making. To accomplish this goal, the statistician and decision maker need to interact to an increasing degree."* (Rice, 1977, p. 101).

While these quotes were all made nearly 25 years ago, very little has happened since then with respect to the role that statisticians play in decision making. Pleas for change (Tukey, 1976; Stangl, 1995; Paltiel & Stinnett, 1996; Rice, 1977; Lindley 1997, 1998; Tan & Smith, 1998) and theoretical decision-making books (DeGroot, 1970; Berger, 1985; Lindley, 1985) are available, practical applications are rare.

The most important gap in current methodology follows from the fact that while decision-making is the incentive for most research efforts, the decision process usually remains informal and ad hoc.

The statistician's role has been to provide statistical synthesis and has not typically included promoting formalised decision-theoretic techniques. The disjuncture between statistical synthesis and decision making is an unnatural and undesirable one, because it undermines the impact of quantitative information.

Adopting a Bayesian perspective provides a natural bridge for this gap. But how does one teach Bayesian methods to persons with little mathematical training? There has been discussion of main didactical problems in teaching Bayesian inference at the undergraduate level, in particular, students' difficulties and misconceptions (Albert, 1997; Berry, 1997; Moore, 1997).

David Moore argues that it is, at best, premature to teach the ideas and methods of Bayesian inference in a first statistics course for general students. He argues that:

1. Bayesian techniques are little used;
2. Bayesians have not yet agreed on standard approaches to standard problem settings;
3. Bayesian reasoning requires a grasp of conditional probability, a concept confusing to beginners; and
4. An emphasis on Bayesian inference might impede the trend toward experience with real data and a better balance among data analysis, data production, and inference in first statistics courses.

Similar arguments could be given for not teaching medical researchers Bayesian inference; however, these arguments are equally specious (Stangl, 1998) for medical researcher education as for undergraduate education. It is my own belief that hiding from our students and research colleagues the fact that there is more than one interpretation of probability, and the fact that there is no universal solution to the problem of inductive inference, is irresponsible.

Also, based on my own teaching of Duke undergraduates and medical researchers, students find it much more difficult to understand sampling distributions and significance levels than they do Bayes theorem and predictive distributions. Distributions for future outcomes of interest are much more easily understood than distributions for test statistics. Empirical research to test these impressions is in the planning stages. Another author, Iversen (2001), also addresses these issues in this volume.

6. SPECIFIC FEATURES OF THE INTERNET COURSE

Many authors have proposed the use of case studies and course maps (Barnes et al., 1994; Bryant & Smith, 1995; Chaterjee et al., 1995; Bradstreet, 1996; Schau & Mattern, 1997; Par & Smith, 1998; Nolan & Speed, 1999). This course will use both.

A first arguments for case studies is that cases require complex problem solving and this translates into longer retention of statistical concepts. Second, because students like cases, they become engaged in learning and are willing to take self-responsibility for their learning. And third, cases impart reality, since cases send the message that data analysis is process rather than a static numerical exercise, and cases more effectively

communicate the need for statistical methods in the 'real' world. Students report increased confidence in their ability to use statistical methods in working with case studies.

A primary strength of this Internet resource will be the use of several health-related case studies. One such case is built around the GUSTO study (GUSTO investigators, 1993; Brophy & Joseph, 1995; Hively, 1996), a controversial study that has been analysed from both frequentist and Bayesian perspectives. It clearly highlights the differences between the paradigms, and the differences in answers to research questions that can arise when using different approaches.

The cases will be more in-depth than the examples provided in the general explanatory materials, will raise scientific and/or policy questions, provide rich relevant background material, contain data sets, and present analyses for addressing questions. The goal is to encourage and develop quantitative inductive and deductive reasoning skills within the context of health-related research.

The premise for course mapping is that understanding the relationships among statistical concepts is a prerequisite for effective statistical reasoning and problem solving. The model assumes that to be accessible from long-term memory, knowledge must be organised (or structured) and the relationships between different concepts clear and explicit. Teaching via the Web mandates organisation and explicit linkage. Because the teacher has less control over the student's navigational choices, the site's organisation is critical.

This course will use the concept of course mapping at multiple levels of resolution. At the lowest resolution general concepts will be linked with no technical detail. This level of resolution will be aimed at someone who wants the 'big picture' differences between Bayesian and frequentist inference. Bayes theorem will be presented not as a mathematical formula, but instead in words as a way to learn from data. Prediction and decision-making will also be presented as general concepts without mathematical formula.

A middle level of resolution will add some technical detail. Some probability and simple mathematics will be presented, but knowledge of calculus will not be assumed, and where calculus is used, it will be explained. This level of resolution will be aimed at the user who wants to be able to work textbook-level problems, and very simple real-world applications. It will demonstrate Bayes Theorem for discrete spaces and conjugate set-ups.

At the highest level of resolution there will be a great deal of technical grit, and knowledge of calculus will be assumed. This highest level of resolution will be aimed at the researcher who not only wants to understand the concepts behind Bayesian inference, but who also wants to be able to implement it in realistically complex applications.

7. COURSE CONTENT

The topics covered in this Internet course are the same as those covered in my 'live' courses except that time constraints are reduced on the Internet, and I can offer the student more choice in terms of technical detail.

At an introductory level, my teaching philosophy is much like that of Freedman, Pisani and Purvis (1998). I want deep thought without being too technical. My course is

very applied and I teach mostly by example. In this particular course, examples will include clinical trials and observational studies from a wide variety of medical topics. A general outline of the topics covered follows:

1. Distinguish the conceptual differences between the Bayesian and frequentist paradigms of statistics including the definition of probability and the likelihood principle (Berger & Berry, 1988);
2. Explain Bayes theorem and examine each of the components of a Bayesian model, including prior, posterior, and predictive distributions, and likelihood and utility functions;
3. Introduce approaches to the elicitation of prior distributions;
4. Teach how to calculate posterior and predictive distributions for simple conjugate distributions, and demonstrate techniques for calculating posterior and predictive distributions for more complex cases;
5. Teach how to use predictive distributions within a decision-theoretic framework;
6. Teach sensitivity analysis for prior distributions and utilities;
7. Present case studies of published health-related research that use Bayesian methods;
8. Demonstrate software useful for Bayesian analysis and explain how to gain access to this software.

8. EXPERIENCES TO DATE

This paper discusses my own conceptual thoughts and others' expert wisdom on designing an internet course for teaching Bayesian methods to medical researchers. It is based on my teaching experiences with diverse students including undergraduate premed students at Duke, applied biostatisticians and non-statisticians taking LearnStat courses through the American Statistical Association, and medical researchers at Duke University Medical School.

While the middle resolution content (slides and script) are ready, this summer will be my first chance to focus on structure. This will include pulling things together, mapping and linking. By the end of the summer I will begin piloting the project, and I will discuss marketing with the American Statistical Association. During the coming year I will focus on developing lower and higher levels of resolution to broaden the market of appeal.

REFERENCES

Albert, J. (1997). Teaching Bayes' rule: A data-oriented approach. *The American Statistician*, 51(3), 247-253.

Barnes, L. B., Christensen, C. R., & Hansen, A. J. (1994). *Teaching and the case method: Text, cases, and readings.* Boston: Harvard Business School Press.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. 2nd ed. New York: Springer.

Berger, J. O., & Berry, D. A. (1988). Statistical analysis and the illusion of objectivity. *American Scientist*, 76,159-165.

Berry, D. A. (1997). Teaching elementary Bayesian statistics with real applications in science.

The American Statistician, 51(3), 241-246.

Berry, D. A. (1996). *Statistics: A Bayesian perspective*. Belmont, CA: Duxbury Press.

Bradstreet, T. E. (1996). Teaching introductory statistics courses so that nonstatisticians experience statistical reasoning. *The American Statistician*, 50, 69-78.

Brophy J. M., & Joseph, L. (1995). Placing trials in context using Bayesian analysis: Gusto revisited by Reverend Bayes. *Journal of the American Medical Association*, 273(11), 871-875.

Bryant, P. G., & Smith, M. A. (1995). *Practical data analysis: Case studies in business statistics, vols. I and II*. Chicago, IL: Richard D. Irwin.

Chaterjee, S., Handcock, M. S., & Simonoff, J. S. (1995). *A casebook for a first course in statistics and data analysis*. New York: Wiley.

Cumming, G., & Thomason, N. (1998). Statplay: Multimedia for statistical understanding. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 947-952). Singapore: International Association for Statistical Education and International Statistical Institute.

Degroot M. H. (1970). *Optimal statistical decisions*. New York: McGraw Hill.

DiCiaccio, A. D. (1998). Hypermedia and WWW for the teaching of statistics. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 953-960). Singapore: International Association for Statistical Education and International Statistical Institute.

Freedman, D., Pisani, R., & Purves, R. (1998). *Statistics*. New York: Norton.

Galmacci, G. (2001). The impact of the internet on training researchers. In C. Batanero (Ed.), *Training of Researchers in the Use of Statistics* (pp. 159-169). Granada: International Association for Statistical Education and International Statistical Institute.

Gartner Group (1998 a). Interactive media scenario. Paper presented at the *Gartner Group Symposium: ITxpo98*, October, 1998, Lake Buena Vista, Florida.

Gartner Group (1998 b). Internet age publishing. Paper presented at the *Gartner Group Symposium: ITxpo98*, October, 1998, Lake Buena Vista, Florida..

Gartner Group (1998 c). Strategic technology planning: Defining the role. Paper presented at the *Gartner Group Symposium: ITxpo98*, October, 1998, Lake Buena Vista, Florida..

Giusti, A., Petrucci, A., & Polverini, F. (1998). Visual statistics: A multimedia statistics facilitator. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 969- 974). Singapore: International Association for Statistical Education and International Statistical Institute.

The GUSTO Investigators, (1993). An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. *New England Journal of Medicine* 329, 673-682.

Hively, W. (1996). The mathematics of making up your mind. *Discover*, May, 90-97.

Iversen, G. (2001). Should we even teach statistics: A Bayesian perspective. In C. Batanero (Ed.), *Training of Researchers in the Use of Statistics* (pp. 103-112). Granada: International Association for Statistical Education and International Statistical Institute.

Korpela, J. (1997). *Publishing on the web is different*. www.hut.fi/u/jkorpela/webpub.html.

Laake, P. (1998). Teaching statistics to medical doctors via the world wide web. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 287-288). Singapore: International Association for Statistical Education and International Statistics Institute.

Lindley, D. V. (1985). *Making decisions*. Chichester: Wiley.

Lindley, D. V. (1997) The choice of sample size (with discussion). *The Statistician*, 46, 129-166.

Lindley, D. V. (1998). Decision analysis and bioequivalence trials. *Statistical Science*, 13(2),

136-141.

MacRae, S. (1998). Introducing statistical principles by interactive multimedia. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 975-976). Singapore: International Association for Statistical Education and International Statistical Institute.

Moore, D. S. (1997). Bayes for beginners? Some reason to hesitate. *The American Statistician*, 51(3), 254-261.

Morris, E., & Le Voi, M. (1998). Sorting the wheat from the chaff: Evaluating a large corpus of computer-based statistics teaching resources. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 977-984). Singapore: International Association for Statistical Education and International Statistical Institute.

Nolan, D., & Speed, T.P. (1999). Teaching statistics theory through applications. *The American Statistician*, 53, 370-375.

Parr, W., & Smith, M. (1998), Developing case-based business statistics courses. *The American Statistician*, 52, 330-337.

Phillips, B, Fancis, G, & Hutcheson, M. (1998). Statistics for workers in social and health sciences by flexible learning modes. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 410-408). Singapore: International Association for Statistical Education and International Statistical Institute.

Paltiel, A. D., & Stinnett A. A. (1996). Making health policy decisions: Is human instinct rational? Is rational choice human? *Chance*, 9(2), 34-39.

Puranen, J. (1998). WWW and teaching statistics - a teacher's point of view. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 993-998). Singapore: International Association for Statistical Education and International Statistical Institute.

Rice, D. (1977). The role of statistics in the development of health care policy. *The American Statistician*, 31(3), 101-106.

Schau, C, & Mattern, N. (1997). Use of map techniques in teaching applied statistics courses. *The American Statistician*, 51(2),171-175.

Stangl, D. (1995). Prediction and decision making using Bayesian hierarchical models. *Statistics in Medicine*, 14(20), 2173-2190.

Stangl, D. (1998). Classical and Bayesian paradigms: Can we teach both? In L. Pereira-Mendoza, L .S. Kea, T. W. Kee, & W. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 251-258). Singapore: International Association for Statistical Education and International Statistical Institute.

Tan S. B., & Smith, A. F. M. (1998). Exploratory thoughts on clinical trials with utilities. *Statistics in Medicine*, 17, 2771-2791.

Tukey, J. W. (1976). Discussion of: Role of statistics in national health policy decisions. *American Journal of Epidemiology*, 104(4), 380-385.

Talbot, M., Horgan, G., Mann, A., Alonso, R., Badia, J., Bishop, G., & Quednau, D. (1998). SMART: Introducing specialist statistical techniques via the web. In L. Pereira-Mendoza, L .S. Kea, T. W. Kee, and W. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 409-414). Singapore: International Association for Statistical Education and International Statistical Institute.

The GUSTO Investigators (1993). An international randomised trial comparing four thrombolytic strategies for acute myocardial infarction. *New England Journal of Medicine* 329, 673-682.

Watson, J. (1998). Professional development of teachers using cd-rom technology. In L. Pereira-Mendoza, L .S. Kea, T .W. Kee, and W. Wong (Eds.), *Proceedings of the Fifth*

International Conference on Teaching Statistics (pp. 921-927). Singapore: International Association for Statistical Education and International Statistical Institute.

Dalene Stangl
Duke University,
Institute of Statistics and Decision Sciences,
P.O. Box 90251, 223C Old Chemistry,
Durham, NC, 27708-0251, USA.
dalene@stat.duke.edu

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 159-169.

Printed in Granada, Spain.

GIANFRANCO GALMACCI

THE IMPACT OF THE INTERNET ON RESEARCHERS' TRAINING

The paper explores the reasons that have made the Internet even more important for the training of researchers, both for pure statisticians and scientists who use statistics for analysing experimental data. The author discusses various aspects of the role of the Internet and focuses on those contributions that populate the network with valuable tools and services (electronic journals, electronic books, virtual labs, data archives and examples of data analysis, etc.). An extensive bibliography has been provided to guide new users, through the millions of sites of the chaotic World Wide Web, towards the most interesting places for people involved in statistics.

1. INTRODUCTION

Most people have discovered the Internet because of the World Wide Web and often WWW is used as a synonym for the network itself, and vice-versa. So it is usually believed that the Internet is no more than five or six years old.

On the contrary, the history of the Internet is quite long, especially if related to computer history. Robert E. Kahn and Vinton G. Cerf published their first paper on TCP protocol in 1974, so in 1999 the 25th anniversary of the Internet was celebrated (TCP/IP is the basic communication protocol of the Internet).

However, the network really became important in the 80s although, for about ten years, it involved mostly academic and research environments (mainly technological and scientific areas). The crucial innovation that made the Internet one of the most important and popular events in communication (perhaps similar to the telephone, radio or television) was the HTTP protocol introduced by Tim Berners Lee in the early 90s to implement the World Wide Web.

Today's popular image of the Internet consists mainly of fancy images, animation, sound, movies, advertisement, all kinds of information, easy downloading and many more tools for fun. However, few people know that all of this has been realised thanks to the thousands of researchers who have contributed anonymously for many years to improve the network's technical characteristics as well as to enrich it with software tools, procedures, data, discussion lists, etc. (all of them free of charge) and who have changed the way academics are working today. In fact, all the main services that have become fundamental for the network have arisen from the scientific Internet community (e-mail, discussion lists, Usenet, anonymous ftp archives and the World Wide Web itself).

The Internet Software Consortium has estimated that in January 2000 there were more than 72 million computers connected to the Internet all over the world. This number is not so interesting in itself, but it becomes much more significant when we consider that in 1993, just before the WWW era, the number of computers connected

was only 1,313,000. Furthermore, the "Messaging Online" estimated 569 million e-mail accounts globally at year-end 1999, up 83 percent on the previous year.

Many facts are helpful for understanding the importance of what is happening. Today we use the term "new economy" to talk about the business born around the network; almost all the Universities have their own network connected to the Internet (even though the situation in some developing countries cannot be compared with that of the more advanced ones). In the 1996 State of the Union Address President Clinton said:

> *"In our schools, every classroom in America must be connected to the information superhighway, with computers and good software, and well-trained teachers. We are working with the telecommunications industry, educators and parents to connect 20 percent of California's classrooms by this spring, and every classroom and every library in the entire United States by the year 2000".*

During the last State of the Union Address (January 27, 2000) the President said:

> *"We know we must connect all our classrooms to the Internet, and we're getting there. In 1994, only 3 percent of our classrooms were connected. Today, with the help of the Vice President's E-rate program, more than half of them are. And 90 percent of our schools have at least one Internet connection".*

All these elements are drawing a scenario for the near future where the networks will become a fundamental medium for most activities: commercial, social and, of course, scientific.

2. THE INTERNET FOR STATISTICS

The Internet can be considered as the biggest database ever built and it grows day by day. Information spreads out in all directions and it is available in real time. Today we can say that there is no scientific field that has not yet been influenced in some way by the network facilities. Sometimes the benefits may simply be a more powerful way of communicating and of reducing distances. Sometimes, there are more important advantages depending on the specific field.

However, considering the Internet resources for statistics, we must take account of two different kinds of people: actors and spectators. In the first category we find mainly statisticians who provide knowledge on statistical methods, software tools, examples, etc., showing various kinds of useful ways to profit from this technology.

The second category includes people looking for such information (i.e. researchers interested in experimental fields, teachers, and young statisticians).However, while some of them are enthusiastic, some others do not care much about it. In any case there are many important facts that must be taken into account, especially if we consider training for young researchers.

The future tasks of young researchers consist of research itself and teaching activities. Both areas have recorded improvements using the network, even though in different forms and with different effects. Faster communication, electronic journals, discussion lists, data and case study archives, paper archives, distributed processing, scientific software distribution, etc. all contribute to improving research activities.

2.1. ELECTRONIC JOURNALS

Many scientific journals accept only papers in electronic form and the most important publishers have their own templates for several text-processors so that authors can submit articles in the right form ready for printing. The proceedings of many national and international conferences are often published only on CD-ROM. Moreover, the main journals make online versions available at reduced subscription rates and, since 1993, a growing number of electronic journals are available completely free or at a very low cost. Electronic journals are not well known yet, so it is useful to provide a list of some of them concerning Statistics and/or Education in Table 1.

Table 1. Statistics and Education Electronic Journals

Journal	First issue	URL
The Educational Technology Journal	1991	http://www.fno.org/
Journal of Statistics Education	1993	http://www.amstat.org/publications/jse/
Compute-Ed	1995	http://computed.coe.wayne.edu/
ESAIM: Probability and Statistics	1995	http://www.edpsciences.com/ps/
Interstat	1995	http://interstat.stat.vt.edu/interstat/intro.html-ssi
Electronic Communications in Probability	1996	http://math.washington.edu/~ejpecp/
Electronic Journal of Probability	1996	http://math.washington.edu/~ejpecp/
Journal of Statistical Software	1996	http://www.stat.ucla.edu/journals/jss/
Studies in Nonlinear Dynamics and Econometrics	1996	http://mitpress.mit.edu/e-journals/SNDE/
The Electronic Journal of Science Education	1996	http://unr.edu/homepage/jcannon/ejse/ejse.html

In addition, many newsletters are published by Scientific Associations both as paper and electronic versions, such as the *Newsletter of the Statistical Computing and Graphics* (http://cm.bell-labs.com/cm/ms/who/cocteau/newsletter/index.html), the *Newsletter of the ASA Section on Statistical Education* (http://renoir.vill.edu/~short/StatEd/), the *Newsletter of the International Study Group for Research on Learning Probability and Statistics* (http://www.ugr.es/~batanero/iase.html), the *IASE Statistical Education Research Newsletter* (http://www.ugr.es/~batanero/sergroup.htm). Furthermore, many departments of statistics have now a repository for technical reports, internal drafts and PhD thesis, freely available in PostScript or PDF formats. This organisation offers a solution to an old problem. As everyone has experienced, till a few years ago this material was mostly unavailable or unreachable without unacceptable delays, even though it often represented a valuable part of the scientific literature.

2.2. STATISTICS AGENCIES

National and International Statistics Agencies are investing in many directions to profit from the network technology. Among the fields more involved in this effort are the channels used to disseminate information: Podehl (1999) has recently written that electronic publishing via the Internet has supplemented and is now starting to overtake paper publishing. Furthermore, the World Wide Web is a very efficient tool FOR PRESENTING statistical information in a new fashion which contributes to attracting

THE interest of ordinary people: the classic social facts (economy, demography, etc) are often presented exploiting the "technological appeal" as much as possible.

Data, so boring for non-insiders, are indirectly used to describe a country, a population, etc., from a new perspective. The discovering of this information by computer is something like a game which contributes to disseminating a statistics mentality and to making the new languages of the so-called information society more friendly. The effort required to implement these ideas is greater, the simpler and friendlier the final result actually is .Furthermore it requires a serious research effort in technological, statistical and socio-psychological areas.

Many agencies are today investing strongly in order to set up Web environments which allow a quick search of statistical information, online computation and graphing, webcasting, etc. Some good examples are already available: The U.S. Census Bureau has carried out a very efficient, advanced and attractive mapping system known as TIGER (Topologically Integrated Geographic Encoding and Referencing system). Another example is the American FactFinder (Reference Maps service, for displaying maps at different scale level, from the whole country level to the street level, and Quick Thematic Maps service, for creating of geographic patterns for commonly requested statistics).

Statistics Canada has set up E-Stat, a very friendly educational environment, and CANSIM (CANadian Socio Economic Information Management System), a professional data-searching engine. Eurostat is working on VIROS (Virtual Institute for Research in Official Statistics) and is experimenting a number of alternative approaches with the purpose of facilitating access to information. They rely on advanced technologies and, as a consequence, specific technical requirements must be met to view them. A remarkable Eurostat project is the Virtual European Statistical Laboratory, which involves state of the art techniques in many fields of computer science. The goal is to facilitate web-based collaboration and teleconferencing by setting up an intuitive three-dimensional multi-user environment (see the Eurostat pages and look for Miscellaneous in the research section).

2.3. ELECTRONIC BOOKS AND VIRTUAL LABS

Statistical education seems to be another field where network technologies can provide new settings: distance learning, virtual labs, electronic texts and books, automatic student assessment and much more. Conferences on teaching statistics always host sessions for discussing new proposals and experiences and during the last few years various experimental sites have revealed what can be done in this field. The description of all the different approaches and technologies used in this field would be too long and not so useful (a list of the most important sites has been included among the WWW References). Therefore, we will discuss the main ideas only to underline those elements which justify efforts in this direction because of their significant contribution for some categories of Web user.

Some educational sites have been produced for teaching purposes, having in mind college or university students, and specific course requirements. Moreover, authors often decided to invest in this direction to prepare practice materials for autonomous use by the students wherever they are. More recently, on the other hand, various sites offer a "virtual lab" developed as a part of a specific teaching project. Sometimes they also take into account pedagogical and psychological aspects besides the statistical methods (a good example is TILE, a project carried on at the Statistics Department, Berkeley).

Today many electronic texts, electronic books and virtual labs are available from the network. The scientific literature has not yet reported on experiments which show whether such tools are effective and really improve the student's learning processes; that is, there are no data which can be used to test them for teaching effectiveness. However they provide valuable support for non-statisticians who need statistics in their experimental research. Generally this kind of material takes more care of applications than many classical books do. From this point of view the Web represents an important starting point for identifying examples and for guided training on data analysis using appropriate software.

Several sites provide this kind of support either for elementary statistics and probability or for advanced methods (i.e. analysis of variance, linear models, etc.). The organisation of online material varies from site to site depending on the aim of the project. Virtual labs consist of software packages that can perform statistical computations directly from the WWW pages; generally written in Java, XlispStat, Perl, etc., they often use the dynamic graphic approaches and allow a fully online interaction with graphs. Some of them (Webstat, Statlets, etc.) are also able to get input data from everywhere on the network just declaring the URL. Similar features are absolutely innovative and make these kinds of tools much more friendly and more intuitive than most of the commercial products (XploRe, Härdle & others 1995, is the only one that has been built using some of these approaches). It seems reasonable to say that these prototypes are establishing new paths for future statistical software architectures.

Electronic texts and books may consist either of sets of simple notes on course topics with examples or of much more complex environments which provide both theoretical aspects and built-in software tools. Some of them (i.e. HyperStat) exploit other sites (mainly the virtual labs) for practical sessions by using links to these Web pages. Moreover, videos and 3D Virtual Reality examples are often available with charming examples.

As a final consideration on this matter, we have to keep in mind that the network actually offers all the best we can do with new technologies at the present time. Not everything may be considered useful, important and scientifically relevant; some are only attractive because of technology fashion. However the World Wide Web represents today a very important place for finding information and support.

2.4. DATA AND SOFTWARE ARCHIVES

The growth of the Internet during the '80s, when it was an academic network only, occurred for many reasons: First of all the technical characteristics of the TCP/IP protocol, secondly the new kind of services made possible by the protocol itself. During this period one of the most important services was the anonymous FTP, which allowed for the setting up of archives where researchers collected their software prototypes for free downloading. Perhaps most people today no longer understand the meaning of "anonymous FTP" because now this service is part of the facilities provided directly by the World Wide Web and hence is almost completely hidden as an autonomous service, even though it remains one of the most important ones.

The first relevant general purpose archive for statisticians is Statlib, set up by Mike Mayer at the Carnegie Mellon University at the beginning of the '90s. It is a public repository for statistical issues open to everyone both for storing and getting information. Now Statlib collects lots of famous data sets, which have appeared in

statistical journals or books, computer programs and macros for the most common statistical packages, and much more.

Nowadays all the universities, and perhaps all their departments, have a Web site so that people often prefer to control and to personalise their home pages. Therefore the number of statistical archives has quickly increased and it is day by day more difficult to have a realistic overview of what is available on the network. For many reasons (see the next paragraph) search engines are often unable to retrieve what you are looking for, and in any case the research is time wasting. A good solution is still offered by those sites containing ordered and classified lists of statistical resources. These sites are courteously maintained by people who, having a wide knowledge of the network, have freely decided to give a contribution to the scientific community by sharing their knowledge (such sites are listed among the "World Wide Web References" in the section "Links to Statistical Resources").

Data, case studies, software and all the related information represent an invaluable scientific heritage for the researcher's training as well as for some teaching purposes.

3. STATISTICS FOR THE INTERNET

Engineers and computer scientists have recently coined the new term *datamind* for embracing all the strategies concerned with non-homogeneous databases. They are working to make information compatible, but there are lots of statistical problems in this field that concern statisticians only.

The Internet is a container for all kinds of information, completely confused, randomly reachable and, for these reasons, unreliable for more advanced uses. Even though it could seem strange, this is a substantial advantage for the "health" of the network itself. In fact it stimulates researchers to develop new efficient methods for managing information coming from hundreds of millions of documents distributed on tens of millions of computers and contributes to the improvement of the Internet and the quality of the services offered.

Today, thanks to the WWW, it is very easy to create Web sites and archives on whatever computer or operating system is available so that anyone can publish their own information by himself. Unfortunately, the information spreads out all over the world and it is impossible to "organise" it in a reasonable way. Though Web searching engines are becoming more and more efficient, selective and powerful, they fail when the documents do not contain enough elements for quick retrieving (i.e. keywords and HTML meta-data). In fact, the easiness of HTML document building often encourages people to organise their sites without caring for the basic rules of the WWW and ignoring how to make the information retrievable.

Some computer scientists and statisticians are now trying to extend data-mining technologies to this area in order to deal with all the problems connected with the network growth.

4. CONCLUSIONS

The researcher's background must always include a good knowledge of the Internet facilities for different reasons. First of all, the network provides information and various kinds of materials that are very important for personal research activities.

Secondly it is necessary for everyone to be able to participate in this effort sharing their own knowledge. Moreover, the quick revolution of the past few years has already shown the advantages of the new means of communication so, in this period of transition, the task of exploring and drawing new paths to exploit the network opportunities is entrusted to researchers.

REFERENCES

Bowman, A. (1997). Computer-based learning and statistical problem-solving. In E. J. Wegman, & S. P. Azen (Eds.), *Proceedings of the 2nd World Congress of the IASC* (p.312). Pasadena, CA: Interface Foundation of North America.

Cerf V. G., & Kahn R. E. (1974). A protocol for packet network interconnection. *IEEE Transactions on Communications Technology*, 5, 627-641.

Currall, J. (1997). Computer aided statistics teaching: Real advance or technological fashion? In E. J. Wegman, & S. P. Azen (Eds.), *Proceedings of the 2nd World Congress of the IASC* (p.321). Pasadena, CA: Interface Foundation of North America.

Härdle W., Klinke S., & Turlach, B.A. (1995). *XploRe: An interactive statistical computing environment*. New York: Springer.

Galmacci, G. (1995). Statistics in the era of Internet. *Proceedings of ISI 50th Session, Bulletin of the International Statistical Institute* (Tome LVI, Book 1, pp. 227-237). Beijing: International Statistical Institute.

Galmacci, G. (1998). Network tools for teaching statistics. In L. Pereira-Mendoza, L .S. Kea, T. W. Kee, & W. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 961-967). Singapore: International Association for Statistical Education and International Statistical Institute.

Galmacci, G. (1999). Output management. *Proceedings of the Satellite Conference on Statistical Publishing* (pp. 52-53). Warsaw: International Statistical Institute and Polish Central Statistical Office.

Podehl, M. (1999). Managing information for electronic and paper publishing. *Proceedings of the Satellite Conference on Statistical Publishing* (pp. 48-51). Warsaw: International Statistical Institute and Polish Central Statistical Office.

Saporta, G. (1999). Teaching statistics with Internet: A survey of available resources and the St@tNet project. *Proceedings of the52nd Session of the ISI, Bulletin of the International Statistical Institute* (Tome LVIII, Book 2, pp. 233-236). Helsinki: International Statistical Institute.

Unwin A. (1997). Using your eyes. Making statistics more visible with computers. In E. J. Wegman, & S. P. Azen (Eds.), *Proceedings of the 2nd World Congress of the IASC* (p.305). Pasadena, CA: Interface Foundation of North America.

West W. (1997). Statistical application for the World Wide Web. *Proceedings of the52nd Session of the ISI, Bulletin of the International Statistical Institute* (Tome LVIII, Book 2, pp. 237-240). Helsinki: International Statistical Institute.

WORLD WIDE WEB REFERENCES

Documents on the Internet

Berners L. T. (1989), *Information Management: A Proposal*, http://www.w3.org/History/1989/proposal.html.

Clinton W. (1996), *The President's 1996 state of the Union address* http://library.whitehouse.gov/WH/New/other/stateunion-top.html, January 23, 1996.

Clinton W. (2000), *The President's 2000 state of the Union addres*,

http://www.whitehouse.gov/WH/SOTU00/sotu-text.html, January 27, 2000.
Internet Software Consortium, (2000), *Internet domain survey*, http://www.isc.org/.
Messaging online (April 2000), *Featured stories*, http://www.messagingonline.com/.
Nua (2000), *Usage patterns*, http://www.nua.ie/surveys/.
US Department of Education (2000), *The education record, 1993-2000*, http://www.ed.gov/inits/record/.

Statistical Agencies
Eurostat, *VIROS - Virtual Institute for Research in Official Statistics*, http://europa.eu.int/comm/eurostat/research/.
Statistics Canada, *Educational resources*, http://www.statcan.ca/english/kits/index.htm.
U.S. Census Bureau - *TMS (TIGER Mapping Service)*, http://tiger.census.gov/.
U.S. Census Bureau - *American FactFinder* http://factfinder.census.gov/java_prod/dads.ui.homePage.HomePage.

Centers for Statistical Education
ASA, *Center for Statistics Education*, http://amstat.org/education/index.html.
CIRDIS, *Centro Interuniversitario di Ricerca per la Didattica delle Discipline Statistiche*, http://www.stat.unipg.it/CIRDIS/.
CTI Statistics (University of Glasgow), http://www.stats.gla.ac.uk/cti/.
IASE, *International Association for Statistical Education*, http://www.cbs.nl/isi/iase.htm.
ISEC, *The International Statistical Education Centre*, http://www.isical.ac.in/~isec/.
NCES, *National Center for Education Statistics*, http://nces.ed.gov/.
NCTM, *National Council of Teachers of Mathematics*, http://www.nctm.org/.
PME, *Psychology of Mathematics Education*, http://www.unifr.ch/psycho/pme/pme.html.
Research Group on Statistical Education (University of Granada), http://www.ugr.es/~batanero/.
Royal Statistical Society Centre for Statistical Education, http://science.ntu.ac.uk/rsscse/.
Royal Statistical Society Education Committee, http://www.maths.ntu.ac.uk/nd/www/rssec.html.
The Korean Statistical Education & Consulting Section of The KSS, http://stat.anyang.ac.kr/sec/.
SIIP, *The Statistical Instruction Internet Palette*, http://research.ed.asu.edu/siip/.

Links to Statistical Resources
Weiss N.A., Arnold.T, *Internet Projects for Elementary Statistics*, http://hepg.awl.com/weiss/e_iproject/index.htm.
Weiss N.A., Arnold.T, *Internet Projects for Introductory Statistics*, http://hepg.awl.com/weiss/i_iproject/index.htm.
Arsham H. (University of Baltimore), *Computational Probability and Statistics Resources*, http://ubmail.ubalt.edu/~harsham/statistics/REFSTAT.HTM.
Friendly M. (York University, Toronto, Ontario, Canada), *Statistics and Statistical Graphics Resources*, http://www.math.yorku.ca/SCS/StatResource.html.
Galmacci G., *IASC - International Association for Statistical Computing*, http://www.stat.unipg.it/IASC/.
Smyth G. (University of Queensland, Brisbane, Australia), http://www.maths.uq.oz.au/~gks/webguide/index.html.
Statsnet, *Statistics Network*, http://www.statsnet.net/.

On-line Course Materials
Baker R. (University of Saskatchewan, CA), *Basic principles of statistical analysis*,

http://duke.usask.ca/~rbaker/stats.html.
CNAM - Conservatoire National des Arts et Métiers & Université Virtuelle Francophone, *St@tNet*, http://www.cnam.agropolis.fr/statnet/.
Dear K. & others (University of Newcastle), *SurfStat Australia*, http://surfstat.newcastle.edu.au/surfstat/.
de Leeuw J. (University of California, Los Angeles), *Statistics UCLA*, http://www.stat.ucla.edu/textbook/.
Department of Statistics (University of South Carolina), *Interactive Statistics and the GASP Initiative - Globally Accessible Statistical Procedures*, http://www.stat.sc.edu/rsrch/gasp/.
Glass G. (Arizona State University), *Intro to quant methods*, http://glass.ed.asu.edu/stats/.
Hopkins W.G. (Sportscience), *A New View of Statistics*, http://www.sportsci.org/resource/stats/.
Institute of Phonetic Sciences (IFA, University of Amsterdam), *Statistical tests* http://fonsg3.let.uva.nl:8001/Service/Statistics.html.
Lane D.M. (Rice University), *HyperStat*, http://davidmlane.com/hyperstat/index.html .
Lane D.M. (Rice University), *Rice Virtual Lab in Statistics*, http://www.ruf.rice.edu/~lane/rvls.html.
Lowry R. (Vassar College, Poughkeepsie, NY), *VassarStats*, http://faculty.vassar.edu/~lowry/VassarStats.html.
Martínez de Lejarza J. & Martínez de Lejarza I. (Universidad de Valencia) http://www.uv.es/~lejarza/estadistic.htm.
MSTE (University of Illinois at Urbana-Champaign), *K-12 Statistics*, http://www.mste.uiuc.edu/stat/.
Newton H.J., Carroll J.H., Wang N., Whiting D. (Texas A&M University), *Statistics 30X Class Notes*, http://stat.tamu.edu/stat30x/trydouble2.html.
NWP Associates, Inc. (The Pennsylvania State University), *Investigating Statistics*, http://espse.ed.psu.edu/statistics/investigating.htm.
Puranen J. (University of Helsinki), http://noppa5.pc.helsinki.fi/.
Statistics Department (University of California at Berkeley) and Duncan Temple Lang (Lucent Technologies, Bell Labs Innovations), *Tools for an Interactive Learning Environment (TILE)*, http://www.stat.berkeley.edu/users/nolan/TILE/.
Rao B., McAndrews P., Sunkara A., Patil V., Quisumbing G., Le H. L., Zhou Z. (Center for the New Engineer & Defense Acquisition University), *The Statistics Line*, http://www.cne.gmu.edu/modules/dau/stat/.
Snell J.L., Peter Doyle, Joan Garfield, Tom Moore, B. Peterson, N. Shah (Dartmouth College), *CHANCE*, http://www.dartmouth.edu/~chance/.
Stirling W. D. (Massey University, NZ), *CAST - Computer Assisted Statistics Teaching*, http://cast.massey.ac.nz/.
Stockburger D. W. (Southwest Missouri State University), *Introductory Statistics*, http://www.psychstat.smsu.edu/sbk00.htm.
Stockburger D. W. (Southwest Missouri State University), *Multivariate Statistics*, http://www.psychstat.smsu.edu/MultiBook/about.htm.
Stark P.B. (University of California, Berkeley), *SticiGui: Statistical Tools for Internet and Classroom Instruction with a Graphical User Interface*, Berkeley http://www.stat.berkeley.edu/~stark/SticiGui/index.htm.
StatSoft, Inc. (Tulsa, OK), *Electronic Statistics Textbook*, http://www.statsoft.com/textbook/stathome.html.
Swinscow T.D.V. (BMJ Publishing Group), *Statistics at Square one*, http://bmj.com/statsbk/.

WWW Interactive Software

Arnold D.N. (University of California, Berkeley), Graphics for the Calculus Classroom, http://www.math.psu.edu/dna/graphics.html.
Behrens J. (Arizona State University), *Research Methods Resources*, http://research.ed.asu.edu/.
Bonnier M. (Värpinge, Lund, Scania, Sweden), *Statiscope*, http://www.df.lth.se/~mikaelb/statiscope/statiscope-enu.shtml.
Claremont Colleges' "Web Interface for Statistics Education", *Statistics Applet made for the WISE Project*, http://www.grad.cgs.edu/wise/appletsf.shtml.
de Leeuw J. (University of California, Los Angeles), *Statistics UCLA*, http://www.stat.ucla.edu/textbook/.
Härdle W., Klinke S., Turlach B.A., *XploRe*, http://www.xplore-stat.de/index_js.html.
Lewis B. (Kent State University), *Elementary Statistical Java Applets and Tools*, Marden J. (University of Illinois at Urbana-Champaign), The CUWU Statistics Program, http://www.stat.uiuc.edu/~stat100/cuwu/.
NWP Associates, Inc. (The Pennsylvania State University), *STATLETS*™, http://espse.ed.psu.edu/statistics/statlets/statlets.htm.
Siegrist K. (University of Alabama in Huntsville), *Virtual Laboratories in Probability and Statistics*, http://www.math.uah.edu/stat/.
Stark P.B. (University of California, Berkeley), *SticiGui: : Statistical Tools for Internet and Classroom Instruction with a Graphical User Interface*, http://www.stat.berkeley.edu/~stark/SticiGui/index.htm.
Statistics Department (University of Glasgow), *STEPS - STatistical Education through Problem Solving*, http://www.stats.gla.ac.uk/steps/home.html
StatPoint, LLC, *Internet Statistical Computing Center*, http://www.sgcorp.com/index.htm.
West R. W. & Ogden R. T. (University of South Carolina), *WebStat*, http://www.stat.sc.edu/webstat/.
Young F. W. (University of North Carolina), *Vista - The Visual Statsistical System*, http://forrest.psych.unc.edu/research/index.html.
Zielman B. (University of), *Statistical Page*, http://huizen.dds.nl/~berrie/.

Data and Case Studies

American Statistical Association, *JSE Data Archive*, http://amstat.org/publications/jse/archive.htm.
Behrens J. (Arizona State University), *Dr. B's Data Gallery*, http://seamonkey.ed.asu.edu/~behrens/classes/data_gallery/.
Cornell University, DASL - The Data and Story Library, http://lib.stat.cmu.edu/DASL/.
de Leeuw J. (University of California, Los Angeles), *Case Studies*, http://www.stat.ucla.edu/cases/.
Department of Statistics (Carnegie Mellon University*), StatLib*, http://www.stat.cmu.edu/.
Harvard-MIT Data Center, *Quantitative social science data*, http://hdc-www.harvard.edu/.
Office of Information Technologies (University of Massachusetts Amherst), *Statistical Datasets*, http://www-unix.oit.umass.edu/~statdata/.
Statistical Engineering Division & Mathematical and Computational Sciences Division (University of Florida), *Statistical Reference Datasets*, http://www.itl.nist.gov/div898/strd/.
Stukel T.A. (Dartmouth College, USA), *Introduction to Linear Models*, http://www.dartmouth.edu/~matc/math30/CourseTopics.html.
Teaching and Educational Development Institute, *OzDASL - Australasian Data and Story Library*, http://www.maths.uq.edu.au/~gks/data/.
University of California, San Diego, SSDC - *Social Sciences Data Collection*, http://ssdc.ucsd.edu/index.html.

University of Essex, ESRC, JISC, *The Data Archive*, http://www.data-archive.ac.uk/.
University of Wisconsin-Madison, DPLS - *Data and Program Library Service*, http://dpls.dacc.wisc.edu/.
Wiggins, D. (*TRAMSS - Training & Dissemination Project*, http://tramss.data-archive.ac.uk/.

Miscellaneous
Binker J. (Key Curriculum Press), *Statistics?! Don't Make Me Laugh!*, http://www.keypress.com/fathom/jokes.html.
Harnisch D. (University of Illinois at Urbana-Champaign), *Statistical Humor*, http://eval1.crc.uiuc.edu/edpsy390/statjokes.html.
Ramseyer G. C. (Illinois State University), *The First Internet Gallery Of Statistics Jokes*, http://www.ilstu.edu/~gcramsey/Gallery.html.
Verhagen J. (University of Utah), *ScienceJjokes*, http://www.xs4all.nl/~jcdverha/scijokes/index.html.
Verhagen J. (University of Utah), *Statistician Jokes*, http://info.business.utah.edu/~bebrblf/statjoke.html.

Gianfranco Galmacci
Dipartimento di Scienze Statistiche,
Università degli Studi,
Casella Postale 1315 / Succ. 1,
06123 Perugia, Italy.
E-mail: glm@stat.unipg.it

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 171-173.

Printed in Granada, Spain.

KENSEI ARAYA

DISCUSSION: HOW TECHNOLOGY AFFECTS THE TRAINING OF RESEARCHERS

As is generally recognised, the rapid progress of information technology typically represented by the Internet has had a tremendous effect on areas as various as industry, education and social science.

I shall start this discussion with some general comments about the essential factors of information in messages involving two different players namely sender and receiver of the message. While the sender wants to transmit rapidly a large amount of information with as low cost as possible, the receiver wants to get the most valuable and newest information for himself. The modern bilateral communication tool has successfully bridged these gaps between them by constructing visual and compressive technologies.

In the educational area the sender of information is a teacher and the receivers are students in the classroom where special kinds of communication beyond the usual conversation levels are requested, focusing on how to effectively learn contents of subjects. One major problem for the classroom teaching rests with the teacher who has to develop his or her own techniques and ability to communicate adequately.

At the present stage of statistical education, personal computers have entered providing splendid ability for computation and access to Internet technology, which can instantly transmit a huge amount of information to the whole world. With respect to this point it should be noted that these functions of personal computers correspond perfectly to the quantitative analysis and large data processing evident in statistical analyses.

1. OVERVIEW OF PAPERS

In this session we have five papers and among them three are directly related to the application of information technology.

First, Galmacci's paper summarises valuable and fruitful information and teaching materials on statistics education by referring to homepages, electronic journals, statistics agencies, electronic books and software archives mainly produced by statisticians and statistics educators. The importance of this kind of work is increasing so much by the day that researchers and consultants are hard pressed to keep up with advancing studies. This paper should be recognised as a pioneering contribution in this respect.

In Stangle's paper, special attention is paid to two questions, why Internet-based instruction, and why the Bayesian paradigm. In designing an Internet course for medical researchers, these questions occupy essential positions for constructing progressive interaction between simulation demonstration by the Internet and the decision side of

Bayesian statistics. Bayesian statistics especially plays an important role in building user- oriented image processes, carrying out visual and sophisticated changes of a posterior distribution according to the choice of prior distribution and the likelihood based on the sample data. But, in order to get optimal decisions, there remain crucial problems and we have to construct a payoff or loss table reflecting medical situations. There are have great expectations of further development according to the author's discussion.

On the topic of teaching biostatistics to medical personnel with the aid of computer technology, Lee concludes from a sample survey that the new teaching method of utilising Internet facilities has comparative advantages over the traditional way in health science courses in Korea National Open University. It should be added, however, that the traditional way of keeping close contact between teachers and students by using handy graphic calculators exhibits considerable effectiveness at least in the primary stages of training.

At this stage I am reminded of the time in the meeting, when I displayed several applets in the *Journal of Statistics Education* v.6, n.3, presented by W. West and T. D. Ogden, which by one click operation, show the effect of changing width intervals in histograms, the effect of influential points in regression and the effect of the number of trials on limit theorem. Many participants expressed strong doubt about the effectiveness of these images for teaching undergraduate students. In addition to these direct uses of the Internet, the Internet has the additional use of permitting communication among students through e-mail and promoting co-operation in learning statistics. The Internet will become the most powerful tool for better teaching and learning in the near future.

In his paper Professor Shimada warns students against blind reliance on software packages and emphasises the importance of finding a suitable model with its solution for given situations. Today we have a lot of software packages so well organised and sophisticated that we can not easily rewrite or improve them without expert knowledge. As the paper shows, it may be profitable enough for undergraduate students even though established software packages have a risk of leading to incorrect conclusions. But for graduate students or researchers it becomes important to develop new models corresponding to given situations. At this point the paper shows the specific results in which the parameters of the proposed model are successfully estimated with some additional processing.

The issues addressed by MacDonald have implications for training and the skills needed by social science researchers when they consider undertaking research using official statistical micro data in New Zealand. The following points will support the argument of this paper: First, as decentralisation is taking place in every area, many social science researchers are becoming concerned about local or special field studies that need the use of official micro data. In general, statistical acts are preventing them free usage of such micro data and in addition it needs very complicated treatment. It takes a long time for a researcher to get permission to use official statistics for private purposes. They are also not likely to understand the environment of micro data and the state sector in the course of study.

Secondly, since the rapid development of network systems has changed modern industrial society structurally, many social scientists need to analyse these new mechanisms with official micro data. They need a new kind of statistical micro data reflecting new trends. At the same time, the necessity of a new kind of official data is becoming greater. Even local governments also are responsible for providing such

official micro data. In all these respects the paper contributes to establishing practical and educational principles for researchers to share official micro data.

2. TRAINING OF RESEARCHERS IN JAPAN

Finally I would like to include some comments on the situation of Japanese statistics education partly because this meeting is being held in Tokyo. Since the Japanese school system has been regulated by the Ministry of Education teachers and schools have not been able to teach much statistics. Therefore, when we talk about improving for statistics education, most of our concerns have likely fallen into the area of system-regulated strategies, such as how to get the contents of statistics included into the guideline of the Ministry of Education, or how to mix statistics with mathematics in university entrance examinations in order to make statistics more important in the teaching of natural sciences. In addition to these organisational limitations, we have another handicap for promoting statistics education, that is, no university has had a department of statistics.

So we have left the education of statistics to private and official sectors. Consequently, each statistics researcher has had little opportunity or encouragement to present his or her experiences that may be useful for better teaching. However, now we are looking forward to a decentralised society and the information industry, which needs the concentration of hard efforts of young researchers. Considering this background, I would like to proclaim that this meeting will help establish a new trends in Japanese statistics education.

REFERENCES

Galmacci, G. (2001). The impact of the internet on the training of researchers. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 159-169). Granada: International Association for Statistical Education and International Statistical Institute.

Lee, T. R. (2001). Teaching biostatistics for medical personnel with computer based supplement. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 139-145). Granada: International Association for Statistical Education and International Statistical Institute.

McDonald, S: (2001). Practical and educational problems in sharing official microdata with researchers. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 119-128). Granada: International Association for Statistical Education and International Statistical Institute.

Shimada, T. (2001). Precaution against errors in using statistical software. In C. Batanero (Ed.), *Training Researchers in the Use of Statistic* (pp. 129-137). Granada: International Association for Statistical Education and International Statistical Institute.

Stangl, D. (2001). Design of an Internet course for training medical researchers in Bayesian statistical methods. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 147-157). Granada: International Association for Statistical Education and International Statistical Institute.

West, W., & Ogden, T. D. (1998). *Journal of Statistics Education* 6(3), http://www.amstat.org/publications/jse/v6n3/west.html

Professor Kensei Araya
Fukushima University
7-9 Higyodan Kitasawamata,

PART 4

NEEDS AND PROBLEMS IN TRAINING RESEARCHERS IN SPECIFIC AREAS

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 177-195.

Printed in Granada, Spain.

JOHN HARRAWAY, BRYAN MANLY, HILARY SUTHERLAND,
& ALLAN MCRAE

MEETING THE STATISTICAL NEEDS OF RESEARCHERS IN THE BIOLOGICAL AND HEALTH SCIENCES

The results of a survey on the use of statistics in research in five subject areas representative of the biological and health sciences are reported. The main component of the survey is a review of statistical methods in 2927 research papers published during 1999 in 16 high impact journals from botany, ecology, food science, marine science and nutrition. A factor analysis establishes that research papers in the different subject areas use different methods. The opinions of research staff and postgraduate students working in these areas are also reported. To support these opinions we provide details of five postgraduate studies involving advanced statistical analyses, which have either resulted in publication or should result in publication in the near future.

Discussion develops recommendations about topics important in a statistics curriculum for research students, where statistics courses should be taught, what is needed in terms of level of theory, the use of short courses and workshops, and the value of project work.

1. INTRODUCTION

Serious problems for research students in the sciences often revolve around inadequate training in study design principles and statistical methodology, both essential for planning research and analysing data, which are frequently collected in the field over long periods of time.

In many cases the research student has attended a first course in traditional statistical procedures involving data summary methods, the binomial and normal distributions, *t* tests, simple linear regression, correlation, and the analysis of variance. But these methods alone are insufficient for the scientists of today and frequently are taught early in an undergraduate degree with the result that many of the ideas are forgotten by the time they are needed.

This paper identifies methods currently being used by scientists through a survey of 2927 research papers published in 16 high impact journals during 1998 and 1999. The opinions of PhD students at the University of Otago are also canvassed and five of their case studies, which are leading to publication, are reported to provide support for the results of the survey.

The concluding discussion will describe how the statistical needs are presently being met and will make a number of additional recommendations about how to develop statistics training for researchers beyond the standard procedures in a first course.

2. LITERATURE REVIEW

Several studies, both general and subject specific, have identified curricula content for statistics programmes designed for researchers and graduate students.

Brogan and Kutner (1986) recommended a two-course statistics sequence with graduate students needing to learn about general linear models including multiple linear regression, analysis of variance and categorical data analysis as well as the use of statistical computer packages.

Elmore and Woehlke (1988, 1998) have focussed on the education research area by conducting statistics content analysis of 1906 papers published in three education research journals covering the period 1978 to 1997. Excluding descriptive statistics, the predominant methodology used was analysis of variance and covariance in 12% of the papers surveyed, followed by multiple regression (8%), multivariate procedures including factor analysis and cluster analysis (8%), qualitative methods (5%), correlation (5%), meta analysis (5%), nonparametric procedures (4%), *t*-tests (4%) and structural equation modelling (3%). The use of qualitative methods and graphical methods has increased substantially over the last ten years in these education journals.

Blumberg (2001) reports on the training of special education teachers in the use of statistics. In addition to the topics reported by Elmore and Woehlke for education research, Blumberg identifies the single-subject design, important because small samples are inherent in studies in special education due to the rarity of many disabilities.

Aitken et al (1990) investigated graduate training in statistics in psychology by surveying 222 psychology departments granting doctoral degrees. The majority of the departments offered courses including analysis of variance, multiple regression, and multivariate analysis of variance, which were viewed as the traditional “old standards of statistics”. “More advanced statistical considerations” including causal modelling, time series, power analysis and repeated measures using multivariate procedures were newer but used less frequently.

They also defined a concept of “in depth” which implied that students could perform an analysis in question themselves. Generally, new PhD’s were competent to handle the traditional techniques but not the newer ones. This idea will be supported by our investigation, which suggests that much of the research in the subjects we have selected can only be successfully completed by enlisting professional statistical advice.

Hammer and Buffington (1994) in a study which is not dissimilar to ours reviewed 1062 articles published during 1992 in six veterinary journals finding that 51% did not include statistical analyses or used only descriptive statistics. They made no attempt to determine whether statistical tests were used correctly and this is the case with our study. The most commonly used tests were analysis of variance and *t*-tests which along with contingency tables, nonparametric tests and simple linear regression permitted access to 90% of the veterinary literature surveyed.

Altman (1998) reports that authors writing for medical journals have poor statistical knowledge and many articles are error prone in relation to statistical content. He comments that there is now much greater use of methods such as logistic regression, proportional hazards regression for survival data, the bootstrap, Gibbs sampling, generalised additive models, classification and regression trees, generalised estimating equations, multi-level models and neural networks with this trend towards greater complexity continuing.

He emphasises the role of the journal editors and statistical referees in the review process. Curtis and Harwell (1998) make this point as well. We also believe this effect

has an influence on the extent of occurrence of statistical procedures in the journals that we have reviewed. In particular it will be noted that the *Journal of Vegetation Science* has a different pattern of statistical use when compared with the *New Zealand Journal of Botany* and *Plant Physiology*.

Curtis and Harwell (1998) report the results of a survey of 27 quantitative methods programs in the field of education producing results similar to those of Elmore and Woehlke (1988, 1998). The majority of doctoral students receive training in the "old standards" analysis of variance, multiple regression, and traditional multivariate procedures as well as logistic regression and log-linear models. Nonparametric procedures are now less prevalent. In terms of recent procedures the majority of programmes train their students in meta analysis and structural equation models while some of the programmes include computer intensive methods and multilevel models.

They also report that students have "strong" skills in standard data analysis packages such as SAS and SPSS, but data base management and programming skills are more varied. Mathematical statistics training was ranked midway between "weak" and "strong". Another line of research suggested was a study of the opinions of doctoral students themselves regarding the adequacy of their own training and this issue is taken up in our paper by surveying students at the University of Otago and describing several recent studies.

3. METHODS

The 16 high impact journals for our survey were selected using *the Journal Citation Reports* (Science Edition) and were confirmed by the opinions of University of Otago staff in the five target departments who publish in these journals. Altogether, 2927 papers published in late 1998 or during 1999 were reviewed and details of statistical procedures present in each paper were recorded.

The journals surveyed, with numbers of papers analysed, are summarised in Table 1, which also includes the percentages of papers in each journal using some form of statistical analysis. For example, ninety five percent of the papers in the *Journal of Vegetation Science* and eighty two percent of all papers contained statistical methodology.

The methods and tests used in each paper were classified initially into 60 classes, but these were subsequently pooled resulting in the fourteen major categories listed in Table 2. A more detailed list of topics is in Appendix 2.

Several decisions were made when classifying the topics. In the analysis of variance category no attempt was made to distinguish between experimental designs like split plots which arose in the biological journals and cross over designs which were common in the nutrition journals. Prospective or retrospective, cross sectional or longitudinal, cohort or case-control studies and the principle of blinding occurred in many nutrition journals, but no attempt was made to differentiate between these and the random, transect or stratified sampling schemes in the other subjects.

Circular data were classified as spatial. Although meta analysis was developed in the social sciences, and more recently has been used in ecology, the few instances of meta analysis were classified as medical statistics on the grounds that the technique is now used quite widely in the health sciences.

Table 1: Summary of Papers Analysed

	Journal	Subject	Number of Papers	Percentage with statistics
a)	Journal of Vegetation Science (Dec 98 – Oct 99)	Botany	88	95
b)	NZ Journal of Botany (Dec 98 – Sept 99)	Botany	63	35
c)	Plant Physiology (Dec 98 – Nov 99)	Botany	446	51
d)	Oikos (Nov 98 – Oct 99)	Ecology	242	72
e)	Journal of Ecology (Dec 98 – Oct 99)	Ecology	89	89
f)	Ecology (Dec 98 – Oct 99)	Ecology	246	96
g)	Journal of Dairy Science (Nov 98 – Oct 99)	Food Science	317	86
h)	Journal of Sensory Studies (Dec 98 – Sept99)	Food Science	28	89
i)	Food Quality & Preference (Jan 99 – Sept 99)	Food Science	237	84
j)	Journal of Food Science (Nov 98 – Oct 99)	Food Science	51	92
k)	Marine Environmental Research (Feb 99 – Dec 99)	Marine Science	54	93
l)	Canadian Journal of Fisheries & Aquatic Sciences (Oct 98 – Sept 99)	Marine Science	213	96
m)	Marine Ecology Progress Series (Oct 98 – Nov 99)	Marine Science	359	87
n)	American Journal of Clinical Nutrition (Nov 98 – Oct 99)	Nutrition	241	96
o)	European Journal of Clinical Nutrition (Oct 98 – Sept 99)	Nutrition	146	95
p)	British Journal of Nutrition (Sept 98 – Nov 99)	Nutrition	107	95
	Total		2927	82

Table 2. Description of Categories

1	Analysis of Variance (anova)
2	Posthoc Tests and Contrasts (posthoc)
3	Simple Regression and Correlation (simpreg)
4	Regression and Modelling (regmodel)
5	Logistic Regression (logreg)
6	Contingency Tables and Log Linear Modelling (loglin)
7	Multivariate Methods (multivar)
8	Basic Tests and Procedures (basic)
9	Medical Statistics (medstat)
10	Population Estimation (popestim)
11	Spatial Analysis (spatial)
12	Computer Intensive Methods (compinte)
13	Stochastic Processes (stochpro)
14	Field Specific (fieldspe)

Each journal had a small number of specialised methods, which would not generally be taught in a general statistics paper, and these were classified as field specific; two examples were diversity analysis and genetics. A small number of papers involving some form of mathematical analysis were excluded on the grounds that these methods would not be taught in a statistics paper.

The category classifications made for *Plant Physiology* by one of the authors were made also by a second author with agreement being found between the two classifications thus confirming classification reliability. Counts were made of the occurrences of the 14 categories in each of the 16 journals and the counts were converted to the percentages of the total number of papers for each journal shown in Tables 3 and 4.

For example, of the 88 papers analysed in the *Journal of Vegetation Science*, 38% involved some analysis of variance procedure, 24% involved a post-hoc analysis and so on for the remaining entries in the table.

Table 3: Percentage Occurrence of Topic Category in each Journal

Journal	anova	posthoc	simpreg	regmodel	logreg	loglin	multivar
1	38	24	42	45	11	26	44
2	13	6	13	8	5	3	3
3	8	9	14	16	0	1	5
4	36	12	31	37	6	16	12
5	38	18	47	46	17	12	17
6	52	26	55	64	7	18	19
7	44	30	29	36	7	5	5
8	50	21	25	21	0	0	50
9	67	33	22	25	0	20	47
10	45	40	23	41	0	2	5
11	39	26	37	37	2	8	31
12	41	22	44	61	5	16	17
13	43	23	42	46	3	11	12
14	54	27	53	43	12	20	4
15	38	14	42	42	10	16	2
16	58	33	27	27	5	6	1

Table 4: Percentage Occurrence of Topic Category in each Journal (continuation)

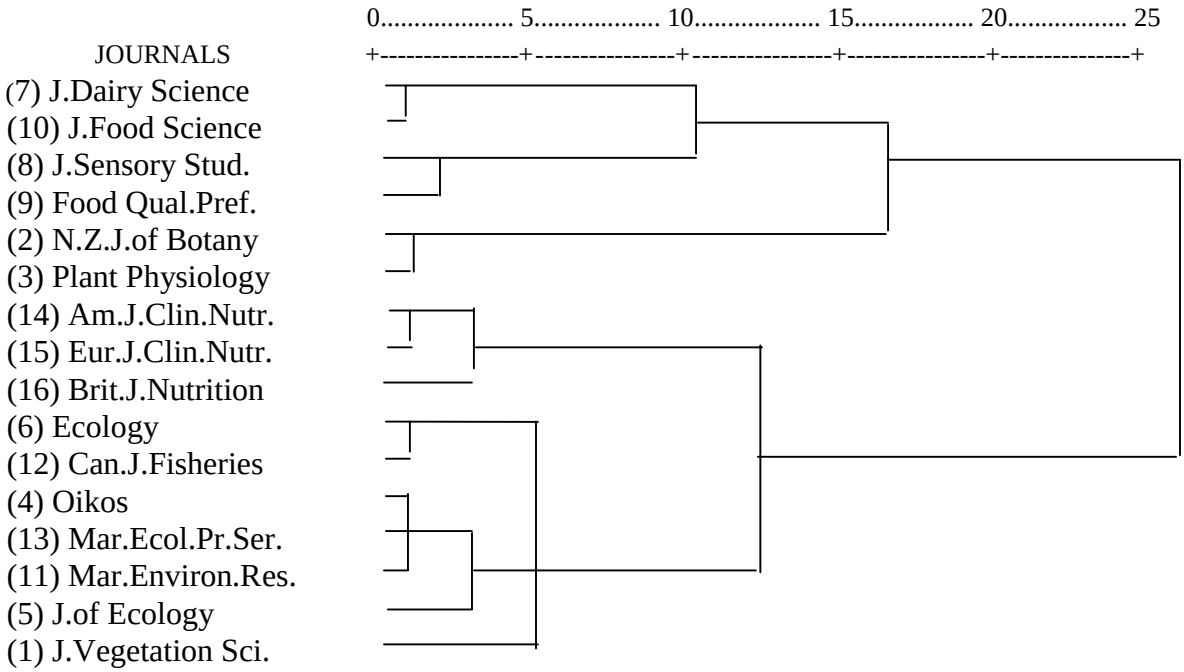
Journal	basic	medstat	popestim	spatial	compinte	stochpro	fieldspe
1	42	1	6	37	26	1	3
2	13	2	8	5	0	0	7
3	20	0	0	0	1	0	6
4	41	1	7	8	9	3	2
5	27	0	6	22	10	6	7
6	54	1	14	13	20	4	1
7	21	1	0	0	3	0	3
8	32	0	0	0	11	0	4
9	41	0	0	0	6	0	4
10	16	0	0	0	2	1	1
11	50	0	7	7	6	2	2
12	49	0	8	3	27	5	4
13	51	0	3	3	10	3	8
14	71	5	0	0	2	0	2
15	73	14	0	0	1	0	1
16	64	6	0	0	0	0	3

4. RESULTS

Cluster analysis using Ward's method identified four clusters of journals shown in the dendrogram in Figure 1.

Cluster 1 included four Food Science Journals, cluster 2 two of the Botany Journals, cluster 3 the Nutrition Journals and cluster 4 the Ecology and Marine Science Journals as well as the *Journal of Vegetation Science*. Both multidimensional scaling and correspondence analysis produced similar journal groupings, but a principal component factor analysis is chosen for further analysis since it shows clearly relationships between the journals and topic categories.

Figure 1: Dendrogram, Ward's method, for 16 Journals



The principal component analysis on the correlation matrix for the 14 categories produced four components explaining 80.74% of the total variance in the percentages. The plot of the scores of principal component one against principal component two is shown in Figure 2.

This confirms the same journal groupings identified in the dendrogram. There is a clear grouping of nutrition journals with high scores on principal component two, and a grouping of food science journals with negative scores on principal component one. The *NZ Journal of Botany* and *Plant Physiology* with low scores on both principal components one and two are similar but the *Journal of Vegetation Science* is more like the ecology and marine science journals, all having positive values on principal component one.

This, we believe, reflects the influence of at least one of the editors of the *Journal of Vegetation Science* who is known to have statistical interests.

A varimax rotation produced the four common factors with loadings greater than |0.100| listed in Table 5. Factor one, explaining 26.0% of the total variance in the

percentages in Tables 3 and 4, is a measure of the extent of regression and modelling techniques, computer intensive methods and population estimation procedures.

Figure 2: Plot of 16 Journals for the First two Principal Components

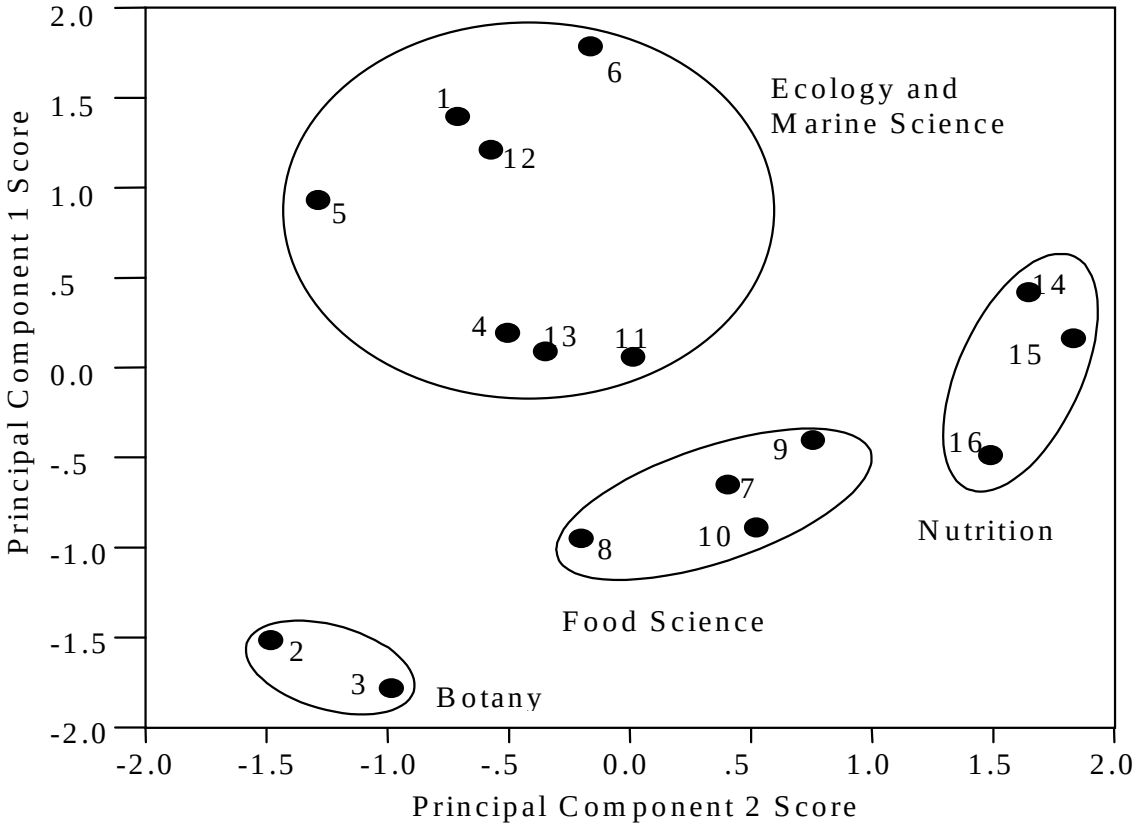


Table 5: Factor Loadings

	Factor			
SUBJECT	1	2	3	4
STOCHPRO	.934			
REGMODEL	.819	.331	.400	
POPESTIM	.756		-.248	.223
SIMPREG	.682	.595	.276	.102
COMPINTE	.674			.602
MEDSTAT	-.341	.827		-.319
BASIC		.766	.421	
LOGREG	.401	.698	-.252	
LOGLIN	.344	.626	.198	.519
POSTHOC		-.144	.895	
ANOVA		.161	.890	.182
FIELDSPE		-.375	-.596	
MULTIVAR		-.235	.239	.881
SPATIAL	.446	.205	-.290	.684

Factor two, explaining 20.7% of the total variance, reflects basic testing procedures, medical statistics, logistic regression and log linear modelling. Factor three, explaining 19.2% of the total variance, is a measure of the extent of analysis of variance and post-hoc procedures with few field specific topics while Factor four, explaining 14.9% of the total variance, reflects the extent of multivariate methods, computer intensive methods and spatial data.

The factor scores for the 16 journals are listed in Table 6. The *Journal of Vegetation Science* has an emphasis on multivariate, computer intensive and spatial analyses with the *NZ Journal of Botany* and *Plant Physiology* having an emphasis on field specific topics, population estimation and spatial analyses with lesser use of analysis of variance and post-hoc procedures.

The *Journal of Ecology* and *Ecology* have emphasis on simple regression, modelling, computer intensive techniques, population estimation and spatial analyses with a lesser emphasis on these topics in *Oikos*.

The *Journal of Dairy Science* has little use of population estimation, spatial analyses and multivariate procedures. The *Journal of Sensory Studies* and *Food Quality and Preference* are low on population estimation, and computer intensive procedures but have extensive use of multivariate methods. *The Journal of Food Science* is high on analysis of variance and post-hoc procedures.

The marine science journals show moderate use of all the techniques identified but the papers in the *Canadian Journal of Fisheries and Aquatic Sciences* has extensive use of simple regression, modelling, stochastic processes and population estimation.

The nutrition journals, in particular the American and European Journals have extensive use of medical statistics procedures, logistic regression and log linear modelling as well as the basic testing procedures.

The occurrences of the different techniques in the five subject areas provide some clues to topics that should be emphasised in a postgraduate statistics curriculum for each of the subjects.

Table 6: Factor Scores for the Sixteen Journals

Journal	Factor 1	Factor 2	Factor 3	Factor 4
1	0.11	0.85	-0.46	2.79
2	-0.69	-0.54	-2.28	-0.33
3	-0.89	-0.92	-1.59	-0.60
4	0.39	0.17	-0.55	-0.01
5	1.48	0.14	-1.14	0.14
6	1.83	0.27	0.71	0.09
7	-0.32	-0.47	0.37	-0.74
8	-1.09	-1.05	0.37	0.98
9	-1.24	-0.57	1.27	1.42
10	-0.09	-1.31	1.23	-1.10
11	0.16	-0.41	0.41	0.18
12	1.73	-0.34	0.32	-0.18
13	0.70	-0.44	0.01	-0.46
14	-0.31	1.63	0.72	-0.51
15	-0.91	2.50	-0.25	-0.82
16	-0.87	0.48	0.86	-0.86

5. THE STAFF SURVEY

The survey of research scientists and staff in the five target departments had 46 respondents. The resident department of each respondent was unfortunately not identified, and therefore the results and conclusions cannot be subject specific. The first question asked staff to nominate, from a given list, topics that they use in their research and topics that they feel their graduate students should know how to use.

The percentage use of each statistical method is presented in Appendix 1. As expected there is a significant correlation between the techniques used by supervisors and techniques seen to be important for students (p-value < 0.001). There is high use of the "old standard" methods involving t tests, ANOVA, simple linear regression, correlation, confidence intervals and multiple comparisons while the more specialised or theoretical methods including matrix methods, time series, survival analysis, canonical correlation, correspondence analysis and multidimensional scaling are less frequently used.

70% of the respondents believe that postgraduate students have inadequate statistical background before embarking on their research. To overcome this problem, 93% feel postgraduate students should attend intensive postgraduate workshops, 26% feel that completing more undergraduate papers could be worthwhile and 30% feel that relying on their own reading during their research could be beneficial. The package Excel is used by 65% of the staff with SPSS (24%), MINITAB (30%) and SAS (20%) being the other most used software.

6. THE STUDENT SURVEY AND CASE STUDIES

There were 43 responses to the postgraduate survey with five additional students being interviewed about their research to illustrate the extent of the problems that arise. The results showed that 86% of the postgraduate students had taken at least one statistics paper as part of their undergraduate degree and 49% had taken at least one mathematics paper. Those who took a mathematics paper were more likely to have taken a statistics paper as well.

Courses or workshops were attended by only 35% as part of their postgraduate study, and 23% did not believe a workshop would be helpful. Of those who would like to attend a workshop/course, modelling techniques and multivariate statistics were the most common topics of interest. 44% of students had training in the use of computer packages before commencing postgraduate study. Excel (60%) was the most common package used in research followed by SAS (19%) and SPSS (26%). About one quarter only of students felt that their knowledge of statistical procedures was adequate for their research.

Several recent studies, which typify the problems confronting postgraduate students in the subjects being investigated in this paper, are described. Each study has involved substantial data analysis assistance from professional statisticians. In each case the analysis has resulted in or is leading to publication.

Case Studies 1 and 2: Marine Science

A biology student with background in algebra, calculus and some introductory probability completed papers for a Master of Marine Science degree, which included a

course on statistics for marine scientists, before commencing field work over two summers on the behaviour of Hector's dolphins in a bay in the South of New Zealand. His research led to the publication of two research articles.

The first (Bejder et al 1998) investigated the association pattern of Hector's dolphins using the Half-Weight Index as a measure of association of two animals. A randomisation test was used to test the extent to which the observed association index values differed from those of a randomly associating population. Association patterns were also investigated using cluster analysis, multidimensional scaling and minimum spanning trees.

The second article (Bejder et al, 1999) reported results from tracking by theodolite the path of a dolphin pod and a tour boat over a period of several hours on 24 days. The direction of movement relative to the boat during the time the boat was in the bay was recorded at 40 second intervals.

In Table 7 the 853 bearings relative to the boat are classified according to the time the boat was present. The proportions of boat directed movements at different times were modelled using logistic regressions. Deviance differences and the Akaike Information Criterion (AIC) established the best model was quadratic in time. This showed the dolphins were attracted to the boat for about 50 minutes after which they lost interest. This conclusion has been used by the Department of Conservation in New Zealand to limit the times tour boats may be in the vicinity of the dolphins.

A second marine science doctoral student with prior training only in basic statistical methods collected data on the presence and absence of Hector's dolphins at 980 locations around the South Island of New Zealand. The sea surface temperature, water depth, water clarity and season of the year were measured at each location sampled.

Table 7: *Bearings Towards Boat Classified by Time*

Time into encounter (min)	Total (n)	Frequency towards boat
0-10	94	32
10-20	133	48
20-30	168	50
30-40	133	52
40-50	111	38
50-60	98	27
60-70	59	16
70-80	34	3
80-90	23	3

The purpose of the study was to establish the preferred environment for dolphins. This entailed fitting habitat selection models using AIC and deviance differences.

A serious problem, which arose with this study, was the result of an inadequate sampling protocol. Initial consultation with a statistician had not taken place. Fortunately, it has been possible to refine the sample and the results are now likely to lead to publication. Although the two students were excellent biologists with good data base management skills, deficiencies in advanced statistical methods, in particular randomisation procedures, multivariate procedures, advanced regression modelling and principles of sampling design would not have allowed completion of their work without consultation with statisticians.

Case Study 3: Food Science

A doctoral student who is a chemistry graduate with first year general mathematics, some advanced calculus but no statistics is working in the area of sensory science involving assessment of taste and smell.

Her research involves training panelists to evaluate mixtures of compounds, with each compound at low, medium and high levels. There are several odour attributes associated with each compound, and the panelists evaluate the mixtures and "score" each of the attributes on unstructured line scales.

Response surfaces for each of these attributes are produced and these indicated how the particular attribute varies with changes in concentration of mixture components. The panel evaluates some intermediate points not used to generate the response surfaces, and checks that these lie on the response surfaces, the idea being that if they do, the response surfaces could be used predictively to establish the attributes for any mixture of these given compounds.

A second study is looking at the perception of trained panelists versus untrained people and the effect on the task they are asked to do – having them evaluate mixtures according to a list of attributes is different to what we do naturally when we consume a food or smell a smell. The trained panel's results from this attribute-based assessment are compared with their results from a task, which involves judging only similarity of mixtures.

An untrained panel perform the similarity task and compare their results with the trained panel. The actual task of the panelists is to select the most similar pair from a set of three samples – so for example in the first experiment there are two compounds either absent, or at low, medium and high levels, giving nine samples. There are 84 possible combinations of these nine samples, and all the panelists see these 84 combinations. These data are analysed using multidimensional scaling.

In interview this student felt she would have benefited from taught work on study design to help formulate the questions to ask a statistician, a rundown on available methods with their strengths and pitfalls, a rundown on appropriate statistical packages, and an opportunity to bring actual data along to any workshop aimed at teaching particular aspects of statistical methodology.

Case Study 4: Nutrition

The nutrition doctoral student involved in this study had completed only an introductory business statistics paper before commencing her research.

Her problem (Taylor et al, 1998) was to assess the relative ability of body mass index (BMI), waist girth and waist hip ratio (WHP) to correctly discriminate between subjects with low and high trunk fat mass or central fat mass as measured using dual-energy X-ray absorptiometry (DEXA).

The subjects consisted of 96 adult women. The DEXA measurements of trunk fat mass and central fat mass were the reference measurements and the anthropometric techniques were the screening tests. Studentized residuals were constructed for all variables. The 75th percentile for each age-adjusted DEXA measurement defined a subject as truly positive. Sensitivity and specificity of 19 percentile cut-offs were calculated for each screening measure and receiver operating characteristic curves (ROC's) were constructed. A bootstrapping procedure in STATA was used to calculate the areas under the curves, the 95% confidence intervals and the differences between the areas.

In a second study (Taylor et al, 2000) a similar analysis was conducted except the conicity index instead of BMI was evaluated in a large group (n = 580) of children and adolescents. Internal z scores were created for each variable. A z score ≥ + 1 identified a subject as truly +ve. Positive and negative likelihood ratios were also calculated in this paper.

Generally in her research this student had used log transformation, t-tests, one-way ANOVA, simple factorial ANOVA, multiple regression analysis, curve estimation, calculation of sensitivity and specificity and construction of receiver operating characteristic curves, calculation of internal and external z scores, positive and negative predictive values, cross-tabulations and chi-squared tests. The SPSS package had been used for most of the calculations. In interview the student commented that it is crucial to increase the statistics knowledge of those involved in research. Although it is possible to learn from trial and error it is a very steep learning curve.

Case Study 5: Botany

This doctoral student had taken no mathematics papers in his undergraduate degree and one introductory statistical methods paper covering the "old standards" which was subsequently supplemented by a further paper on similar material taught within the Botany Department. All the statistical analyses were performed using the programs Teddybear, Statistix and Glim.

The PhD topic was "The comparative ecology of rare and common *Aceana* and *Chionochloa* species". The work comprised glasshouse and garden experiments, where species were grown from each genus together, subjecting them to various treatments relating to intrinsic growth and reproductive traits, and taking account of responses to stress factors and competition.

Experimental responses were analysed using analysis of variance, and correlated response means with the geographic range sizes of the species, separately for each genus. Looking at the patterns within each genus reduces phylogenetic dependence in the data, since many phylogenetically correlated traits 'cancel out' when comparing congeners. Some species were represented by two populations, and because these were non-independent the mean value for each of these population pairs was used in the correlations. Consistent differences between rare and common species were being looked for. Regression methods were used frequently.

A typical experiment comprised replicated randomised blocks of species being treated with one or more controlled factors. For one experiment where the treatments were difficult to apply at the level of single plants, a split-plot design was used. The number of controlled factors and their levels varied. Examples of some of the designs used for *Aceana* are presented in Table 8.

A variety of problems arose with the analyses. There was almost always a small amount of mortality (unrelated to treatments) among the experimental plants, raising complications when it came to statistical analysis, as it was always desirable to extract a block effect. Either the block effect was included in the error term, with a reduction in power, or "missing values" were estimated. In other experiments, the mortality response where it differed between treatments was analysed. Because mortality is an all-or-nothing event, the data had a binomial error distribution. In these cases Glim was used to perform generalised linear modelling, specifying a binomial error structure.

In interview the student suggested the best way to approach the presentation of statistics to botany research students is to give examples of how poor experimental design can invalidate results. He saw two important issues to be pseudoreplication (in

the field and with growth cabinets), leading to non-independence of data, and inadequate randomisation, opening up the possibility of biased treatments. The usefulness of randomised blocks in reducing noise (e.g. if plants vary in size, associate difference in size with the different blocks) should be stressed.

Table 8: *Examples of Experimental Designs Used*

Experiment	Growth rate	Reproduction	Nutrients	Shade	Competition
Design	completely randomised (in growth chamber)	randomised block	fully factorial, randomised block	split-plot	fully factorial, randomised block
Treatments	13 species	13 species	13 species	12 species (subplot)	13 species
	10 replicates	4 replicates	N(3 levels)	Shade (5 levels) (plot)	2 fertility levels
			P(3 levels)	4 replicates (plot)	2 competition levels
			5 replicates		6 replicates

7. MEETING THE STATISTICAL NEEDS

At the present time two papers outside the usual undergraduate statistics programme are provided at the University of Otago for postgraduate researchers in the biological and health sciences.

The first paper, for marine and environmental scientists, has as prerequisite an introductory methods paper including the “old standards”. It has 72 teaching contact hours and covers sampling and estimation (3 weeks), testing hypotheses (2 weeks), regression methods (2 weeks), analysis of variance (2 weeks), BACI designs (2 weeks), power analysis and sample size determination (2 weeks), time series (2 weeks), environmental monitoring (2 weeks), impact assessment (2 weeks), spatial data (2 weeks), bioequivalence (2 weeks) and censored data (2 weeks).

The purpose of the paper is to introduce students to the theory and methods of sampling and modelling in the context of environmental science. The emphasis is on correct and justifiable procedures rather than statistical theory. Project work is used to promote “learning by doing”.

For example, one project involves estimating the biomass of a shell fish in a tidal inlet. The students design a stratified sampling scheme, carry out the sampling and find an estimate of the biomass from the data. An appropriate text for the course is available (Manly, 2001).

The second paper is for postgraduate students who have knowledge of the “old standards” but who need the more advanced analysis methods routinely used in the health sciences. It has 52 teaching contact hours with less emphasis on statistical theory but more emphasis on the relationship between statistical methods and the scientific inference.

From a statistical viewpoint the paper covers regression methods including the generalised linear model (linear, logistic, Poisson regression) and the Cox regression

model. From the scientific aspect the course covers definition of outcome measures, transformations, robustness of linearity assumptions, scientific aspects of model selection, and model checking. The paper relies on several data sets providing examples from a range of health science fields. The class is attended by clinical researchers and statisticians training to be biostatisticians. Bangdiwala (2001) points out that such interaction between these groups is a valuable learning experience for all involved. Aspects of the INCLEN programme described by Bangdiwala are also included in this paper.

8. DISCUSSION

It is apparent from the surveys presented above that a great deal of statistics is used by researchers in the areas of science covered, but the important techniques vary to some extent between, and even within disciplines. Given the time constraints on these researchers, it is apparent that although basic statistics (the 'old standards') can be taught in generic courses, it is highly desirable that advanced learning is through approaches that are targeted to the needs in different areas.

For example, we see that a botany student interested in studying the ecology of plant species in a quantitative manner, in the style of papers in the *Journal of Vegetation Science*, needs to have a good knowledge of multivariate and spatial analyses and computer intensive methods in particular. This might have to be provided by giving less emphasis to analysis of variance, regression methods and other topics. On the other hand, a nutrition student mainly needs a good knowledge of analysis of variance and regression methods.

There are three basic approaches that can be used to provide starting researchers with the statistical skills that they need:

a) They can learn informally from their own reading, with assistance from established researchers;
b) They can attend specialist short courses and workshops provided by statisticians or by established researchers in their disciplines; or
c) They can attend formal lecture courses, again either provided by statisticians or established researchers.

At present (a) seems to be the most commonly used approach. However, we believe that this is not altogether satisfactory because it may result in the researchers having gaps in their knowledge that they are not even aware of.

The best approaches seem to be (b) and (c), or perhaps a combination of these. If these are used then we believe that it is absolutely crucial that there is the involvement of statisticians who are very familiar with applications of statistics in the subject area. Established researchers may be well informed about current uses of statistics, but they will rarely be able to see how these relate to developments of statistical methods in general.

The danger with just using established researchers is therefore that important new developments in statistical methodology will be ignored. The situation is no better if statisticians who are not familiar with the subject area do the teaching. In that case they will not be able to judge very well what topics are important, and probably will bore

their audience with irrelevant examples.

At the University of Otago, approach (c) has been used with success with courses for biological and health science students. We believe that it has been successful largely because the statisticians involved are very experienced at consulting and research in the subject areas.

Statistics staff have also provided a number of short courses on specialised areas of statistics for researchers. These have been intended primarily for scientists outside the university, but research students and university staff have been able to attend. These courses seem to fill a real need for the scientific community by providing an update on the latest developments in a restricted area of statistics. We see such short courses as complementing rather than replacing other modes of instruction.

Although there have been successes at the University of Otago, this does not mean that the situation is satisfactory. Of the staff in our survey, 70% thought that students were inadequately prepared when they started research. Of the students, 75% thought that they were inadequately prepared. One reason for this is that most of the students have not taken more than an introductory course in statistics, although more advanced courses that would be valuable for them are available.

One question that has only been considered rather indirectly in our surveys concerns how well the existing methods available for researchers to learn statistics are actually working in the end. Students may be inadequately prepared when they embark on research, but perhaps this is rectified by the time that they are ready to publish their results. Put another way: is the research that is appearing in scientific journals usually based on correct and appropriate statistical analyses?

It is difficult to know for sure what the answer to this question is because it often depends on the nature of data that are not provided in papers, it is not often possible to decide exactly how analyses were done, and in some cases the meaning of 'correct and appropriate' is debatable. However, there is no question that some very experienced consultant statisticians think that there are problems. Altman (1998) expressed this view with regard to medical researchers. Maindonald (1999) puts it quite forcefully:

> *"My reading of published papers persuades me that serious problems with the design of data collection and with data analysis are common. A cursory overview of papers in major international journals may be sufficient to reveal examples of serious statistical misinterpretation,…, there is a case for treating all published analyses as preliminary, pending scrutiny by researchers with relevant statistical skills!"* (Maindonald, 1999)

If these views are correct, which we suspect is the case, then there certainly are real problems at the moment. So how do we meet the needs of researchers in the health and medical sciences?

We suggest that the only really satisfactory approach is for statisticians with relevant consulting experience to develop structured programmes in collaboration with established researchers. These programmes must recognise the way that statistics is currently being used in the discipline concerned, but must also take into account broader developments in statistical methods so that researchers are able to make use of new approaches where this is desirable.

APPENDIX 1

Percentage Use of Statistical Method

Method	Percent Staff Use	Percent Student Use
t test	91	91
ANOVA	91	93
Simple linear regression	91	93
Correlation	91	89
Confidence intervals	87	87
Experimental design	83	74
Sampling/Power	72	72
Graphical procedures	67	63
Nonparametrics	67	74
Multiple comparisons	67	89
Transformations	63	74
Multiple regression	63	72
Survey design	63	61
Goodness of fit	61	63
ANCOVA	59	67
MANOVA	50	59
Contingency tables	46	50
Principal components	43	39
Cluster analysis	33	35
Contrasts	30	30
Bootstrap/Randomisation	30	39
Discriminant analysis	28	24
Factor analysis	28	24
Fishers exact test	26	39
Population estimation	22	24
Time Series	22	15
Logistic regression	22	39
Non Linear regression	20	35
Simulation	17	26
Correspondence analysis	17	15
Maximum likelihood	15	22
Multidimensional scaling	15	15
Loglinear modelling	13	28
Matrix methods	13	9
Canonical correlation	11	9
Jackknifing	11	20
Survival analysis	9	15

APPENDIX 2: CATEGORY DESCRIPTIONS IN DETAIL

1. *Analysis of variance and experimental designs:* Factorial designs. Blocking. Repeated measures. Split plots. Cross over designs. Random and Mixed Models. Approximate *F*-test. Levene's Test. Kruskal-Wallis and Friedman tests. Ante-dependent repeated measures analysis. Q_w test for within group homogeneity.

2. *Posthoc tests and contrasts:* Multiple comparisons. Student-Newman-Keuls. (Fisher's) least significant difference. Tukey. Duncan. Scheffe. Dunn. Ryan. Dunnett. Least square means.

3. *Simple regression and correlation:* Allometry. Fisher's *z* transformation. Durbin Watson. Pearson. Spearman.

4. *Regression and modelling:* Multiple regression. Response Surface Analysis. ANCOVA. Path Analysis. Non-linear regression. Growth curves. Nonparametric regression. Lowess curves. Ridge regression. Classification and Regression Trees (CART) GLM. GAM. Bayesian inference methods. Partial least squares regression. Structural equations. Polyserial correlation coefficients. Kernal density estimators. Linear standard curves. Split line regression. Spearman Karber 3 D plane regression procedure. Longitudinal regression models. Maximum likelihood and REML.

5. *Logistic regression:* Odds ratios. Survival analysis. Proportional hazards models. Cox regression model. GLM. GAM (for this type of analysis).

6. *Contingency tables and log linear modelling:* Fisher's exact test. Goodness of fit; χ^2. Cochran's C test. Mantel test for trend. Mantel-Haenszel method. McNemar's test. Categorical modelling; CATMOD. GLM. GAM (for this type of analysis).

7. *Multivariate methods :*MANOVA; MANCOVA. Box's M. Mauchly; Pillai's trace. Test of sphericity. Hotelling-Lawley. Principal Component Analysis. Factor Analysis; Procrustes analysis. Canonical correlations analysis. Redundancy analysis. Correspondence analysis. Multidimensional scaling. Principal Co-ordinate Analysis. Methods of ordination. Canonical detrended analysis. Cluster analysis. Distance measures and analysis. Bray Curtis. Mantel. Jaccard's coefficient of similarity. Sorensen's index. TWINSPAN.

8. *Basic tests and procedures:* Descriptive statistics. Exploratory data analyses. Coefficient of variation. Skewness, Kurtosis. Profile analysis. Kolmogorov-Smirnov. Anderson Darling. Lilliefors. Simple confidence intervals and *t*-test. Power. Nonparametric tests. Kruskal-Wallis. Kendall. Freidman. Spearman. Rayleigh. Jonckheere's ordered alternatives test. Dixon's test. Likelihood ratio test. G-test. Bartlett. Shapiro-Wilks. Box's M. Mauchley. D'Agestino-Pearson K^2 test. Gini coefficient. Geary's C coefficient. Wald test statistic. Exact significance probability test. Binomial probability test.

9. *Medical statistics:* Epidemiological methods. Methods of Comparison Procedures; Bland and Altman. Measures of Agreement; Kappa; Cronbach's α. Diagnostic testing; specificity and sensitivity. Repeated measurements; AUC (Area under curve) and summary measures method. Relative risk. Odds ratios. Intention to treat analysis. Meta analysis (which developed in the Health Sciences to a large extent).

10. *Population estimation:* Mark recapture. Transect sampling. Stratified sampling.

11. *Spatial analysis:* Ripley's K Method. Diggle's G and F functions. Spatial pattern analysis. Lattice wombling. Morisita's index. Spatial variograms. Semivariogram. Kriging. Spatial autocorrelation. Moran's I autocorrelation. Mantel correlogram. Circular data involving the

William V-test and the Rayleigh test.

12. *Computer intensive methods:* Bootstrapping. Jackknifing. Simulation. Cross validation analysis. Gibb's sampling. ANOSIM. Hardy-Weinberg proportions. Neural networks. Superposed epoch analysis.

13. *Stochastic processes:* Transition matrices. Random walk. Markov processes. Time series. Periodogram. Cross correlation. Random intervention analysis.

14. *Field specific:* For example, Shannon's index, Shannon-Weaver, Shannon-Wiener and Simpson (measures of diversity). Also area of Genetics involving Phylogenetic Analysis. Parsimony analysis. Consensus tree (evolutionary trees in biology). AMOVA – analysis of molecular variance Linkage analysis; Genotype analysis. Sister-taxon comparison.

REFERENCES

Aitken, L .S., West, S. G., Sechrest, L., & Reno, R. R. (1990). Graduate training in statistics, methodology and measurement in psychology. A survey of PhD programs in North America. *American Psychologist*, 45, 721-734.

Altman, D. G. (1998), Statistical reviewing for medical journals. *Statististics in Medicine*, 17, 2661-2674.

Bangdiwala, S. I., & Muñoz, S. (2001). Training of statisticians worldwide to collaborate as co-investigators within country clinical epidemiology units: The experience of the International Clinical Epidemiology Network (INCLEN). In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 265-275). Granada: International Association for Statistical Education and International Statistical Institute.

Bejder, L., Fletcher, D., & Bräger, S. (1998). A method of testing association patterns of social animals. *Animal Behavior*, 56, 719-725.

Bejder, L., Dawson, S. M., & Harraway, J. A. (1999). Responses by Hector's dolphins to boats and swimmers in Porpoise Bay, New Zealand. *Marine Mammal Science*,15 (3), 738-750.

Blumberg, C. J. (2001), Training regular education and special education teachers in the use of research methodology and statistics. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 231-244). Granada: International Association for Statistical Education and International Statistical Institute.

Brogan, D. & Kutner, M. H. (1986). Graduate statistics service courses. *The American Statistician*, 40 , 252-254.

Curtis, D. A., & Harwell, M. (1998). Training doctoral students in educational statistics in the United States: A national survey. *Journal of Statistics Education*, 6(1), http://www.amstat.org/publications/jse/v6n1/curtis.html

Elmore, P. B., & Woehlke, P. L. (1988). Research methods employed in *American Educational Research Journal*, *Educational Researcher*, and *Review of Educational Research* from 1978 to 1987. Paper presented at the *Annual Meeting of the American Educational Research Association*, New Orleans.

Elmore, P. B., & Woehlke, P. L. (1998). Twenty years of research methods employed in *American Educational Research Journal*, *Educational Researcher* and *Review of Educational Research*. Paper presented at the *Annual Meeting of the American Educational Research Association*, San Diego.

Hammer, A. S., & Buffington, C. A. (1994). Survey of statistical methods used in the veterinary medical literature. *Journal of the American Veterinary Medical Association*, 205 (2), 344-345.

Maindonald, J. H. (1999). New approaches to using scientific data - statistics, data mining and

related technologies in research and research training. Included in Section 2 (*Research Standards: Data Collection and Analysis Standards*) of the discussion document on the web at http://www.anu.edu.au/graduate/papers/gs98_2.html

Manly, B. F. J. (2001). *Statistics for environmental science and management*. London: CRC Press.

Taylor, R. W., Keil, D., Gold, E. J., Williams, S. M., & Goulding, A. (1998). Body mass index, waist girth, and waist-to-hip ratio as indexes of total and regional adiposity in women: Evaluation using receiver operating characteristic curves. *American Journal of Clinical Nutrition*, 67, 44-49.

Taylor, R. W., Jones, I. E., Williams, S. M., & Goulding, A. (In press), Evaluation of waist circumference, waist-to-hip ratio, and the conicity index as screening tools for high trunk fat mass, as measured by dual-energy x-ray absorptiometry, in children aged 3-19 years. *American Journal of Clinical Nutrition.*

John Harraway, Bryan Manly, Hilary Sutherland, & Allan McRae
University of Otago, Department of Mathematics and Statistics
P.O. Box 56, Dunedin, New Zealand
Email: jharraway@maths.otago.ac.nz

Bryan F.J. Manly, Biological & Environmental Statistics
Western EcoSystem Technology
Inc. 217 South 1st Street, Suite 5 Laramie, Wyoming 82070, USA
bmanly@compuserve.com

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 197-213.

Printed in Granada, Spain.

DAVID J. SAVILLE

A HANDS-ON, INTERACTIVE METHOD OF TEACHING STATISTICS TO AGRICULTURAL RESEARCHERS

In this paper, I focus on two topics. The first topic is several day-long workshops, which I run annually for agricultural researchers. These cover the "linear model" methods, based upon the normal distribution, which are very commonly used by agricultural researchers. These methods are analysis of variance, regression and analysis of covariance. I describe the manner in which these workshops are run, cover the content of the workshops, describe the course evaluations, speculate upon why the workshops have proven to be popular, and attempt to draw conclusions.

The second topic is a method of teaching linear model theory using N-dimensional geometry. This method has been successfully used for both second-year and graduate-level university statistics courses.

1. INTRODUCTION

I work as a consulting biometrician/statistician in a government-owned agricultural research institute, which deals with animals, pastures, crops, plants, weeds, insects, diseases and other agricultural entities. Most of my work involves helping researchers with the design of agricultural experiments, the analysis of the resulting data, and the interpretation and writing up of the results.

This work involves continual communication between me (the statistician) and the researcher – the better the communication, the better the agricultural research. On the one hand, the research work is prejudiced if its statistical foundation is shaky. On the other hand, a firm statistical foundation is of little assistance if I have misunderstood the objective of the research, and we have answered a question of little interest. The best results are obtained when we each have a basic knowledge of the other's field in addition to specialist knowledge of our own field. That is, the researcher has a basic understanding of the statistical methods being used, and I have a basic understanding of the research area and the objectives of the research. This point is also made by Bangdiwala (2001).

In prioritising my everyday work, I therefore give highest priority to the teaching of statistical ideas to my agricultural research colleagues, to learning about research projects with which I am involved, and to collaboratively working on project designs. Next priority is given to analysis of designed experiments or surveys, followed by interpretation and writing up of such work. Lowest priority is given to attempting an analysis of data from unplanned work.

I teach statistical ideas to my agricultural research colleagues in two main ways. Firstly, I always make time during my everyday consultations to explain any statistical concepts that are relevant to the topic of conversation. Secondly, I run one-day statistics workshops each winter (the period when most researchers are inside the office). These workshops are specifically designed to teach ideas rather than methods of calculation,

and have enjoyed an unusual level of popularity. These workshops are the first main topic of this paper.

Turning to the second topic, my interest in the teaching of statistics led to two periods of university lecturing in 1984 and 1985. The first course in which I was involved (with Graham Wood) was a second-year applied statistics course at the University of Canterbury in Christchurch, New Zealand.

For this course, we developed a novel teaching method that involved the use of *N*-dimensional geometry. This allowed an easy integration of the mathematical basis for linear models with the practical application of the models and the necessary computing. The second course was on "design, analysis and interpretation of experiments" for graduate students in agriculture at the University of Davis, California, U.S.A. For this course, I further simplified the geometric approach.

The enthusiasm of the students for the geometric approach led to Graham and I writing two statistics textbooks (Saville & Wood, 1991; 1996). The first book was comprehensive, and included all of the topics covered in the California course. The second book was written as an easy-reading introduction to the teaching method, and was aimed at lecturers to whom the method was new, as well as a general audience of statisticians and users of statistics.

In this paper I briefly describe my everyday consultations, then focus on (1) the annual workshops to agricultural researchers and (2) the geometric method of teaching statistics.

2. EVERYDAY CONSULTATIONS

Everyday consultations are of many varying types, and cannot be adequately summarised in a short section. Some involve the design of an agricultural experiment for which the primary variable or its transform can be assumed to follow a "normal" distribution. Here the main hypothesis of interest may correspond to a particular "contrast" between the treatment means, and in this case, attention may focus on ensuring maximum precision for the estimate of this contrast.

This may involve choosing an appropriate experimental layout (e.g., a randomised complete block or a "split plot" design), then carrying out "power" calculations to determine the number of replications. Alternatively, if the primary variable follows a binomial distribution, different power calculations may be carried out. During such "design" consultations, I normally discuss the relative virtues of various competing experimental layouts, the level of replication, and points such as whether some treatments should be replicated more heavily than other treatments.

Before consultations concerning the analysis of data, the researcher would normally enter the data into columns of an Excel spreadsheet, and email it to me before visiting my office with a copy of the field plan. We then have a general discussion about how the experiment went, during which the researcher will inform me of any peculiar problems that were encountered and how this may have affected the data. We discuss the list of treatments in relation to the hypotheses of interest, and in the case of "normal" data I write down the contrasts between the treatment means, which we (hopefully) planned prior to the experiment being conducted (most researchers find this specification of contrasts very difficult).

We then discuss the field layout and how this is taken into account in the analysis of variance or covariance (or, occasionally, regression). Usually I copy the data into an appropriate statistical package and process the data. This processing includes routine

assumption checking such as a histogram of the residuals from the model (to check normality) and a plot of the residuals versus the fitted values (to check for heterogeneity of variance). At this point in time unusual values are often identified, leading to checks of the data entry, possibly a transformation of the data using a logarithm or square root transform, and/or discussions about what is an appropriate course of action. Finally, we usually decide upon how to present the treatment means and indication of precision (e.g., least significant difference), whether in a table, a bar graph, a scatter plot, or other option.

Other consultations may involve the analysis of binomial data, either by simple 2 x 2 chi-squared tables, partitioned tables (c.f., contrasts), or using generalised linear models. Alternatively, the consultation may concern how to describe the statistical design and method of analysis in the research report, how to respond to comments of a statistical nature written about a manuscript by journal referees, or a multitude of other possible topics.

3. ANNUAL WORKSHOPS

The primary aim of the annual statistics workshops is to *instill basic statistical ideas* into the minds of the researchers, not to teach them how to manually do arithmetic calculations nor how to process data using statistical packages. Most modern-day researchers are capable of feeding data into a computer package - what many are *not* capable of doing is choosing the correct menu item from the bewilderingly long list which pops up on the computer screen. My aim is to at least educate researchers to the level that they know the choice is important (and therefore ask for help if uncertain as to the correct choice).

The workshops *do* use simple exercises, which involve arithmetic calculations, but each exercise is designed to illuminate a particular statistical idea. The exercises ensure the people attending each workshop are involved in a hands-on, participatory manner. During each exercise, each person randomly selects experimental units (e.g., sheep or plots of pasture) using a set of random numbers, so that each person has their own data set to work with. In effect, each person is doing a “simulation.”

To be more specific, I shall describe the first workshop in some detail (§3.1), and the remaining workshops in brief (§3.2). I then describe how the workshops were invented and how they have since developed (§3.3). I will describe the course evaluations (§3.4), give my philosophical thoughts on what I consider to be the essential ingredients of the workshops (§3.5), and conclude with a few general comments (§3.6).

3.1. WORKSHOP #1 IN DETAIL

Workshop #1 (Basic Statistics/Analysis of Variance, Day 1) starts at “zero assumed knowledge” and finishes by carrying out a simple analysis of variance. It runs for a whole day, from 9 am until 4.15 p.m. Workshop members are rested often (they need it!). Lunch lasts for one hour and there is a 20-30 minute break for refreshments halfway through both morning and afternoon sessions.

The basic equipment which people bring on the workshop is paper, pencil, eraser and battery-operated calculator. In terms of the venue for the workshop, I try to use a room which will comfortably hold enough tables for 20-30 people, allowing each person plenty of table-top space and allowing room between the tables for me to wander

around and help people who are experiencing arithmetic problems.

The workshop commences with introductions. We each say our name, where we work/study, and what our area of research is. Then I put up a "plan for the day" in the form of a flow diagram on a large sheet of paper, which I attach to the wall. I go through the items on the plan and explain that we shall initially go quite slowly for the purpose of consolidating the "basics."

The first exercise is called "Why Use Statistics?" from Bishop (1980). I describe an experiment with 12 sunflowers grown in pots on a windowsill for nine weeks. There are three replicate pots for each of four liquid fertiliser treatments (A, B, C and D), with pots assigned in a completely random manner to treatments. I give the weights of the sunflowers at the end of the experiment. I then ask each class member to think of the three weights for each treatment as a sample of a larger population of weights for other sunflowers grown with the same treatment.

I ask: "*On average over this larger population*, do sunflowers grown with treatment A weigh more than sunflowers grown with treatment B?" Class members are asked to write either Yes, No or Don't know on a sheet of paper, along with reasons for their answer. They are also asked to do the same for the five other pair-wise comparisons (A versus C, A v. D, B v. C, B v. D, and C v. D). After 5-10 minutes to ponder the questions, I ask the class members to swap pieces of paper before I tabulate the responses in terms of Yes/No/Don't know (plus the nil responses). Then I ask people to give reasons for their own answers. This generates discussion of sample means, overlap or otherwise between sample values, variability within each sample in comparison with the size of the differences between sample means, and so on. It also serves to "break the ice" and set the scene of the workshop as a participatory event.

The next section of the workshop is mainly me talking. I give each person an empty workshop folder plus the first 17-page section of a booklet, which summarises the work that we cover during the rest of the day (Saville 1980). Included is a population of the gains in weight of 50 lambs during a three-month period (the data were from lambs 1-50 in an experimental group of 150 lambs in a local New Zealand experiment).

I then describe various alternative ways of measuring the centre and spread of this population of weight gains. For measures of centrality I discuss the median, mode and mean, and for measures of spread I discuss the range, the inter-quartile range, the mean absolute deviation and the mean squared deviation (or variance) plus its square root, the standard deviation. By chance, lamb 50 had a very unusual weight gain, so I discuss the influence of this one weird value on the various measures. I then explain that the methods we most commonly use are based on the mean and variance, and that this is because of the simple mathematical solutions which are possible by courtesy of Sir R.A. Fisher and Pythagoras' theorem in N-dimensional space.

This takes us through to the morning break at about 10.30 am. After the break, at about 11 am, I hand everyone a Xerox copy of some random numbers, and we embark on a sampling exercise. The idea is that we know the "truth" about our population of 50 weight gains in that we know the true mean and true variance; however, we pretend that we have only a small sample from this population, and set about "estimating the truth" by random sampling.

Firstly, this is done for a sample of size three. As in all class exercises, I run through the exercise myself on a white board, then each person does it for themselves. In the current exercise, each class member shuts their eyes, stabs their random number table with a pencil, then takes the next three random numbers in the range 1-50, and writes down the three corresponding weight gains. The sample mean and sample variance (with divisor n-1) are then calculated.

When everyone is finished I do a histogram of the sample means from the class on a "printable" white board. The class discusses how well they did in estimating the true mean. I then do a histogram of the sample variances to the right of the first histogram, and the class discusses its shape and approximate centre. I also make some points about the long-run shapes of the histograms. The class repeats the exercise using a sample of size six. While they are working on the arithmetic, I draw up the outline of the "n=6" histograms, immediately below the "n=3" histograms and using the same scales. Each person then comes up and enters their own value on each histogram. The class then discusses the effect of increasing the sample size on the spread of each histogram and I make some comments upon how the shapes change. I print off the white board (containing the four histograms), we break for lunch, and I Xerox a copy for each person's folder during the lunch hour.

After lunch I consolidate "where we're at" by drawing bell-shaped normal curves to represent the "parent" population of 50 lamb weight gains, plus increasingly narrow bell-shaped curves to represent the two "derived" populations of sample mean weight gains for samples of size three and six respectively.

We then embark upon a further exercise concerning the variability of sample means, using the same population of 50 lamb weight gains. The idea here is to generate several sample means, then directly calculate their variance (thinking of the several means as a sample from the population of all possible sample means). Each class member uses their random numbers to select four samples of size two, and calculates the four sample means and the variance of these four means.

They then add their "estimated variance for means of samples of size two" to a histogram on a fresh page of the printable white board. They redo the exercise for four samples of size four, adding their "estimated variance for means of samples of size four" to a second histogram directly below the first on the same scale. After this, the class discusses their results, hopefully noting that the centres of the histograms differ. After the discussion has dried up, I mention the theoretical "rule" that the variance for sample means of n values is the variance of the parent population divided by n.

To see that this is reasonable, we calculate the theoretical values for n=2 and n=4 and add these to the two histograms. [As an aside, I mention also that on the original scale (kg of weight gain), the "rule" is equivalent to saying that the "standard error (=deviation)" of a mean of n values is the standard deviation of the parent population divided by the square root of n.] I print this page of the white board and Xerox it for distribution to class members.

Throughout the day, I explain that the above exercises are intended to illuminate the ideas that are required for an understanding of an "analysis of variance." We now embark upon two analyses of variance. In both, there are three experimental treatments, each with four lambs randomly assigned to them. The treatments are "untreated," "treated once with anthelmintics for worm control," and "treated twice with anthelmintics for worm control."

The first analysis of variance is a "dummy run" in that each class member draws the three sets of four lambs from the same population (above), thereby simulating an experiment in which all three treatments are equal. An "F ratio" is calculated as the ratio of the "actual" variance of the means (calculated directly from the three treatment means as above) and an "expected" variance of the means.

The latter is the variance of the means which one would "expect" if all three treatment means were identical. It is calculated by calculating the variance *within* each treatment group, averaging these three numbers to obtain the best estimate of the average variance of the three parent populations, then applying the theoretical rule by

dividing by n=4. Note that if there are no treatment effects, the "actual" and "expected" estimates will both average to the same quantity, while if there are treatment effects, the actual estimate will be inflated while the expected will be unaffected; hence the F ratio is increased by the presence of treatment effects.

After completing the necessary calculations, each class member writes their actual and expected variances and the resulting F ratio on a fresh page of the printable white board, and adds their F value to a histogram of F values. We now discuss the class results. It is hopefully clear that the two variance estimates have a similar average, though one is more variable than the other. At this point I introduce the $F_{2,9}$ reference distribution, and hand out a table of F "critical" values which contains its tabulated 95 and 99 percentiles (4.26 and 8.02). The histogram of F ratios generated by the class tells us the approximate shape of the $F_{2,9}$ distribution, and it is interesting to see how many values are outside the "95% normal range" of 0 to 4.26. This page of the white board is again printed and xeroxed for distribution.

After a refreshment break, we embark upon the second analysis of variance, "the Real McCoy," meaning an experiment in which there *are* treatment effects. For this exercise, the first set of four lambs ("untreated") is randomly selected from the population used above. The second set of four lambs is randomly selected from a second population which I provide (the data were from lambs 51-100 in the same experimental group of 150 lambs as described above, except that I added 5 kg to each data value). Similarly, the third set of four lambs is randomly selected from a third population, which I provide (the data were from lambs 101-150 in the same group of 150 lambs, except that I added 10 kg to each data value).

In effect, each person is simulating an experiment in which the second and third treatment means are 5 and 10 kg greater, respectively, than the first treatment mean (these effects are very large in relation to the variability in these data). The same calculations are carried out as described above. In this case, each person finds that the "actual" variance of the mean is many times larger than the "expected" variance of the mean, with the result that the F ratio is large (usually greater than 10-20) and usually "highly significant" statistically. As above, each class member writes their actual and expected variances and the resulting F ratio on a fresh page of the printable white board, the class results are discussed, and a xerox is made for distribution.

To wrap up the day, three things remain to be done. Firstly, I write out the more usual "analysis of variance table" for the Real McCoy data set which I had worked through on the white board prior to the class doing their own simulation. The purpose here is to relate the class method to the more usual computer printouts. Here the usual "treatment" and "error" mean squares are obtained by multiplying the "actual" and "expected" variances of the mean by the sample size (n=4 here). Secondly, I draw up a 2 x 2 table with the "truth" along the top (No differences, Differences) and the experimental decision down the side (No differences, Differences). Two of the four cells constitute correct decisions, while the other two cells correspond to Type I and II errors (the class find this easy to relate to, since they have had plenty of discussion of such errors during the afternoon). Thirdly, I quickly go over the plan for workshops #2 and 3 (the two remaining workshops in the first series).

3.2. OTHER WORKSHOPS IN BRIEF

The first three workshops (*basic statistics/analysis of variance, days 1 - 3*) form a series in that the workshops build on each other and people cannot join in partway through. The fourth and fifth workshops (*simple regression and analysis of covariance*

(ANCOVA)) form an independent second series in that people can attend them without having attended workshops 1-3 (in practice, though, only a few people exercise this option). The *ANCOVA* workshop builds on the regression workshop, so the latter is a pre-requisite.

The workshops have been run most winters since 1978, except for the more specialised *ANCOVA* workshop, which is run about every second year. There is also a third independent series of two workshops which has been run on three occasions; this is entitled "Statistics done properly" and is based upon the *N*-dimensional geometry which is at the heart of these methods.

Successive workshops are normally run with at least one day's break between each workshop, to allow people's brains to recover. I only break this rule if my travel/accommodation costs make it necessary – on such occasions, I notice that the class members become quite tired by the end of the second day. Tuesdays and Thursdays are the days I aim to use for workshops.

I shall now briefly outline the contents of workshops 2-5. The second workshop (*basic statistics/analysis of variance, day 2*) starts by revising the work of day 1 using a second set of data.

Here the population consists of 64 plots in a pasture arranged in an 8 x 8 grid, with the data being pasture dry matter yields expressed in kg/ha (these are data made up by myself, and are not real data). The pasture is highly variable, with yields ranging from 360 to 930 kg/ha (unlike the weight gains used on day 1, which were relatively low in variability). Three treatments are hypothesised (untreated, "superphosphate fertiliser (super)" and lime), and three plots randomly assigned to each treatment by each class member.

To generate known treatment effects, 200 and 100 kg/ha are added to the yields for the second and third treatments respectively (that is, the "true" super response is 200 and the true lime response is 100 kg/ha). The analysis of variance is carried out, and most class members fail to "detect" any treatment effects using the overall *F* test. This demonstrates the low "power" of such an experiment. The exercise is then carried on to estimate super and lime responses and the significance of these responses. This also leads to the calculation of the "least significant difference (LSD)" between any two treatment means and a discussion of multiple comparison procedures (Saville, 1990). The class results are then used to show how one would combine data from a series of identical trials. Presentation of results for a report or journal article is then covered, as are sample size calculations.

The third workshop (*basic statistics/analysis of variance, day 3*) starts off by introducing the idea of a statistical model, using the day 2 design as an example. This sets us up for dealing with a randomised block design experiment.

As a class exercise involving such a design, the pasture data from day 2 is re-used in conjunction with "scores" of the pasture plots which were assigned by the experimenter before the experiment was set up (these contrived scores range from 3 to 9, and are very highly correlated with the subsequent yields).

Each class member designs their own experiment by choosing blocks of three plots with similar scores within each block, and randomly assigning treatments within each block. They then use the corresponding yields, adding 200 and 100 as in day 2, and fit a randomised block model. The corresponding analysis of variance then uses "actual" and "expected" variances of the mean as in days 1 and 2. Failing arithmetic errors, class members notice a dramatic increase in precision as compared to their day 2 results. As in day 2, responses are estimated, the LSD(5%) is calculated, and there is a discussion of power for both overall and more specific tests. The analysis of variance table is

related to the workshop method, and there is a full discussion of assumption checking. Lastly, the "whys and hows of blocking" are discussed.

The first workshop in the second series (*simple regression*) again uses a data set that I invented for the purpose of the class exercise. I first set up a hypothetical straight line ($y = 80 + 10x$). Then, for each of $x = 2$, 5 and 8, I generated 20 y values from a normal distribution centred on the appropriate point on the line with a standard deviation of 20.

Each class member uses random numbers to select two y values for each of $x = 2$, 5 and 8 from the 20 which I provided, then calculates the intercept and slope of the fitted line using formulae which I provide. The class histogram of slopes gives people an idea as to the standard error of the slope. The latter is then estimated for each person's regression by calculating the variance about the line and using this in the appropriate formula. Hypotheses about the true slope are tested, and the equivalence of F, t and r (correlation) tests are discovered by the class. As time permits, 95% confidence bands for the "true line" and for predictions are calculated. The pitfalls of the method are discussed, as are Anscombe's Quartet, the influence of outliers, the dangers of extrapolation, and why a "high r value does not ensure the curve is a good fit."

The more specialised final workshop (*analysis of covariance*) deals with the fitting of parallel straight lines. The same basic data set is used from the simple regression workshop, except that course members draw their data as two "x-y" samples, and learn how to fit two parallel lines. For the first line, a single y value for each of x=2, 4, 6 and 8 is used, while for the second line, a single y value for each of x=2, 4, 4 and 8 is used. To offset the second line, each class member adds 60 to each y value for the second set of data. In the analysis, hypotheses of interest are whether the common slope is significantly different from zero, and whether one line is at a significantly different elevation to the other. The common uses of this method are also discussed in relation to real examples.

Workshops 1-4 cover the basic material which is of most interest to researchers, so are the most popular. Workshop 5 is useful to some researchers, but not others, so attendance is always lower for this workshop.

Workshops 6-7 (*statistics done properly, days 1 and 2*) explain the mathematics that is at the heart of the methods covered "heuristically" in workshops 1-5. In my opinion, these methods cannot be "properly" explained without the use of geometry.

During workshops 1-5, I mention words such as "degrees of freedom" for which I can give only an intuitive "hand waving" explanation – on these occasions, I comment that the real explanation lies in the geometry. Such comments arouse the curiosity of some of the class members, so that usually about one third to one half of them are keen enough to learn more by attending workshops 6-7. In the event, about half of those who do workshops 6-7 find them really rewarding. The workshops are largely based on simplifications of the four class exercises in Saville and Wood (1996). For reasons of space, I shall not detail these workshops here.

Over the years, I have also run other workshops on more specialised topics. I ran one on "orthogonal contrasts in analysis of variance," but found that most people could not grasp the concept of orthogonality (which is basically the idea of right angles in N-dimensional space, so hard to teach without mentioning N-space). I ran another one on "combining data from a series of trials," but concluded that this specialised topic was not one that warranted a workshop.

On two occasions, I set up workshops on "design of experiments" (meaning designs more advanced than the ones covered in the basic statistics workshops). The first was an informal, interactive workshop for a small group of 7 researchers all working on similar problems at a distant research station – this went well. The second was a workshop for

about 16 researchers from diverse fields of research – I felt this was poorly received, since the designs used in different fields are quite different, and any particular design that I spoke about was only of interest to about one third of the people in the room, meaning each person was bored about two thirds of the time! My conclusion was that more specialised topics are more efficiently taught during one-to-one consultations, and that only the core ideas are best covered by a workshop.

3.3. INVENTING THE WORKSHOPS

In 1978, when I invented the workshops, I was unaware of any of the literature on statistics education. However, I did have the benefit of advice from a school teacher friend. She told me to focus on the stumbling blocks, which are the things which researchers find hard to understand about statistics.

These blocks seemed to be everywhere, but especially in the very basic material, so it seemed to me that the only sensible approach was to cover the material *from the beginning*, working on each stumbling block in turn, and taking time to overcome each stumbling block.

In some cases, I had to think hard about how to overcome a particular obstacle. For example, researchers take a sample of observations and estimate the mean and standard deviation; then they calculate the *standard error of the mean*; this is hard for them to understand - they have only one mean, so how can it vary? This is one reason why I set up an exercise in which the students took several samples, calculated the mean for each sample, and then calculated the variability of these means (after lunch on day 1, as detailed in §3.1).

I had various other ideas in my mind when I designed the workshops. As a pure mathematician who learnt statistical methods "on the job," I found that I needed to go through the calculations for each type of statistical analysis by "long hand." I therefore felt that it was important for each researcher to go through each example long hand. At the same time, I thought it was a good idea if everyone was in effect doing a "simulation" so that we could summarise the class results in histograms and other ways which would illustrate statistical principles and results. Lastly, in the case of analysis of variance, I worked out a method of doing the arithmetic (involving the variance of the mean) which tied in with a gradual build up of ideas via a sequence of simple class exercises.

The workshops have slowly evolved and improved over the years based on feedback from class members. Also, during the 1990s a fellow statistician and ex-secondary school teacher, Mrs Lesley Hunt, highlighted for me the value of the "Discovery" method. This led to me more consciously standing back to let the class do the talking as much as possible ("discovering" statistical truths for themselves), with me just filling in any gaps at the end. She also suggested drawing up a "plan" for each day so that class members could see where they were at each stage of each day. The course evaluation sheets were a further improvement suggested by Lesley.

The "basic statistics" series originally consisted of two workshops. However, the material included in these workshops slowly increased over the years. In addition, the Discovery method, while an improvement, meant that discussions were encouraged more actively, and lasted longer. This meant that it was difficult to cover all of the planned material in two days, and some material was covered in a rush (especially the Day 2 material). This led to some 1998 class members suggesting that the material should be spread over three days. I decided that this was long overdue, so I redesigned the series before the 1999 workshops. The same material was covered in the same

manner, but I spread the old Day 2 over two days and added new material on assumption checking.

3.4. WORKSHOP EVALUATIONS

From 1978 to 1996, the workshops were run without any formal evaluations. People who attended the workshops seemed to be very pleased with what they learnt, and feedback was generally very positive (people did suggest minor improvements, so I felt that negative feedback would have been forthcoming if it had been warranted).

In general, I was always pleased that I had made the effort to run the workshops, and felt motivated to re-run them the following winter (except for an occasional year, such as the year I was overseas and the year we were restructured from a government department to a new government-owned institute). People attending the workshops recommended them to their colleagues, and there was a continuing "supply" of attendees (though fluctuating, in the range 6 - 26 per annum), so I concluded that the workshops were filling a continuing need.

For the last three years (1997 – 1999) the attendees have filled out evaluation sheets (Figure 1; note that the lines of text are evenly spaced over an A4 sheet). In general, most respondents have been very positive about the workshops, with similar responses in all three years. To indicate the sort of responses given, I shall summarise the responses for the most recent year (1999).

In response to question 1, most people expected to gain a better understanding of statistical ideas and methods, some hoped to gain increased confidence and/or increased ability in interpretation, and some had no or low expectations. The responses in terms of whether expectations were met are summarised in Table 1.

The second column of Table 1 is the total number of different people who attended one or more workshop. Only 1 out of the 29 respondents said that the workshops failed to meet their expectations and a further respondent said they only partially met their expectations. On the other hand, 69% said the workshops met expectations and 21% said that their expectations had been exceeded (replies like "Yes, and more," "Very much so," "Definitely," "Yes, very well"). Amusingly, one person had two expectations. The first, "enlightenment and understanding," was met. The second, "swamped with figures/formulas and left confused," was not met (to my relief). Another person wrote that their expectation (from the workshop advertisements) was "not to learn anything new," but that this expectation was not met!

Figure 1. Course Evaluation Sheets Used in 1997, 1998 and 1999.

Statistics Courses 1999
Feedback and evaluation sheet

Course name *Your name (optional)*

1. What were your *expectations* of the course? Were these met?
2. What did you *learn* from the course?
3. What did you *like* about the course?
4. What did you *dislike* about the course?
5. What could be *improved* about the course?

Use reverse of page if you run out of room

Table 1. Responses to the Question "Were These (Expectations) Met?" in 1999.

Location	Number of different people on workshops	Number of different people responding	Number of people responding as follows			
			No and/or partially	No expectations	Yes	Stronger than yes
Lincoln	12	8	0	0	5	3
Palmerston North	36	12	0	0	9	3
Wallaceville	16	9	2	1	6	0
Total	63	29	7%	3%	69%	21%

The subsequent questions had many, varied responses, so I shall just summarise the Palmerston North responses, which were about average. To question 2 (what was learnt), replies were:

"Insights from Dave about analysis options"; "Basic trial design and how to think about design"; "Reinforced the theoretical that was learnt previously with practice"; "The fundamental idea behind statistical equations, which makes them seem less daunting and more understandable";

"(ANCOVA workshop) How to correct treatment means for covariate effects, test for significant treatment effects (very useful), interpret F, t, r and so on (very useful), understand importance and relevance of these values and how false interpretations can occur (pitfalls)";"The hows and especially the whens of regression and ANCOVA".

"Basic analysis of variance, how to understand the different terminology, how LSDs are useful, how to understand ANCOVA (needed for person's own trial which had a nuisance factor)". "LSD, how to calculate sample size, "what the numbers I have been computing for so long actually mean."

"(Geometry workshops) Greater understanding of the geometric methods behind stats". "The applications and ability to calculate regression and ANCOVA myself. In addition I now really understand where F comes from!"

To question 3 (what people liked about the courses), replies were:

"Emphasis that "simple is best." Generous provision of time for exercises, limiting the hassle (anxiety) associated with some courses where basic skills and confidence are a bit low".

"Practical working"; "Hands on. Prepared notes"; "The chance to hand work the analysis – we can become too dependent on computers doing this for us"; "Hands on approach to learning (best way to learn stats in my opinion). Time spent on answering specific questions relating to material (improves relevance to own work)".

"Good number of people"; "It was done at a pace that I could keep up with, using very understandable language"; "Good practical examples. Honesty about limitations of stats". "Worked examples made things easier to follow; "The informality. Smaller class for workshops 4-5"; "Slow pace, reinforcement of examples. Class aggregation of data to demonstrate principles"; "Dave's easy, relaxed method of teaching. It was magic to see and hear someone with a bit of humour about stats!".

To question 4 (dislikes), six people left the section blank or wrote "nothing." Other replies were:

"Attendees who ask too many questions"; "Some of the verbal explanations were a little

inadequate – but due to time constraints – understandable"; (Workshops 1-3) *"Too much of routine calculations following the "rules" versus not much discussion of ideas and concepts behind the rules".* (Workshops 4-7) *"No dislikes".*

(Regression and ANCOVA)*" Sometimes went too quickly over difficult areas";* (Workshops 1-3) *"Time spent waiting for class results".* (Workshops 4-5) *"Got lost at one point in the algebra –still can't figure out why we did it as we didn't use it later."*

(Geometry workshops) *"Last hour a bit rushed – otherwise not a lot (of dislikes)".*

To question 5 (improvements), four people left the section blank or wrote "nothing." Other replies were:

"Request for an informal, spiral-bound course text with sequential exercises worked through with explanations, followed by appendices with Dave's supporting papers, stats tables, etc". "Limit on number of people, say 20. At end of course: single handout sheet of formulas plus explanations and symbols plus meaning, to be used as a prompt for returning to the notes". "Addition of information on chi-squared tests and partial correlation coefficients".

(Workshops 1-3) *"To have more discussion on concepts and ideas".* (W 4-7) *"Nothing". "Go quickly over initial stuff and more slowly over difficult stuff". "Less people, smaller class, go faster".*

(Geometry) *"Perhaps a fully worked example of one of the more complex procedures (e.g., ANCOVA or RCBD). Additional session with computing".*(Regression and ANCOVA) *"Speed about right. More depth on 3 covariates – but appreciate that this was beyond the remit of the course".*

Further evidence of attendee satisfaction is provided by the low attrition rate. Researchers are encouraged to attend the first workshop in the "Basic Statistics" series with the assurance that there is no obligation to attend the other workshops if they do not find the first workshop useful to them (in line with this, non-AgResearch staff only pay for the workshops that they attend, with invoicing being done after the workshops rather than beforehand). In Table 2 we present the numbers of people starting and finishing the "Basic Statistics" series of workshops (2-day series, 1995-1998; 3-day series in 1999), the 1-day Regression workshop and the 1-day Analysis of Covariance (ANCOVA) workshop (which was not always offered).

Table 2: Numbers of People Starting and Finishing the Different Workshops

		Basic series numbers		Numbers attending	
Location	Year	First day	Last day	Regression	ANCOVA
Lincoln	1995	22	21	18	-
	1996	26	25	18	11
	1997	18	16	16	9
	1998	22	19	17	-
	1999	10	9	9	8
Palmerston North	1998	24	24	19	-
	1999	26	24	18	15
Wallaceville	1999	16	11	-	-

As seen from Table 2, however, most researchers complete the entire "Basic" series,

and many return for the Regression workshop (though the numbers include a few people for whom this is the first workshop). Lower numbers return for the ANCOVA workshop, but this is largely because not all researchers perceive this method to be important. Note that sickness is one component of the attrition rate. Note that the high attrition rate at Wallaceville was largely because 7 of the attendees (all from one division) were called away to a hastily scheduled restructuring meeting halfway through the afternoon of Day 2, with only 3 of these people returning on Day 3.

3.5. ESSENTIAL INGREDIENTS

There are various aspects of the workshops which I feel contribute to their ongoing popularity. It is hard to know which aspects are the most important – however, I shall list the ones I feel are most crucial. Two of the most vital ingredients are:

1. Start at the *beginning*;
2. *Go slow.*

These two ingredients require nerves of steel in the presenter! No one likes to be considered stupid or uneducated, so the natural inclination is to exhibit one's intelligence and/or vast knowledge, preferably as soon into the course as possible! The worry is that the attendees will decide that they cannot possibly learn anything from a workshop which starts off dealing with such elementary material, and walk out "en masse," to the presenter's never-to-be-forgotten embarrassment. I have had an occasional individual who has made this very decision and departed one hour into the first workshop, but luckily, the contagion has never spread!

Yet, if you lose your nerve and start anywhere except at the beginning, or go fast, you will certainly lose your class (especially if they are volunteers, as in my classes). In reality, most researchers have had bad experiences with their previous statistics courses, so approach the idea of another statistics course with fear and trepidation. They are therefore, in the main, relieved and eternally grateful that the workshop material is understandable, starts at the beginning, and goes only at their pace.

Two points that are also essential are:

3. Provide *hands-on* work with agricultural data sets;
4. Encourage *participation* and *interaction*.

The hands-on work with data is really appreciated by attendees, and is favourably commented upon in many of the evaluation sheets. I feel it is important because the class members get "the feel" of how to do statistical calculations, and this gives them a confidence in their own statistical ability. Participating in a group learning experience is also something most people enjoy and benefit from. The group as a whole quickly start to interact among themselves and with myself, ideas are bounced around, jokes made, and a nice friendly atmosphere develops. The class members develop a sense of ownership of their own data set, and mentally identify with the hapless researcher who has obtained such variable data.

Once "the scene is set" (i.e., once each class exercise is set up), class members *discover* many things for themselves, and this sinks in much better than if I told them the answer. The discussions of each set of class results can be quite wide-ranging as people raise all sorts of statistical issues which have been worrying them – some of these issues can be discussed naturally at the time they are raised, and others I need to

postpone till later in the workshops and/or discuss outside the workshop time. I feel that the attendees perhaps learn as much from these impromptu discussions between a practising biometrician/statistician and a group of agricultural researchers as from the more formal parts of the workshops. What they subconsciously assimilate is "how a biometrician thinks," which tells them a great deal about statistics.

Yet two more elements that are crucial are:

5. *Experiencing the variability;*
6. Learning to *live with uncertainty.*

Variability is at the heart of statistics, and *experiencing the variability* is an important thing that the researchers get out of the workshops. They are always highly impressed by the fact that their result is *not* statistically significant while that of their neighbour *is* significant. When there are truly no differences between populations, they are "right," while when there are differences, they are "wrong." This leads to good-natured joking between the class members, and leads to a good atmosphere in the workshop, while simultaneously explaining the idea of "Type I and II errors" and the concept of "power."

An issue which worry researchers and which is always raised is "How do I know whether I've made the right decision?" My answer is that it is like being a jury in a court of law. You never really know whether you've hung an innocent person (Type I error), let a guilty person go free (Type II error), or made the correct decision – all you can do is design your experiment so that you minimise the chances of both types of error, and/or carry out a second experiment to confirm your apparent findings.

Throughout the workshops, this "uncertainty" in the results is continually discussed in relation to the various scenarios which we simulate. In some scenarios, we know there are truly no differences, yet some class members decide there are differences; in other scenarios, there are differences whose magnitude we know, yet some class members obtain estimates which are quite different from the true value. Since researchers in practice never know the truth in terms of the responses that they are estimating, they find it a useful exercise to see how good or bad the statistical methods are when they do know the truth.

Two further vital ingredients are:

7. *Confidence* building*;*
8. Building *interest in statistics*.

Many attendees come to the workshops with a dislike for statistics and a lack of confidence in their own ability in the subject. I like to think that they go away with a greatly increased level of confidence and a much higher level of interest in statistics. During the workshops people learn only the very basic facts of life "statistically speaking." They also learn during the discussion periods that there are many things which are outside the scope of the workshops – however, their basic grounding means that they are in a much better position to consult with a statistician concerning their more difficult designs or analyses. An improved appreciation of statistical ideas means that people are also more likely to seek statistical advice when they need it.

3.6. GENERAL COMMENTS

With the widespread availability of modern computers, one might wonder whether

there is still a place for workshops based upon simulations done by "calculator arithmetic." Would it not be better to set up a room full of PCs so that each student can click the "simulate" button and save themselves a great deal of "mindless" arithmetic? My feeling is "no," though I would be interested in anyone's attempts to prove me wrong! My reasons are two-fold. Firstly, I feel that the mind is working whilst the arithmetic is being done, and that things are being learnt. The "no pain, no gain" phrase comes to mind – clicking a button is very easy, and perhaps as easily forgotten. Secondly, the PCs could be a distraction from the main event, which is thinking about statistics.

Another question, which is periodically raised, is: "Do the workshops need to be updated?" This is an alias for the thought: "You've run these courses for 22 years now – surely they must be out of date?" My answer is three-fold. Firstly, the methods covered by the workshops are still the ones most commonly used in agricultural research, and there is still a high demand for courses which explain the basic ideas behind these methods. Secondly, if the workshops work, why change them? After all, humans have been eating and drinking for centuries – some things take a while to go out of fashion! Thirdly, there is the automation issue, which I discussed in the first paragraph.

The value of the geometry-based workshops 6-7 is also periodically questioned, with the idea that researchers have a pragmatic view of statistics, and would not be interested in the mathematical ideas which underlie the methods (c.f., "they just need to be able to drive the tractor, not understand how it works"). I find that this is true of some researchers, just as it is true of some statisticians. However, equally, some researchers are fascinated by the elegance of the geometry and the way in which it unifies a large number of apparently unrelated, seemingly "ad hoc" statistical methods (this is also the fascination for myself). For example, one such researcher attended all of my workshops and told his colleagues that the geometry workshops were the most interesting of all. In addition, several people who have attended the geometry workshops have been interested enough to buy one or other of our books (Saville and Wood, 1991 and 1996).

In summary, workshops 1-5 for agricultural researchers are suitable as a "heuristic" way of introducing the statistical ideas which underpin the most commonly used "frequentist" methods. Workshops 6-7 and/or our two textbooks constitute an alternative, more unified approach in which theory, practice and computing are all viewed from the same perspective.

4. THE GEOMETRIC APPROACH

As mentioned above, Graham Wood and I have pioneered a new method of teaching statistics geometrically to university undergraduate statistics students and agricultural science postgraduate students. This has been the subject of two books (Saville & Wood, 1991, 1996) and a review paper (Saville & Wood, 1986). The first of these books served as the textbook for a second year applied statistics course at the University of Canterbury, Christchurch, New Zealand until the end of 1998.

This course was jointly taught by Graham and myself for the first year (1984), then by Graham on his own for several years, then by Dr Murray Smith for several years, and finally by Dr Peter Heffernan for several years. All of these lecturers were enthusiastic about the geometric approach, and the course was well liked by the students throughout the 15 years during which it was taught (sadly, the course was discontinued when the statistics courses were re-arranged in 1999). In general, the students found the material relatively easy to understand, and they found it satisfying because they were learning

something relatively concrete (how to design and analyse experiments) following a method which unified theory, practice and computing.

As evidence of its popularity, I shall present the course and teacher evaluations for the final year that the course was run (Table 3). The first survey rated the course, while the second survey rated the teaching. The number enrolled in the course was 59, and the numbers responding to the two surveys were 33 and 35 respectively. The 1 – 5 rating scale is defined at the top of the table.

Table 3. Percentages of Responses To Two Surveys Carried out by the University of Canterbury Education Research and Advisory Unit at the End of the 1998 Course.

Student Rating of Course (1-5 points scale; 1 strongly disagree; 5 strongly agree)				
Standard Questions	Mean	<3	=3	>3
This was a well-organised course	4.3	0	3	97
This course helped to further stimulate my interest in the course area	4.0	6	15	79
The overall workload in this course was reasonable 1 = too light / 3 = reasonable / 5 = too heavy	3.2	9	64	27
The level of difficulty of this course was reasonable 1 = too easy / 3 = reasonable / 5 = too hard	3.2	9	70	21
Overall, this was a good quality course	4.2	0	24	76
Supplementary Questions				
The lectures were a valuable aid to my learning	4.1	3	16	81
The tutorials were a valuable aid to my learning	4.2	3	9	88
Student Rating of Teaching				
Standard Questions				
The classes were well organised	4.3	0	11	89
The lecturer was able to communicate ideas and information clearly	4.3	3	9	89
The lecturer stimulated my interest in the subject	4.1	6	20	74
The lecturer's attitude towards students has been good	4.8	0	0	100
Overall, the lecturer is an effective teacher	4.5	0	14	86
Supplementary Questions				
I found that the teaching methods used in this course were effective in helping me to learn	4.3	6	11	83

The ratings were excellent. For the questions for which 5 was the most favourable rating, the mean ratings were all between 4 and 5. For the two questions for which the most favourable rating was a 3 (workload and level of difficulty), the mean ratings were both close to this on 3.2. Of course, in such ratings the ability of the teacher is paramount in that a wonderful method of teaching will prove a dismal failure if the teacher is not equally wonderful! Nevertheless, some credit at least should go to the geometric method of teaching this material. The success of this course serves to contradict the opinion that students would not be able to cope with a geometric approach (such an opinion was expressed in several journal reviews of our 1991 book).

The geometric method of teaching statistics is especially enlightening for, though not restricted to, university students who have taken first-year courses in linear algebra (also called vector geometry) and statistics. The linear model statistical methods are seen as a simple application of linear algebra, and the linear algebra is made less abstract by being applied to a concrete problem. However, the linear algebra requirements of the method are quite minor, and could be easily learnt by Californian graduate students in agriculture during a one hour lecture.

Our second book (a primer) on the geometric approach also includes material on the way in which R. A. Fisher came to invent the linear model methods. This is material on the *p* value which is included in an appendix (D).

ACKNOWLEDGEMENTS

I would like to thank the attendees and other colleagues who have contributed to the development of my workshops over the years and AgResearch for sponsoring my attendance at the IASE Round Table conference in Tokyo, Japan, in August, 2000.

REFERENCES

Bangdiwala, S. I. (2001). Training of statisticians worldwide to collaborate as co-investigators within country clinical epidemiology units: The experience of the International Clinical Epidemiology Network (INCLEN). In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 265-275). Granada: International Association for Statistical Education and International Statistical Institute.

Bishop, O. N. (1980). *Statistics for Biology: A practical guide for the experimental biologist, 3rd edition.* Harlow, Essex, U.K: Longman.

Saville, D. J. (1980). Analysis of variance in agriculture (revised). *Ministry of Agriculture and Fisheries Biometrics Bulletin* 1, 1-49.

Saville, D. J. (1990). Multiple comparison procedures: The practical solution. *The American Statistician* 44, 174-180.

Saville, D. J., & Wood, G. R. (1986). A method for teaching statistics using n-dimensional geometry. *The American Statistician* 40, 205-214.

Saville, D. J., & Wood, G. R. (1991). *Statistical methods: The geometric approach*. New York: Springer.

Saville, D. J., & Wood, G. R. (1996). *Statistical methods: A geometric primer*. New York: Springer.

David J. Saville
AgResearch, Statistics Group
Canterbury Agriculture and Science Centre,
P.O. Box 60, Lincoln, New Zealand
E-mail: Dave.Saville@AgResearch.CO.NZ

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 215-226.

Printed in Granada, Spain.

GLENYS BISHOP & MIKE TALBOT

STATISTICAL THINKING FOR NOVICE RESEARCHERS IN THE BIOLOGICAL SCIENCES

Postgraduate students from non-statistical disciplines often have trouble designing their first experiment, survey or observational study, particularly if their supervisor does not have a statistical background. Such students often present their results to a statistical consultant hoping that a suitable analysis will rescue a poorly designed study. Unfortunately, it is often too late by that stage. A statistical consultant is best able to help a student who has some grasp of statistics. It is appropriate to use the Web to deliver training when required and that is the mechanism used in this project to encourage postgraduate students to develop statistical thinking in their research. Statistical Thinking is taught in terms of the PPDSA cycle and students are encouraged to use other Web resources and books to expand their knowledge of statistical concepts and techniques.

1. INTRODUCTION

Much current statistical education of researchers is focused on training in specific techniques such as how to use statistical packages and interpret results. Not enough attention is devoted to such strategic issues as what technique is most appropriate and how to organise data collection. This paper describes a project that aims to fill this gap and to provide training in strategic research skills from a statistical viewpoint.

For instance, Day (1998) has tackled the problem of guiding young researchers through the research process. We wish to focus here on the statistical thinking required in the research process. This is not something that can be tacked on at the end but is intertwined with the contextual thinking throughout the whole project.

We have adopted the notion that statistical thinking takes account of variation and uncertainty, that it uses models for reasoning and that it brings together statistical understanding and contextual knowledge.

The inspiration for this project came from a paper by Pfannkuch and Wild (1998) and later, Wild and Pfannkuch (1999) who, following earlier work by MacKay and Oldford (1994), propose a cyclic model to describe the major elements of a statistical investigation, based on Problem-Plan-Data-Analysis-Conclusions (PPDAC). The model is cyclic because, at each stage, one must look forward as well as backward and the conclusions relating to one problem lead on to a new problem.

They propose zooming in at each stage of the cycle to gain more details of the stage at deeper and deeper levels. It is their view that statistical thinking occurs at the interface between statistical activity and context matter knowledge. They suggest that if the cyclic model targets a particular subject area, more expertise for synthesising the context knowledge and statistical knowledge can be incorporated into the deeper levels.

With a slight change of terminology, we have used this approach in this project. We prefer to name the cyclic elements, Problem-Plan-Do-Study-Act. The act of zooming in

to gain more and more insight into a particular stage is particularly suitable for use on the World Wide Web. To that end we have sought in-depth information about each stage of the cycle and concentrated on the biological sciences for context matter as the application area.

Other authors have used the cyclic approach. Box, Hunter and Hunter (1978) give an excellent example of an investigation into water filtration that highlights using the conclusions of one experiment to form the next problem. We have taken the liberty of including several cycles from their example. Connolly (1996) has devised a research methods course for young researchers in the biological sciences. He considers the four phases of biological research to be Planning, Execution, Analysis and Reporting with various aspects of statistical thinking required in all stages. He illustrates the connections between the phases by using examples of defective experimental designs and emphasising the reasons to apply statistical methods and when.

Many authors have analysed and described what statisticians do and these can assist with deep descriptions of each stage of the cycle. Chatfield (1995) devotes much of his attention to the analysis (study) stage, and stress the importance of analysing data from properly planned studies. He distinguishes five main stages of a typical analysis, namely, look at the data, formulate a sensible model, fit the model to the data, check the fit of the model and utilise the model to present conclusions. Coleman and Montgomery (1993) provide a detailed method for planning an industrial experiment. Some of their details, such as the discussion of response variables and the tutorial on interactions, translate readily to the biological sciences.

There are a variety of papers discussing the concepts of statistics and statistical thinking. Bartholomew (1995) defines Statistics as being "concerned with the real world through the information that we derive from classification and measurement. Its distinctive characteristic is that it deals with variability and uncertainty which is everywhere." Pfannkuch and Wild (1999) summarise statistical thinking as taking account of variation, constructing and reasoning from models, transnumeration, and the synthesis of problem context and statistical understanding.

Gal (1998) distinguishes two contexts for interpretative skills, namely reporting and listening or reading. The novice researcher must be able to interpret, create, communicate and defend opinions based on statistical argument. Both contexts are relevant for researchers but data producers, at whom this project is aimed, are particularly concerned with the reporting context. Therefore, they must be able to defend the use of an experimental or survey design or the implications of an analysis of experimental data. On a similar theme, Moore (1998) argues that statistical thinking should be considered as a liberal art and that we should therefore spend more time on broad ideas rather than technical content.

There are many Web-based materials available for teaching introductory statistics course. Puranen (1998) and Ganesh & Ganesalingam (1998) discuss some of these. Many of these materials are aimed at first-year undergraduates. However, there are some designed for more advanced users. Talbot et al (1998) have developed an infrastructure, SMART, for publishing training materials in advanced statistical methodologies. SMART is aimed at established researchers in the biological sciences and, thus, it assumes that the researcher understands the research process and the place of statistics in the same. It became clear during the development of SMART that novice researchers would need more than an introduction to some of the modern methods of statistical analysis. They would need to be introduced to the statistical design process. To some extent, this has been addressed by including a checklist, and a link has been provided in the Statistical Thinking facility to this checklist.

Bishop (1998) tackled the issue from another angle by developing Web-based material for illustrating the basic concepts and terminology of experimental design. Two real experiments are used as illustrations with photographs and video clips while the student is able to experiment interactively with techniques such as randomisation and blocking. These experiments are a useful introduction to design and are based on lectures given to Statistics majors but do not include the whole statistical process. This introduction to experimental design has been linked to the statistical thinking facility.

The self access project at the University of Adelaide is developing schemes for enhancing undergraduate and postgraduate students' development as independent lifelong learners by means of an integrated series of Web-available learning resources in a number of different learning contexts. This will demonstrate to students a model of independent, resource-based learning on demand, involving self-access to a linked series of Lotus Notes databases in six main content areas with a common user interface and front page arrangement. The Statistical Thinking facility forms part of this project.

2. AVAILABLE STATISTICAL HELP

At the time that this project was instigated, there were several sources of statistical help provided by the Department of Statistics to University of Adelaide students. Undergraduate Science students were encouraged to take Statistical Practice I, a first year subject, equivalent to a one-quarter load for one semester. This subject used the textbook by Moore and McCabe (1999, and earlier versions). A second year subject, Statistical Practice II, following on from the first year subject was seldom taken by Science students, mainly because of timetable and funding problems. Some departments, such as Zoology and Geology, offered applied statistics subjects to their own students. It was possible for a postgraduate student who lacked statistical knowledge to audit undergraduate subjects. At the time of writing, all of these subjects are still offered.

However, other avenues of help are no longer available. The Department of Statistics offered a fee-paying course to the general public. This was a 72-hour course spread over 12 weeks and postgraduate students were encouraged to attend with the fee subsidised by the University. This was an introduction to statistical practice and not a course in research design. Postgraduate students were also entitled to five free hours of a statistical consultant's time. The five hours were usually divided between discussing the project with the consultant and the consultant performing some analyses; formal reports were not written. More than 100 students per year used this service.

One of the authors (Bishop) participated in this consulting service and observed that many postgraduate students who came to consult lacked sufficient statistical knowledge to have a meaningful discussion about the role of Statistics in their projects. Students were requested to write 200-500 words summarising their project before they came. Those that did usually gained a lot from the consultation. Bishop perceived that many of those who did not write the summary needed basic statistical explanations before the discussion could begin. It was departmental policy that the statistical consulting service should not become a dumping ground for data collected by students who had no idea how to analyse them.

Another area of concern was that many students collected data without any consideration of the statistical validity of the experimental or survey design. It would be a good idea to build on a student's basic statistical knowledge by demonstrating the many facets of statistical thinking. A book was not required, rather a succinct summary

readily accessible when needed. Therefore, the Statistical Thinking facility was developed for the University of Adelaide Intranet.

3. DESCRIPTION OF THE STATISTICAL THINKING FACILITY

3.1. GENERAL OVERVIEW

The first screen of Statistical Thinking gives the user information about the purpose of the facility, the focus on biological sciences and how the first-time user should begin. There are ten headings or main topics:

- Introduction;
- Problem;
- Plan your experiment;
- Plan your survey;
- Plan your observational study;
- Do;
- Study;
- Act;
- Help with extra reading;
- Useful web sites.

One can see that the five steps of a statistical investigation are covered, with Plan having three options, depending on the type of data collection planned, i.e. experiment, survey or observational study. The first-time user is advised to begin with the Introduction, which explains the cycle with examples and gives instruction on using the facility. Other topics provide help with extra resources, either web-based or printed materials. The user can choose to expand all or some of the topic headings. This gives the next level of information and by choosing one of the subtopics, the user will open a screen of information about that subtopic. For instance, the topic, *Plan your experiment*, has the following subtopics:

- State the objectives for this experiment;
- Turn an objective into a hypothesis;
- Choose the response variables carefully;
- Select the treatments;
- Select the experimental units;
- Choose blocking variables;
- Plan collection of new data;
- Time plan;
- Plan analysis;
- Piloting and adjustment;
- Checklist.

The topics, *Plan your Survey* and *Plan your Observational Study*, have not yet been developed but they appear in Statistical Thinking as an indication of their usefulness.

3.2. DETAILS OF PLAN YOUR EXPERIMENT

We have identified these components as the necessary steps in planning an experiment. The most difficult step to write in a general sense was *Turn an objective into a hypothesis*. At the next level of detail it is as follows:

- *Remember that your experiment is conducted on a sample of experimental units so that you can infer the properties of the population* (usually called parameters) from which the experimental units were drawn. Any hypothesis relates to the population and its parameters.
- *Break your objective down into components*.
  - *You may want to know whether the mean response differs among the levels of a factor*. You can extend this to each of the other factors. For instance, you may want to know whether the mean wheat yield differs between fertiliser types, A and B.
  - *You may be interested in how two factors interact*. For example, are the responses to rates of application different between A and B? Alternatively, do the differences among wheat variety means depend on the application rate of fertiliser?
  - *Continue adding a factor until all factors are included*. For example, you may want to know whether the wheat varieties respond differently to fertiliser type and application rate.
- *Form a null hypothesis*, for example that the differences of interest do not exist in the population. The analysis of the data you collect will test whether the observed differences in means are strong enough evidence to reject the null hypothesis.

The user selecting *Choose the response variables carefully*, will open a screen with the following information:

Choose the Response Variables Carefully

A good response variable*:*

- *Is easy to record.* Imagine weighing a live pig.
- *Encapsulates the maximum possible information on the outcome of the investigation.* Continuous variables usually provide more information than binary or ordinal variables. Care should be taken that it is possible to analyse the data using the chosen form of the response variable. For instance, it is difficult to count large numbers of lesions on a leaf but is also difficult to obtain a meaningful result if the counting system in Table 1 is used.

Table 1. Example of Counting System

Number	Variable value
0	0
1 or 2	1
3-10	2
>10	3

- *Can be measured objectively on a generally accepted scale.* If a characteristic of interest is directly measurable, then that characteristic can be a response variable, e.g. plot yield or weight gain of an animal are directly measurable. Some characteristics are not directly measurable, e.g. size of an insect, intelligence,

prompt service. Therefore, variables that represent that characteristic must be measured instead.

- *Is measured in appropriate units.* Units may be absolute, such as kilograms, degrees Centigrade or days, or they may be relative such as percentage mortality.
- *Takes values that discriminate well.* For instance, it is difficult to distinguish between mortality rates of 99.5% and 99.8%. It is hard to detect and distinguish contamination levels near zero.
- *Preferably has constant variance over the range of experimentation.* It will be much easier to analyse such a variable using standard statistical methods.

The above subtopic was collated from several sources, particularly Chatfield (1995) and Wild and Pfannkuch (1999).

3.3. SCIENTIFIC METHOD AND ROLE OF STATISTICS

The Introduction contains the subtopics:

- The Scientific Method;
- The Role of Statistics;
- The Five Stages of an Investigation.

The Scientific Method

Although there are potentially many 'scientific methods' in modern science, there is only one generally accepted approach to the solving of problems or the answering of questions. It is based around:

a. *Observation*: classification and measurement
b. *Laws*: accepted generalisations
c. *Theories:* working hypotheses
d. *Experimentation:* testing hypotheses
e. *Inference*: drawing conclusions and making predictions

Statistics has an important role to play in a, d and e.

The Role of Statistics

Statistics is the science of collecting, organising and interpreting numerical facts called data. To think statistically means that one can:

1. Read data, critically and with comprehension;
2. Produce data that provide clear answers to important questions;
3. Draw trustworthy conclusions based on data.

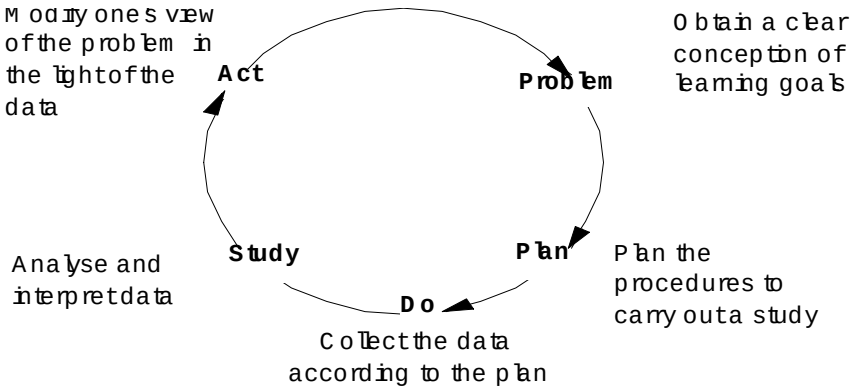
Statistics play a role in the:

1. Planning of surveys and experiments;
2. Collection and presentation of the data;
3. Analysis and interpretation of the data.

The Five Stages of an Investigation contains a diagram of the general PPDSA cycle, shown in Figure 1, and an example showing how the cycle is used in a specific experiment. Navigation through Statistical Thinking can be sequential where the user starts at some point, usually the beginning, and after reading a screen, clicks on the forward arrow. In addition, a back arrow follows the sequence in reverse. A user wishing to have more control can click on the up-arrow to return to the main topic list

and then choose another topic or sub-topic.

Figure 1. The Five Stages of An Investigation



3.4. USEFUL WEBSITES

In some subtopics, useful web site addresses have been given and the user can link directly to them from within Statistical Thinking. For instance, the subtopic, *Adopt an Experimental Design*, discusses determination of number of replicates and contains a link to power analysis freeware available on the Web. A complete list of these web sites is also given as a separate topic. They are shown below.

1. For help with writing you thesis, *Writing and Presenting Your Thesis or Dissertation (http://www.canr.msu.edu/aee/dissthes/guide.htm),* provides many helpful hints on thinking about an appropriate topic, preparing the proposal, writing the thesis and defending the thesis (if your university requires a defence). There is a list of associated bookmarks, some of which will be useful in countries other then the USA.
2. An *Introduction to Experimental Design* is accessible via http://www.maths.adelaide.edu.au/people/gbishop/smart/mintroed/edframe.htm. It contains a brief explanation of the difference between observational studies and experiments, terminology used in experimental design, examples of real experiments and descriptions of how to carry out some experimental design procedures.
3. For an introduction to surveys, there are several documents available from the American Statistical Association. Start with *What is a Survey? (http://www.stat.ncsu.edu/info/srms/survwhat.html),* and follow links to documents about planning a survey, collecting survey data and privacy in surveys.
4. A comprehensive list of software for performing *power calculations* is available at (*http://sustain.forestry.ubc.ca/cacb/power)*. Some is freeware and can be downloaded, some can be used interactively on the web. There is general-purpose software that can be used for power analysis. Some software is dedicated to performing power analyses and some calculate sample size or probabilities but not power. Quite a few have been reviewed and links are provided to the reviews.
5. One site described at the power calculation site, performs *on-line randomisation* and is at: *http://www.stat.ucla.edu/calculators/powercalc*.
6. A *checklist* for making sure that you have covered everything in your experimental plan. (*http://www.bioss.sari.ac.uk/tele/strat/stratdesexp.htm)*

7. The US Environmental Protection Agency is a rich source of documents for carrying out scientific investigations of environmental issues such as air and water pollution. These documents show the rigour required in setting up the study and the role of Statistics in the overall plan. Start with *Guidance for the Data Quality Objectives Process (http://es.epa.gov/ncerqa/qa/qad-docs/epaqag4.pdf).* This process has been updated and applied to a specific type of investigation in *Data Quality Objectives Process for Hazardous Waste Site Investigations (http://es.epa.gov/ncerqa/qa/qad-docs/g4hw-final.pdf*). In particular page 72 of this document shows the iterative nature of the process in a similar manner to the PPDSA cycle.

4. EVALUATION

The first draft of the Statistical Thinking facility was available on the University of Adelaide Intranet at the end of August 1999. However, it was not publicised widely until February 2000. At that stage, postgraduate students who were prepared to use the facility and answer a questionnaire were offered a free statistical consultation. This was thought to be sufficient encouragement as the University's free statistical consulting service had been discontinued to all but Agricultural Science students several months beforehand. Free consultations were offered for a three-week period in February 2000. Twelve students applied for the consultation but some were beyond the deadline imposed by external constraints and so only six received the free consultation. The analysis of these students' responses was qualitative.

The six people who attended for a consultation all offered further comments to their original questionnaire responses. One student's responses were very negative, partly because he had had a lot of trouble accessing the Intranet. Most of his criticisms were directed at the Intranet screen layout, something over which I (Bishop) have no control. I was able to explain the way it worked during the consultation. However, he had not really worked through any of the material in Statistical Thinking, possibly because of his frustration. He wanted to know how many animals he should look at. Unfortunately, there were no published data to give any idea of variation among the animals of interest for the trait under study.

Two other students had not worked through the material either. One was at a very early stage of his thesis and needed help determining what variables to measure. The other could state his overall problem quite well but was unable to progress to describing the population about which inferences were to be made or to defining treatments. Both were engineering students and may have been put off by terms such as block and plot. One requested a downloadable pdf file of all the material and this was later included in the facility.

Two consultees had already completed their theses and both commented that they would have liked the facility at an earlier stage because their supervisors were unable to offer much help with Statistics. Both had problems with more advanced analysis, one requiring a mixed effects model and the other an examination of multicollinearity.

The sixth student was a Mathematics student with a background in biochemistry. He had been studying as much Statistics as he could and so found this facility more of a refresher than an introduction but did state that it had improved his ability to think statistically. He wanted more information on curve fitting.

Generally, the results from this qualitative study were disappointing. There were several contributing factors. The students who responded were desperate to receive the free consultation. They perceived their problems as specific to their own projects rather

than ones that could be addressed by more general information. Statistical Thinking does not address the concerns of the last three consultees described above and one of them specifically requested that the facility be expanded to include various aspects of curve-fitting. This will be discussed later. The author (Bishop) offering the free consultation was about to leave the University of Adelaide and because time was short, found it difficult to insist on more thorough use of Statistical Thinking before the free consultation would be offered. Students were requested to bring a completed checklist when they came for a consultation but none complied. The third problem was the short time frame over which the evaluation took place.

These comments should be viewed in the context of a generally warm response to the facility from postgraduate co-ordinators within the University, a geology lecturer who taught geostatistics, the postgraduate students association and the university's educational advisory centre, who all helped to publicise the facility.

Usage of the facility was logged from August 29 1999 to April 17 2000. During this period, 69 users accessed the Statistical Thinking facility. The amount of time spent on the facility by each user is shown in Table 2.

Table 2. Number of Users Accessing Statistical Thinking for Each of the Time Periods.

Minutes of access	Number of users
0	6
1-5	27
6-10	16
11-15	4
16-20	4
21-25	3
26-30	1
31-35	3
36-40	2
41-45	1
46-50	1
51-55	0
56-60	1

Those who spent zero minutes merely hit the title page and left. Some people spent much longer on each screen than others. For instance, one user made 106 screen changes in 4 minutes 13 seconds and another made 169 screen changes in 21 minutes 53 seconds. At the other end of the spectrum, one user made only eight screen changes in 8 minutes 23 seconds. While there is a large number of short-time users, some of these are repeat users. The mean time spent accessing the facility was 10.08 minutes and the standard deviation was 12.87 minutes.

Most users started with the subtopics of the *Introduction* topic and worked through the facility in sequential order. Some previous users started where they had left off previously. Jumps out of sequence were usually to the *Useful Web sites* topic. Most users worked through part of, or the entire *Plan your Experiment* topic.

5. IMPLICATIONS FOR AND RELATIONSHIP TO THE TRAINING OF RESEARCHERS

A number of issues have arisen during the course of this project. Some relate to the

specifics of the Statistical Thinking facility while others are more general.

Postgraduate students from non-statistical disciplines often have trouble designing their first experiment, survey or observational study, particularly if their supervisor does not have a statistical background either. Such students often present their results to a statistical consultant hoping that a suitable analysis will rescue a poorly designed study. Unfortunately, it is often too late by that stage. One aim of this project is that postgraduate students who use the facility when embarking on a research degree will be made aware of the need for statistical principles in the planning of their projects. Others have used different approaches to the same aim. For instance, Svenson (2001) emphasises the need for communication between the statistician and researcher, while Glencross and Mji (2001) use an integrated approach to research design and statistical analysis in their training courses.

The large number of hits made to the site, which is only accessible by staff and students within the University of Adelaide, indicates that postgraduate students and a few staff are keen to have statistical help. This is supported by co-operation received from various bodies within the University and by the audience response to talks about statistical help given at postgraduate student orientation days over the past few years. It should be noted that use of this facility has not been restricted to biological science students but is open to postgraduate students from all disciplines.

If one can generalise from experience with six students, students prefer to have advice directed at their particular problems rather than to read material that is more general. However, 69 people have accessed Statistical Thinking and the six who received a free consultation may not be typical of the others.

Harraway et al. (2001) have found that a variety of statistical techniques are used in just five branches of biological sciences. Rather than include a large number of analytical methods in this facility, it would be more fruitful to include links to descriptions of analytical methods that are available on the Web. There is a need for brief descriptions of methods that are not usually taught in basic Statistics courses. Talbot et al (1998) have addressed this need in part but more methods with examples are required.

A substantial proportion of users accessed the topics *Design your Survey* and *Design your Observational Study*, even though they were listed as under development, indicating that there is a need for these topics to be developed.

One indication that students are thinking statistically about their own projects is their ability to complete the checklist. the one linked to Statistical Thinking is long and involved and covers practically every aspect of designed experiments, surveys and observational studies. The reluctance shown by the trial group to complete this could be because it is just too difficult and perhaps a simpler checklist would be better initially. The longer one could be tackled when the student feels happier about some of the basic points.

Some students, once they have worked through the checklist, will be able to proceed directly with the experiment. Other students will be in a better position to reap the benefits of a consultation with a statistician because they will have a clearer plan for their experiment and a better idea of the questions to ask the consultant. One of the case studies in Harraway et al. (2001) indicates a student preference to be taught about study design to help formulate the questions to ask a statistician.

The other major issue that arises from this project is that of using Internet or Intranet facilities. Stangl (2001) lists four interrelated reasons that increase the chance of success of the Internet as a teaching tool over other technologies such as video and CD-ROM. One of these is interactivity and she points out its drawbacks of complexity and expense

in developing and delivering the materials and of human factors such as time and the feeling of information overload. Both of these drawbacks were highlighted in this project.

One way that people accessing the Web overcome the problem of information overload is to download printable material and think about it later. Statistical Thinking is essentially text-based and so a pdf file with topics and subtopics in a table of contents has been provided. However, in using the text in this way, students lose some of the structure that aims to give overviews at different levels of complexity.

Some students who followed the link to the Introduction to Experimental Design link encountered problems with the interactivity provided by Java applets, video clip examples and photographic examples. Many students' computing resources do not support the first two of these and are very slow at downloading photographs. As Stangl (*ibid*) says, student access is the biggest hurdle. Furthermore, the inclusion of interactive applets, video clips and examples that rely on colour make printing for later reference difficult, if not impossible.

Harraway et al. (2001) refer to three basic approaches that postgraduate students may use to learn Statistics. They are attending formal lecture courses, attending specialist short courses and informally through their own reading. I would add a fourth approach to this list. The fourth approach would use the basic framework provided by Statistical Thinking, downloadable as a pdf file, much shorter than a book, easily updated and with structure that enable the student to follow a topic to finer detail when required. The facility would be rich in links to other web sites that illustrate poor design and its consequences, sites with examples of good design, sites that describe measurement methods, and sites that outline analytical methods with examples of appropriate use.

REFERENCES

Bartholomew, D. J. (1995). What is statistics? *Journal of the Royal Statistical Society, Series A,* 158(1), 1-20.

Bishop, G. (1998). Experimental design training on the world wide web. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics* (pp. 205-210). Singapore: International Association for Statistical Education and International Statistical Institute.

Box, G. E. P., Hunter, W. J., & Hunter, J. S. (1978). *Statistics for experimenters. An introduction to design, data analysis and model building*. New York: Wiley.

Coleman, D. E., & Montgomery, D. C. (1993). A systematic approach for planning for a designed industrial experiment. *Technometrics,* 35(1), 1-12.

Chatfield, C. (1995). *Problem solving: A statistician's guide*, second edition. London: Chapman.

Connolly, J. (1996). Integrating the teaching of statistics, computing and research methods for biologists. In J. M. Ocerin, A. M. Guzman, & J. M. Serrano (Eds.), *Design of experiments and statistical education in agriculture. III Harma Meeting.*

Gal, I. (1998). Assessing statistical knowledge as it relates to students' interpretation of data. In S. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K-12* (pp. 275-295). Mahwah, NJ: Erlbaum.

Galmacci, G. (2001). The impact of the Internet on researchers' training. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 159-169). Granada: International Association for Statistical Education and International Statistical Institute.

Ganesh, S., & Ganesalingam, S. (1998). Teaching basic statistical concepts using computer display tools. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics* (pp. 1001-

1008). Singapore: International Association for Statistical Education and International Statistical Institute.

Glencross, M. & Mji, A. (2001). The role of a research resource centre in the training of social science researchers. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 245-257). Granada: International Association for Statistical Education and International Statistical Institute.

Harraway, J., Manly, B., Sutherland, H., & McRae, A. (2001). Meeting the statistical needs of researchers in the biological and health sciences. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 177-195). Granada: International Association for Statistical Education and International Statistical Institute.

MacKay, R. J., & Oldford, W. (1994). *Stat231 course notes*. Waterloo: Canada: University of Waterloo.

Moore, D. S. (1998). Statistics among the liberal arts, *Journal of the American Statistical Association*, 93(444), 1253-1259.

Moore, D. S., & McCabe, G. P. (1999). *Introduction to the practice of statistics*, third edition. New York: Freeman.

Pfannkuch, M., & Wild, C. J. (1998). Investigating the nature of statistical thinking. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics* (pp. 459-466). Singapore: International Association for Statistical Education and International Statistical Institute.

Puranen, J. (1998). WWW and teaching statistics-a teacher's point of view In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics* (pp. 993-998). Singapore: International Association for Statistical Education and International Statistical Institute.

Stangl, D. (2001). Design of an Internet course for training medical researchers in Bayesian statistical methods. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 147-157). Granada: International Association for Statistical Education and International Statistical Institute.

Svennson, E. (2001). Important considerations for optimal communication between statisticians and medical researchers in consulting, teaching and collaborative research with a focus on the analysis of ordered categorical data. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 23-35). Granada: International Association for Statistical Education and International Statistical Institute.

Talbot, M., Horgan, G., Mann, A., Alonso, R., Badia, J., Bishop, G., & Quednau, H. (1998). SMART: Introducing specialist statistical techniques via the Web. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics* (pp. 409-414). International Association for Statistical Education and International Statistical Institute.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry, *International Statistical Review* 67(3), 223-265.

Glenys Bishop
Statistical Services Branch,
Australian Bureau of Statistics,
PO Box10, Belconnen 2616, Australia
Email: glenys.bishop@abs.gov.au

Mike Talbot,
Biomathematics & Statistics Scotland, JCMB, King's Buildings
Edinburgh, EH9 3JZ, United Kingdom
Email: mike@bioss.sari.ac.uk

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 227-230.

Printed in Granada, Spain.

GILBERTE SCHUYTEN

DISCUSSION. RESEARCH SKILLS: A CLOSELY CONNECTED TRIPLET OF RESEARCH AREA, RESEARCH METHODOLOGY AND STATISTICS

In the discussion document of the theme of the *IASE Round Table Conference* it is stated that: 'Statistics is an important component in the training of new researchers within masters and doctorate courses'. Indeed within many of these programs in non-statistical disciplines courses such as statistics and research methodology are included.

Winer's book (1962) is an old standard book in the biosciences, which focuses on statistical manipulation of experimental data. This book belongs to the Fisher tradition (Fisher, 1925). The book of Cook and Campbell (1979) is another old standard book well known in psychology and social sciences. Here the focus is on general methodological issues and statistical techniques are described in terms of different research designs. This book belongs to the McCall (1923) tradition with focus on the methods of securing adequate and proper data to which to apply statistical procedure.

At the end of those programs students are supposed to possess adequate research skills. It is often heard that they master these skills in an insufficient way. In addition to this lack of mastery of statistical skills by the end of programs, researchers need to update their statistics knowledge in order to keep in touch with recent developments in statistics. This line of thought leads us to consider different types of researchers and the methodological and statistical thinking skills required in the research process. The intertwining of general research methodology and statistics is illustrated in Section 4 by the different treatment of the key concept 'error'. Finally, some considerations are given concerning the place of statistics and/or general research methodology courses in student programs and instruction type used.

1. OVERVIEW OF PAPERS

Table 1 summarises the type of researchers addressed by the papers of Bishop and Talbot, Harraway, Manly, Sutherland and McRae, and Saville.

Table 1 Overview of Papers

	Bishop	Harraway	Saville
Target Populations	Students; Postgraduates	Students; Postgraduates; Professionals	Professionals
Statistics & Methodological Components	Methodological components are emphasised	Relation between statistics methods & scientific inference	Both components are stressed
Training	When required Web delivered	Special short courses Workshops Formal lectures	Workshops Projects Essential ingredients

The attention given to statistical and methodological competencies and the way the training is organised.

Bishop discusses the training of master's and doctoral level students and emphasises the methodological component; the training is web delivered and done as required. The basics of statistics are included in the background materials.

Harraway trains the three types of researchers. He organises the training in different ways: specialised short courses, workshops and formal lectures. He stresses the relationship between statistical methods and scientific inference.

Saville trains the professional researcher and organises his training in workshops and projects.

The papers by Harraway and his colleagues, Saville and Bishop and Talbot all reflect the intertwining between research area, general research methodology and statistics and discus how the training of research skills can be organised.

My comments below are organised into four topics: 1) the researcher type, 2) research skills, statistical and methodological competencies and the research process, 3) concept of ‘error’ in research methodology and statistics, and 4) training in research skills.

2. RESEARCHER TYPE

Three types of researchers are considered in the three papers: (1) the students, other than a doctoral student as researchers, (2) the doctoral student as researcher, and (3) the professional researcher.

The dimensions for assessment of mastering research skills are not really different for the master and doctoral students but the expectations or standards for a doctoral student are higher. Both have to deliver a research project: The first to obtain a masters degree, the second to obtain a Ph.D.. Master's students are required to activate and apply their statistical and methodological knowledge. The needs of professional researchers working at research institutes are of a different kind; here updating of their research methodology and statistical techniques are the core matter. Doctoral students are between the two, as they are required to activate and apply their knowledge and to acquire new developments in statistics as imposed by the research questions. Activating, applying and updating of statistical and methodological knowledge are core activities involved in a research process.

3. RESEARCH SKILLS

Research skills can be broadly categorised into two types (Schuyten, 1991):

1. Skills needed to read and evaluate surveys, experiments and other studies dealing with substantive problems in the research area;
2. Skills needed to do research while planning a study, analysing the data, interpreting and generalising the results. To this second category ‘reporting the results’ can be added (see in the different papers included in this book). Both categories of skills rely on statistical and methodological competencies.

Empirical research is generally characterised by an empirical cycle with five phases: Describing phenomena, constructing a theory, formulating hypotheses, testing hypotheses and adapting the theory. In this cycle researchers start in the real world by

collecting data, move to the theoretical world of the specific discipline and confront their findings in the theoretical world again with the data collected in the real world. In bridging the gap between the two worlds general research methodology and statistics play an important role.

Let us focus on the research process while moving from the theoretical towards the real world. Here the researcher goes from formulating (phase 3) to testing hypotheses (phase 4); the interplay between general methodological competence and statistical competence is crucial. First a conceptual framework, which results from answering the methodological questions 'why' and 'what', has to be worked out. Here there is a strong interplay of specific discipline and general research methodology. From this conceptual framework the hypotheses have to be operationalised such that the collection of data can start and the answers to the methodological questions 'where' and 'how' are given by the design. In answering the 'where' and 'how' questions researchers need knowledge about the connections between design and the forthcoming statistical analyses. The choice of statistical techniques is induced by the design.

Once the statistical analysis and interpretation of the results is done, the researcher moves again from the real world to the theoretical world by generalising the results and/or collecting new data. This generalising depends again on the design. In this description conceptual framework and design are important components in the research process. The conceptual framework links the specific discipline with general research methodology; the design links general research methodology with statistics. The conceptual framework induces the choice of the design and the design induces the choice of the statistical techniques. In this description of the research process, research methodology is strongly emphasised and statistics is seen as the servant. Nevertheless both are strongly connected. We will discuss one important key concept 'error' to illustrate this.

4. THE CONCEPT OF ERROR IN RESEARCH METHODOLOGY AND STATISTICS

Variation in the real world makes research a challenging endeavour. In danger of oversimplification we consider two types of sources of variation: an intended one -the true variation we want to study- and the non-intended one which can be split again into systematic and random error. Statistics deals with random error. The non-intended systematic variation can be reduced by more adequate operationalisation, which means appropriate instrumentation, sampling and appropriate design. Research methodology helps in avoiding systematic error; statistics helps in handling random error.

The following illustration clarifies this non-intended error in the case of the statistical technique of 'analysis of variance' applied with two different designs: a random and a block design. Suppose that we have three treatments and one dependent variable. The within variance is taken as the error variance in the calculation of the F-value. This error variance can contain a lot of unknown non-intended systematic error variance that can be taken out by using a block design. In general using more appropriate designs reduces the non-intended systematic variation. Researchers are not always fully aware of the consequences of choosing a specific design. In statistics classes the focus is on techniques and the design is subordinated. Why a particular design has been chosen is often not fully addressed. In methodology classes the focus is on the design and the appropriate statistical technique is subordinated. This brings us to the question 'how should we train people in these research skills?' such that integration of methodological and statistical competencies can be done.

5. TRAINING IN RESEARCH SKILLS

Nowadays a lot of discussion is going on in the health sciences and the social sciences about the problem-based, inductive and deductive approach of training research skills. By the inductive approach we generally mean a problem-based approach based on project-like work and co-operative work

I would like to call this 'the triplet approach' where substantive theory, research methodology issues and statistics are all three emphasised and linked. In a research problem dealing with correlation, for instance, focus is not on the statistical theory of correlation but on the consequences of using a specific instrumentation on the choice of appropriate statistical technique. A functional procedural knowledge of statistics is needed rather than an analytical, conceptual one. This approach is mainly problem driven.

By a deductive approach we mean a theoretical approach with examples and applications. This approach is driven either by statistics or by research methodology. In this approach there are usually two courses, one statistics course and one methodology course. Starting training with a triplet approach may cause a lot of problems. A deductive generic course dealing with the 'old standards' of statistics, as well as a deductive generic course dealing with the 'old standards' of research methodology provided with sufficient bridges between both is needed. Building on these two generic courses where the basics of statistics and research methodology are dealt with, research skills can be practised in a problem driven course; it is only by doing research that research skills can function in an integrated way.

REFERENCES

Bishop, G., & Talbot, M. (2001). Statistical thinking for novice researchers in the biological sciences. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 215-226). Granada: International Association for Statistical Education and International Statistical Institute.

Fisher, R. A. (1935). *Statistical methods for research workers*. London: Oliver & Boyd.

McCall, W. A. (1923). *How to experiment in education*. New York: Macmillan.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-Experimentation*. Boston: Houghton Mifflin.

Harraway, J., Manly, B., Sutherland, H. & McRae, A. (2001) Meeting the statistical needs of researchers in the biological and health sciences. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 177-195). Granada: International Association for Statistical Education and International Statistical Institute.

Saville, D. S. (2001). A hands-on interactive method of teaching statistics to agricultural researchers. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 197-213). Granada, Spain: International Association for Statistical Education and International Statistical Institute.

Schuyten, G. (1991). Statistical thinking in psychology and education. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (pp. 486-489). Voorburg, The Netherlands: International Statistical Institute.

Winer, B. J. (1962). *Statistical principles in experimental design*. New York: McGraw-Hill.

Gilberte Schuyten
University of Gent, Department of Data Analysis
Henri Dunantlaan 1, 9000 Gent, Belgium
E-mail: gilberte.schuyten@rug.ac.be

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 231-244.

Printed in Granada, Spain.

CAROL JOYCE BLUMBERG

TRAINING REGULAR EDUCATION AND SPECIAL EDUCATION TEACHERS IN THE USE OF RESEARCH METHODOLOGY AND STATISTICS

The purpose of this paper is to discuss the needs of primary and secondary level teachers seeking Master's degrees (second degrees) in terms of knowledge and use of research methodology and statistics, with special consideration of the needs of Special Education teachers. In particular, the following topics will be discussed: goals and organisation; descriptive statistics; inferential statistics; specific hypothesis testing procedures; experimental and quasi-experimental designs; survey research and sampling techniques; qualitative data collection techniques; and reliability and validity. The paper ends with a discussion of what special education teachers need to know concerning meta-analysis and single-subject designs and with some miscellaneous comments.

1. INTRODUCTION

The purpose of this paper is to discuss the needs of primary and secondary level teachers seeking Master's degrees (that is, second degrees) in terms of knowledge and use of research methodology and statistics, with special consideration of the needs of special education teachers. In the USA, the term special education teacher refers to any teacher who has taken extra training (usually as part of a first degree) to be able to teach students with disabilities, such as mental retardation (the exact term used actually varies from state to state in the USA), learning disabilities, emotional problems, and physical disabilities. The term research methodology and statistics is used here to mean the traditional statistical methods, as well the related areas of statistical thinking, logic of hypothesis testing, experimental and quasi-experimental design, survey research design, sampling, qualitative data collection techniques, reliability and validity.

To the best of the author's knowledge, almost all Master's degree programs in Education or Special Education in the USA and England include such courses. In this paper the discussion will be in terms of all teachers who are pursuing second degrees in Education or Special Education, except where otherwise noted. However, since there is a special emphasis in this paper on Special Education teachers, many of the references will be in terms of special education even though other, more general, references were sometimes available.

2. TEACHERS AS RESEARCHERS

While it may seem strange at first to label a classroom teacher also as a researcher, it is not. All teachers are continually collecting and analysing data on students' academic

performance. Special Education teachers, besides collecting and analysing data on academic performance, also must perform well-designed experiments in their classrooms with individual students or small groups of students (most often using single-subject design principles) to assess what "treatments" will be most effective for educating each student (Tawney & Gast, 1990). This is especially important in countries where, by law, each student with a disability must have an individual education plan (often abbreviated, IEP in the United States).

3. GOALS AND ORGANIZATION OF RESEARCH METHODOLOGY AND STATISTICS COURSES

The two major goals of most research methodology and/or statistics courses for teachers completing a Master's degree are to: i) help them critically read, correctly interpret, and decide the validity of conclusions in the published and unpublished literature, and ii) give them the tools necessary to complete research-based projects and/or theses required for their Master's degrees (Cooke, Test, Heward, Spooner, & Courson, 1993; Gall, Borg, & Gall, 1996). For Special Education teachers, there is a third major goal; they should learn how to better collect and analyse data that will be used to help determine each student's IEP. These three goals are exactly the same as those stated by Burrows and Baillie (1997) for the research methodology component of a Diploma in Nursing course in the UK with the phrase "help determine each student's IEP" above replaced by "enhance the quality of care delivered to the client" (Burrows & Baillie, 1997, p. 35).

Only one study of the content of the research methodology and/or statistics courses for Master's degree students in education could be found in the literature. Todd and Reece (1990) conducted a Delphi study on graduate introductory research courses. At the first stage, "...21 experts nominated by... [their peers] generated a list of 114 skills and knowledge areas they considered worthy objectives in an introductory course (p. 2-3). At the second stage, these 21 experts simplified the list to 59 skills and knowledge areas that are "'Essential', 'Essential-Important', or 'Important.'"(p. 3), where the authors had operational definitions for each of these terms.

This list of 59 skills and knowledge areas was made up of phrases, most of which were vaguely stated. For example, two of the areas were "Demonstrates an understanding of research methodology" and "Understands the concept of internal validity" (p. 6). Only a few were specific such as "Can formulate a testable hypothesis" (p. 6) and "Has an understanding of the standard deviation" (p. 7). This present paper, on the other hand, will provide more detailed guidelines as to how a faculty member might approach the content of research methodology and statistics courses for teachers. Also, reference will not be made in the rest of this paper to the number of experts who stated that certain skills and knowledge areas should be included, since doing so would be misleading, because so few of the specific (that is, not vaguely stated) skill and knowledge areas discussed here were included in Todd and Reece's (1990) list.

One major question that the faculty members in each Master's Degree program need to address in the beginning when designing the research methodology and statistics components of their program is whether or not research methodology and statistics should be combined in one course or split into two courses. Each has its advantages and disadvantages. When doing the literature search for this paper, only one article could be found that gave statistics on the number of courses required by various Master's Degree

programs. That article (Calder, Justen, &Waldrop, 1986) was a 1985 survey of 60 Master's Degree programs in Special Education. They found that 6 (10%) of these programs required no research or statistics courses, 29 (48.3%) required one course, 19 (31.7%) required two courses and 6 (10%) required three courses. For those programs that required only one course, they did not specify whether this was a research methodology course, statistics course or some combination of both. They did say that for the programs requiring two courses that most programs had one course each in statistics and in research methodology. No discussion could be found in the literature of the advantages and disadvantages of the various course combinations. So, the advantages and disadvantages listed here are the opinions of the author of this paper based on 20 years of experience in teaching research methodology and statistics.

The advantage of having research methodology and statistics combined into one course is that it is more efficient in terms of delivery of information. That is, unnecessary overlap between courses is eliminated. A second advantage of having one course is that the students are less intimidated before the course by the thought of taking a statistics course. The disadvantage of having the two combined is that is extremely hard to find faculty to teach such a course who are experts in both the research methodology used in education and in statistics. The advantage of separating research methodology and statistics into two courses is that it is easier to find faculty capable of teaching the courses. It is also true that that by separating the course, there will be overlap in the material covered. Although others may look at this as a disadvantage, the author of this paper feels that it is an advantage as long as the faculty members teaching the two courses communicate well with each other to minimise needless overlap.

4. DESCRIPTIVE STATISTICS

Before discussing the specifics of what teachers pursuing Master’s degrees need to know about descriptive statistics and other topics, it should be pointed out that in the USA few first-degree (Bachelor’s) programs require students preparing to be teachers to take a course in statistics. Hence, for this paper the assumption is made that the students enrolled in a Master’s degree course in research methodology and statistics have no knowledge of statistics and probability other than how to compute the mean, median, and mode.

It is extremely important that several hours are spent on the proper use and interpretation of descriptive statistics since these techniques are so widely used in the literature, as well as in the technical portions of standardised test manuals. Elmore and Woehlke (1998) have pointed out, that out of the 1906 articles examined that were published in the three journals of the American Educational Research Association (AERA) from 1978 to 1997, graphical methods were used in 58 (3.0%) and descriptive statistics were used in 246 (12.9%) of the articles. In fact, descriptive statistics were the second most common technique used, with only the use of ANOVA, ANCOVA, and MANOVA combined being more common. It should be noted here that these percentages are artificially low since Elmore and Woehlke used all articles published as their population, whether or not the article used any methods of data collection or data reporting. Baumberger and Bangert (1996) found the same order, but because of the way they reported their results no exact percentages can be given here. Heckenlaible-Gotto and Choi (1993) found that 34% of the 94 articles they examined from the 1990 to 1993 issues of *Topics in Early Childhood Special Education* used descriptive

statistics. Those interested in the percentages of use of statistical techniques in the biological and health sciences are encouraged to read the paper by Harraway, Manly, Sutherland, and McRae (2001).

Graphics: Some discussion needs to be done of how to present readable (in terms of axes and legends) histograms or bar charts of both frequencies and relative frequencies. Also, since boxplots and stem-and-leaf displays are increasing every year in their occurrence in journal articles and in textbooks for students of ages 10 and up, it is important that these teachers know how to interpret both boxplots and stem-and-leaf displays. They should also learn how to make a stem-and-leaf displays by hand. It was two Special Education teachers who mentioned to the author that they had been asked for help with how to construct a stem-and-leaf display by Special Education students who were mainstreamed for mathematics. Neither was able to help the students and both urged the author to include the construction of stem-and-leaf displays in her course. The term mainstreaming refers to when a special education student is taught mathematics (or some other subject matter area) in a regular classroom with either only a regular teacher present or with both a regular teacher and Special Education teacher present.

Measures of central tendency: Some discussion of when each of the measures of central tendency (that is, the mean, trimmed mean, median, and mode) should be used needs to take place. Most teachers can compute a mean, median, and mode. But, few have any knowledge as to when each should and should not be used.

Measures of variation & sampling distributions: Although most teachers regularly see standard deviations and standard errors, they do not know how to compute or interpret them. They are especially confused by the term standard error, even though it occurs in almost all standardised test manuals. So, the idea of a sampling distribution needs to be introduced to help them understand how a standard deviation and standard error are different. They also need to understand the difference between sampling error and other types of errors in data collection.

Hand/computer calculations: As many students in many different courses have pointed out to the author over the years, sometimes you learn some statistics concepts better by doing them step-by-step by hand/calculator first and then learning how to do them on a computer or using the special keys on calculators. Hence, it seems important that teachers be taught to do stem-and-leaf displays, histograms, means, medians, modes, variances, and standard deviation step-by-step by hand and/or by calculator. The inclusion of means here may seem strange, but there have been students in the author's Master's level research methodology class over the years that did not know how to compute the mean both by hand and by a calculator.

Use of computers: Since the computer packages available to teachers at their places of employment or at home change quickly, not much time should be spent training them in the use of a specific computer package. On the other hand, it is important that they have some exposure doing the descriptive statistics discussed in this section of this paper on a computer using a statistics package, so that they can begin to understand the logic of statistics packages. Further, the examination of output from at least two different statistics packages is important so that teachers realise how different the output may be from one package to another.

5. INFERENTIAL STATISTICS

The logic of hypothesis testing: In order to be able to critically read and correctly

interpret research-based articles, teachers must understand the logic of hypothesis testing, including the ideas of null hypothesis versus alternative hypothesis, of population parameter versus sample statistic, and of critical region including the correct use of the terminology of "do not reject the null" and "reject the null". Extensive discussion of the difference between statistical significance and practical (also called, meaningful or educational) significance also needs to take place.

Writing hypotheses: It is extremely important that teachers learn how to write both null and alternative hypotheses that are well defined (that is, operationally defined). By learning how to write hypotheses where all terms are operationally defined, they can better figure out what data needs to be collected and can get more useful help from a statistician before they collect data.

Probability values: There must be a careful discussion of probability values in these courses, since almost all published articles that use hypothesis tests in education present p-values. It is not necessary to have the teachers learn how to calculate probability values. They must, however, be able to correctly interpret probability values and be able to read them off of computer outputs.

Type I error, Type II error, and power: The distinction between Type I and Type II errors needs to be discussed. The concept of power should be introduced informally. They must understand conceptually how sample size, differences between means, and variability affect power and how power and Type II errors are related.

Confidence intervals: Teachers, like other disciplines at the Master's degree level, need to be able to correctly read and interpret confidence intervals. They must understand the correct way to write out and verbally state the interpretation of a confidence interval in language that parents can understand. They should also realise, as with t-tests, there are many different types of confidence intervals and one must be very careful as to what parameter or difference in parameters in being estimated by a confidence interval. In addition, because teachers deal with students as individuals, they must understand informally the difference between confidence intervals on means versus confidence (or prediction) intervals for an individual student's test scores or other measures of interest. The related topic of standard error of measurement also needs to be discussed.

Generalisation: In applied research there are two types of generalisations. The first is from the sample to the population that it represents. This is what inferential statistics does. The second type is the logical extension of statistically significant results from the population that the sample represents to other populations. This second type of generalisation is actually the one done more often in education. It is important that teachers learn to appreciate that both types of generalisation are acceptable as long as the people doing the generalisation make it clear which type they are doing.

Consulting with a statistician: It should be made clear to these teachers that the best way for them to complete their theses or other research papers required for their degree is to consult with a statistician before they do any inferential statistics. In fact, they should be strongly urged to consult with a statistician when they design their studies, before they begin data collection, as changes are being made during data collection, and before beginning their data analysis.

6. SPECIFIC HYPOTHESIS TESTING PROCEDURES

This is a big question in the author's mind as to how much these teachers need to

know about specific hypothesis testing procedures. For all of these procedures they should not have to compute anything by hand/calculator or by computer, but rather be able to interpret results.

Analysis of variance/analysis of covariance: Some discussion of how to read the results from ANOVAs, MANOVAs and ANCOVAs as reported in journal articles is absolutely necessary since these are the techniques used the most often according to Elmore and Woehlke (1998) and Baumberger and Bangert (1996). In fact, 309 (16.2%) articles out of 1906 articles appearing between 1978 and 1997 in the three AERA journals (Elmore & Woehlke, 1998) used these methods. ANOVA was also the most often used primary statistical technique in the special education literature in the period of 1984-85, being used in 27% of the random sample of 104 articles that Swanson and Alford (1987) examined. Teachers taking a research methodology and/or statistics course should have some conceptual understanding of the ideas of Sum of Squares, Mean Squares, and F-ratios. They should also be shown how to get the Sum of Squares, Mean Squares, degrees of freedom, F-ratios, and p-values from the output generated by one or two computer packages. The author's preference is for two computer packages here that present the results of an ANOVA differently.

It is also very important to make sure that teachers understand the ideas of main effects versus interactions and the ideas of covariates. They should understand the importance of correctly interpreting statistically significant interactions since interactions are extremely common in educational settings. For example, Torgesen and Dice (1980) found that in 62% of studies where the student (or subject) by experimental condition interaction was examined, this interaction was found to be statistically significant. Some discussion of how to interpret statistically significant main effects when there are also statistically significant interactions present should also take place, although probably only through the use of graphical displays of the appropriate means for the main effects and the interactions.

Correlation, regression, and contingency table analysis: Correlations and multiple regression are a very popular statistics in education, ranking as the fifth and third most used techniques, respectively, in articles published from 1978 to 1997 in the three journals of the AERA (Elmore & Woehlke, 1998). Contingency table analysis (that is, chi-square tests) was the primary statistical technique used in 7.7% of the articles examined by Swanson and Alford (1987). Correlation is also extensively used when studying reliability and validity of standardised tests. Hence, students must be able to interpret a correlation in terms of direction (positive versus negative) as well as in terms of practical significance and statistical significance. Often, R^2 is reported in the education literature. Hence, some discussion of how to interpret R^2 needs to be done as well as a discussion of the relationship between correlation and R^2. Time must also be spent helping these teachers understand how to interpret the results of regression and of contingency table analyses as reported in journal articles. They should also understand some of the major advantages and shortcomings of correlation, regression, and contingency table analysis. As Estepa and Sanchez Cobo (2001) have pointed out, the interpretation and correct use of measures of association is a difficult one for students taking a first course in statistics.

T-tests: Some discussion must take place about the various types of t-tests so that these teachers know to read and interpret the results carefully for t-tests appearing in research articles. The first several times the author taught a research methodology course to teachers she avoided introducing formulas. However, one time several members of the class asked to see the formula for a t-test. The author then showed them

the formula for an independent samples pooled t-test. It was amazing how much they learned from seeing the formula for the test statistic. Hence, some discussion of the formulas for a one-sample t-test and an independent samples t-test should be part of the course. Some explanation should also be included about how the various pieces of each of the formulas relate to the power of the test.

7. EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGN

Most of the projects and theses that Master's degree students complete have fairly simple designs. Hence time should not be spent teaching them a catalogue of complicated designs. They should also be exposed to the more common true experimental designs used in educational research: the Pretest-Posttest Control Group, the Posttest-Only Control Group, and the Solomon Four-Group (Gall, Borg, & Gall, 1996).

Teachers also need to be aware of the importance and usefulness of multi-factor designs. They also need to know the difference between random sampling and random assignment and why both are important. Further, time should be spent explaining why both random sampling and random assignment are hard to achieve in educational settings. This is especially true in special education because of legal and ethical requirements. For example, the USA (Public Law (P.L.) 94-142--The Education for All Handicapped Children Act), Israel (Michael, 1989), and Italy (Italian Law 517, 8/04/1977 and Italian Law 104, 2/05/1992) have laws that require special education students to be educated in the least restrictive environment possible and be placed in special classrooms only when their needs cannot be met in regular classrooms. The references for the Italian laws are from Balboni and Pedrabissi (2000). Hence, most well-designed research studies in special education settings (and even many in regular educational settings) are quasi-experimental. For example, Swanson and Alford's (1987) examination of 179 research articles in special education for the period of 1984-85 found that 0% were true experimental, 68% were quasi-experimental, and 32% were other types (for example, survey research or case studies). Hence, extensive discussion needs to take place on how to best design research studies using quasi-experimental methods that minimise bias as much as possible.

The proper analysis of quasi-experimental designs is still a matter of much debate among statisticians and can be very tricky (Cook & Campbell, 1979; Elashoff, 1969; Maxwell & Delaney, 1999). So, it is especially important that teachers doing research using quasi-experimental designs be urged to consult a statistician before they collect their data, if they need to make any changes while collecting their data, and before they begin their data analysis.

In education, the terms correlational studies and causal-comparative studies are often used. The term correlational study does not usually mean a study that is analysed using correlation. It is used to designate a study where one can only argue a relationship between two or more variables, but cannot argue that changes in one variable caused changes in another variable. Causal-comparative studies, on the other hand, are those where one can argue that changes in one variable caused changes in another variable. The distinction between these two terms, and why this distinction is important, needs to be made clear to these teachers.

8. SURVEY RESEARCH AND SAMPLING TECHNIQUES

Because of their perceived simplicity, surveys are a very popular instrument used in Master's degree research papers and theses in education as well as in schools and classrooms. For example, of the 19 theses and professional papers done by Master's degree students at Winona State University to which the author had access (see Appendix for a list), surveys were a major method of data collection in ten of them. Swanson and Alford (1987) found that survey research or other similar descriptive techniques were the basis for 18% of the articles they examined. Bruininks, Wolman, and Thurlow (1990) have pointed out that almost all follow-up studies of special education service programs include survey research. A follow-up study is one in which those who have graduated from and/or dropped-out of a special education program are contacted some number of years later and asked a series of questions. The answers to these questions are often the core of reports to local school districts, state agencies, and federal agencies.

Also, since almost all of these teachers will design a survey at sometime in their professional life it is important that substantial time be spent on how to write surveys that are clear and concise and collect useful information. Extensive discussion needs to take place of the various formats for items on surveys such as rating scales, semantic differentials, multiple choice, and open-ended items. Some discussion of the number of alternatives presented on a rating scale needs to take place. For example, should it be a 3-, 4-, 5-, 6-, or 7-point scale and should the respondents be allowed to choose neutral as a response? The issue of why it is essential to collect accurate background demographic information also needs to be discussed.

The related issue of confidentiality needs to be discussed, although classroom teachers are much more aware of the issues of confidentiality than any other researchers the author has worked with over the last 25 years. Further, they need to be made aware of the different ways of collecting data via mail, telephone, in-person, and the computer (both via email and the Internet). Some discussion of the merits and shortcomings of each method in a few different situations also needs to take place.

In addition these teachers need to be taught how to correctly use a random number table either in a book or on a computer. They also need to be shown good examples of the different sampling methods of simple random sampling, systematic random sampling, stratification, and clustering. Some discussion of how to make good strata and clusters also needs to be done. They should also see an example where a combination of sampling techniques was employed.

9. QUALITATIVE DATA COLLECTION TECHNIQUES

Time must be spent showing teachers a variety of qualitative data collection techniques. In fact, qualitative techniques were the fourth most popular techniques used in articles published in AERA journals between 1978 and 1997 (Elmore & Woehlke, 1998). Another reason these techniques should be discussed is that it is important for teachers to think more broadly when designing their research studies.

It is especially important that observational techniques be discussed since observational studies are extremely common in education and in special education, in particular (Greenwood, Peterson, & Sideridis, 1994-95). In fact, the use of observational techniques predates the use of quantitative techniques in special education (Hulek,

1983). Some of the issues that need to be discussed regarding observational techniques are minimal-bias data collection, the variety of recording and data-coding procedures (such as frequency-count recording and interval recording) available and the difficulty of properly analysing observational data.

Techniques such as case studies, cross-case analysis, ethnography, evaluation, and interviews also need to be discussed since, along with observational studies, they are the most popular qualitative techniques used in special education research (Crowley, 1994-95). Some general discussion of data coding and analysis under these techniques needs to occur. It needs to be emphasised that when data is collected using observational or other qualitative techniques, the issue of data analysis is a tough one. Further, it should be explained that often a variety of qualitative and quantitative techniques need to be used together to properly collect data (McWilliam, 1991) that will answer the research questions and/or hypotheses of interest.

10. RELIABILITY AND VALIDITY

Depending on the requirements of the individual State, Province, or Country, teachers may have had from very little exposure to extensive exposure to the topics of the reliability and validity of tests and other data collection instruments when pursuing their first degrees and initial teacher certification. Fairly typically in the States of Minnesota and Wisconsin (where the teachers the author has as students come from), there is little to no training in reliability and validity. It is important for these teachers to learn about reliability and validity for two reasons. First, they will be able to better understand the technical portions of the standardised test manuals. Second, when designing their own research projects they will be able to think better about these two topics. It is amazing how often teachers can make their studies much more useful by thinking about the reliability and validity consequences of the instruments they plan to use before collecting their data.

In terms of content here, they need to see at a minimum the ideas of internal consistency reliability (Cronbach's coefficient α), parallel forms reliability, reliability over time (often called, test-retest reliability), inter-rater reliability, content validity, construct validity, and criterion-related (both concurrent and predictive) validity. It also needs to be emphasised that these terms refer to specific measurements, while the terms internal validity and external validity refer to research studies as a whole.

11.ADDITIONAL TOPICS THAT NEED TO BE COVERED IN COURSES FOR SPECIAL EDUCATION TEACHERS

Meta-analysis: Meta-analysis has become a very popular technique in Special Education journals because of the small sample sizes that are inherent in research studies in special education due to the rarity of certain disabilities, such as autism or the combination of severe hearing and visual impairments or for other reasons. It should be remembered that meta-analysis was popularised first by educational statisticians (Glass, McGaw & Smith, 1981; Hedges & Olkin, 1985) and then spread to other areas of inquiry. As early as 1984 and 1985-86, the *Journal of Special Education* (1984; 1985-86) devoted two issues to meta-analysis and the more general issue of research synthesis. Stanovich and Stanovich (1997) have pointed out meta-analysis is also very

important in the area of Special Education because special education teachers have more faith in a meta-analysis of studies than in individual studies. Hence, Special Education teachers must be able to understand the logic of meta-analysis, have some knowledge of its advantages and disadvantages, and be able to interpret effect sizes in the simplest cases.

Single-subject design: In the USA there is a wide variety in the backgrounds of Special Education teachers in terms of knowledge of single-subject designs from having had an entire course in single-subject design prior to being certified as a Special Education teacher to having never even heard of the term, with most teachers having heard of single-subject design and having been taught how to use a few simple designs in their classrooms (Cooke, Test, Heward, Spooner, & Courson, 1993). In this course, a wide variety of single-subject designs need to be introduced along with how to implement and analyse them properly (Tawney & Gast, 1990). At a minimum, ABA (where A is the first treatment (usually the control treatment) and B is the second treatment), ABAB, other alternating treatment designs, and multiple baseline designs need to be introduced since these are the most popular single-subject designs used in educational research (Gall, Borg, & Gall, 1996). The teachers also need to understand how these designs can be used in rigorously done research as well as in the classroom.

Some discussion of how many observations need to be taken at each phase and how to determine the time intervals for collecting observations should be included. Further, it should be emphasised that the term single-subject design is a bit misleading. Some of the most useful single-subject studies are those where the same single-subject design is implemented on several students and the results discussed both individually and as a group (Tawney & Gast, 1990). Finally, some discussion of the limitations of single-subject design needs to take place including the limitations of the usual method of simply using a visual analysis. Since there has been much discussion over the last 30 years, at least, of how to statistically analyse single-subject designs (see, for example, Kratochwill & Levin, 1992), the topic of the statistical analysis of single-subject designs should probably not be discussed.

12. MISCELLANEOUS COMMENTS

Statistical thinking: The author does not see this as a separate topic in a research methodology or statistics course for teachers. Rather, it underlies the teaching of all of the material. Students often comment that after they have taken a Master's level research methodology or statistics course (the author has taught a Master's level statistics course for nursing many times), they think differently. When they explain what they mean by this it sounds like what is now called by the term statistical thinking.

Bias, external validity, and internal validity: These ideas should be discussed early and often throughout the course. As was pointed out several times in this paper, rarely is true experimental research possible in education and special education. Hence, the research which teachers read and do will contain various types of bias and internal validity problems. These biases and internal validity problems will limit the external validity (that is, the ability to generalise) of these studies. On the other hand, some generalisation can often be done and these teachers need to realise that just because a study contains bias or internal validity problems it is not useless. When reading studies they have to learn how to recognise bias problems and to decide whether or not the biases are too severe for the study to be of use to them. When doing their own studies,

they should try to eliminate as many bias problems as possible. These teachers must also be taught to be honest when reporting their results and to make clear how their sampling was done and admit to any other bias problems of which they are aware.

Replication: Some discussion needs to take place of the good points of doing replication studies. In fact, the only bad point of doing a replication study in the past was that many journals and university promotion and tenure committees in education would not accept them as being worthwhile. The situation is changing, however, as Gersten, Lloyd and Baker with others (1998) have pointed out. The good points of doing replication studies at the Master's degree level is that the teacher has a model to follow closely and hence the design of the study is made easier. Second, by having Master's degree students and others do replication studies, the quality of meta-analyses and other research syntheses will be improved because more studies will be available that can be compared and contrasted (Stanovich & Stanovich, 1997).

Other statistics topics: There are other statistics topics that have not been discussed in this paper because they occur less often in studies in education and special education. For the topics of factor analysis and structural equation modelling, the author has examples from journal articles that she distributes to her class and gives them a quick (10 minute) explanation of how to read the tables and diagrams for each technique. When the students in the class encounter other techniques in articles that they are reviewing as part of the course requirements, the author uses that article (or another one that is clearer if she has one) to explain a very little bit about the technique. Some of the statistical techniques that are used occasionally in special education are, in alphabetical order, ARIMA modelling and other time-series techniques, Bayesian analyses, cluster analysis, discriminant analysis, generalizability theory, log-linear models, multidimensional scaling, non-parametric and distribution-free techniques other than chi-square contingency table analysis (with, as mentioned in Section 6, chi-square contingency table analysis being a popular technique), small space analysis and survival analysis.

Training of first-degree (Bachelor's) students

As mentioned earlier, in the USA, the tradition is still not to include training in research and statistics methodology at the Bachelor’s level in programs preparing students to become teachers. In many colleges and universities, courses in educational measurement and/or evaluation are included. Most of these courses, however, contain very little research or statistics methodology. Interestingly, each year more and more disciplines are requiring courses in both research methodology and statistics methodology.

At Winona State University, for example, the following departments or majors at the undergraduate level require both a research methodology and a statistics course: Communication Studies, Criminal Justice, Environmental Science, Exercise Science, Marketing, Nursing, Psychology, Social Work, and Sociology. A few others require only a research methodology course and several departments require only statistics.

For the period from 1987 to 2000, the number of sections (of approximately 37 students each) of introductory statistics in the Department of Mathematics and Statistics increased from 10 sections to 30 sections. As for teaching majors, only the major in Mathematics requires a statistics course and none of the teaching majors requires a research methodology course. It is important that all programs training teachers, but especially those training teachers in Special Education, require courses in research methodology and statistics as part of their degree programs so that those teaching have

preparation equivalent to those with non-teaching degrees and are better prepared for graduate level courses.

Interestingly, at Winona State University, starting in January 2000, an introductory course in statistics is being strongly recommended by the Department of Special Education for all its first-degree students preparing to be special education teachers. Further, an educational statistician from another institution contacted the author for suggestions for textbooks for their new Bachelor's level course in statistics in the College of Education since that College of Education would begin teaching statistics at the Bachelor's level for the first time starting in January 2001. Hopefully, these two additions of Bachelor's level statistics for education majors are not isolated cases, but indicative of a trend. It is too early to tell.

A final note

In this paper the author has discussed what she feels is the ideal set of topics and depth of coverage for research methodology and statistics courses for teachers. The course outline and various projects that the author uses have not been included because she is only allowed a total of 25 hours to teach the Master's students in special education both research methodology and statistics, of which 1 1/2 hours are spent having a librarian discuss how to use the library and the Internet to locate source material needed for scholarly reviews of the literature. Hence, her assignments, examinations, and projects are compromises between what she believes should be taught and what she can get done in 25 hours of classroom instruction with about 75 hours of extra work outside of the classroom expected of the students.

APPENDIX: PARTIAL LIST OF THESES AND PROFESSIONAL PAPERS UNDER THE MASTER'S DEGREE PROGRAM REQUIRING THE RESEARCH METHODS COURSES AT THE DEPARTMENT OF SPECIAL EDUCATION, WINONA STATE UNIVERSITY

THESES

Dennison, M. I. (2000). *Pre-service teachers self-reported perceptions of their computer skills computer with demonstrated skills*.

Griffin, K. A. (1997*). A validation of guidelines for the Special Education paraprofessional-teacher team and the student teacher*.

PROFESSIONAL PAPERS

Drazkowski, M. J. (1996). *An analysis of a secondary regular education initiative program.*

Foss-Zylstra, R. S. (1995). *Team teaching to meet the needs of students with the dual label of learning disabled and emotionally behaviourally disturbed*

Grace, J. (1993). *A survey of LD, ED, and CD teachers on their use and opinions of peer tutoring.*

Griffin, R. M. K. (1995). *Effects of repeated readings on the fluency, comprehension and attitude of learning disabled and non-learning disabled readers using a standard fifth grade Social Studies text.*

Gu, W. Y. (1997). *A comparison of Chinese and American fifth-grade samples on measures of arithmetic calculation.*

Hansen, H. (1995). *The effect of review cycles on retention.*

Hill, B. B. (1991). *Eye pursuit characteristics of second grade and fourth grade children.*

Holland, D. (1995). *Acquisition and retention of spelling words for students with mild disabilities using pattern generalisation and frequency word lists*

Jordan, J. (1995). *Descriptive study of educational recommendations given by a Midwestern university hospital clinic.*

Junker, J. (1999). *State self-esteem as effected by teacher mood states.*

Kreisel, C. (1993). *A follow-up survey of Special Education program completers at Winona State University from 1973 to 1984.*

Morris, A. N. (1998). *Assessing the effectiveness of an in-class delivery model for serving students with behaviour problems in the regular classroom: The intervention support team model of the Kenai Peninsula Borough School District Special Services Department.*

Navarre, D. (1990). *A comparison of teacher-student perceptions of the student's locus of control.*

Nyre, M. (1995). *The relationship between Attention Deficit Hyperactivity Disorder and learning disabilities.*

Peterson, C. (1993). *Teacher and parent perceptions of the effectiveness of mainstreaming exceptional educational needs students in a rural Midwestern school district.*

Radatz, D. (1991). *A survey of the post-school outcomes of learning disabled students.*

Traxler, K. (1999). *Family eating patterns and school performance.*

REFERENCES

Balboni, G., & Pedrabissi, L. (2000). Attitudes of Italian teachers and parents toward school inclusion of students with mental retardation: The role of experience. *Education and Training in Mental Retardation and Developmental Disabilities*, 35(2), 148-159.

Baumberger, J. P., & Bangert, A. W. (1996). Research designs and statistical techniques used in the *Journal of Learning Disabilities*, 1989-1993. *Journal of Learning Disabilities*, 29(3), 313-316.

Bruininks, R. H., Wolman, C., & Thurlow, M. L. (1990). Considerations in designing survey studies and follow-up systems for special education service programs. *Remedial and Special Education*, 11(2), 10-17 & 46.

Burrows, D. E., & Baillie, L. (1997). A strategy for teaching research to adult branch diploma students. *Nurse Education Today*, 17(1), 39-45.

Calder, J. E., Justen, J. E., & Waldrop, P. B. (1986). Research requirements in special education programs. *Teacher Education and Special Education*, 9(2), 51-54.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally.

Cooke, N. L., Test, D. W., Heward, W. L., Spooner, F,, & Courson, F. H. (1993). Teachers' opinions of research and instructional analysis in the classroom. *Teacher Education and Special Education*, 16(4), 319-328.

Crowley, E. P. (1994-95). Using qualitative methods in special education research. *Exceptionality*, 3(2), 55-69.

Elashoff, J. D. (1969). Analysis of covariance: A delicate instrument. *American Educational Research Journal*, 6(3), 383-401.

Elmore, P. B., & Woehlke, P. L. (1998). Twenty years of research methods employed in American Educational Research Journal, Educational Researcher, and Review of Educational Research. Paper presented at the *Annual Meeting of the American Educational Research Association* (ERIC Document Reproduction Service No. ED 420 701).

Estepa, A., & Sanchez Cobo, F. (2001). Empirical research on the understanding of association and implications for the training of researchers. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 35-51). Granada: International Association for Statistical Education and International Statistical Institute.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction* (6th ed.). New York: Longman.

Gersten, R., Lloyd, J. W., & Baker, S. with others (1998). *Designing high quality research in special education: Group experimental designs.* Reston, Virginia, USA: Council for Exceptional Children (ERIC Document Reproduction Service No. ED 420 957).

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Thousand Oaks, California: Sage Publications.

Greenwood, C. R., Peterson, P. & Sideridis, G. (1994-95). Conceptual, methodological, and technological advances in classroom observational assessment. *Diagnostique,* 20(1-4), 73-99.

Harraway, J., Manly, B., Sutherland, H., & McRae, A. (2001). Meeting the statistical needs of researchers in the biological and health sciences. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 177-195). Granada: International Association for Statistical Education and International Statistical Institute.

Heckenlaible-Gotto, M. J., & Choi, H. (1993). *The use of research methods and statistics in TECSE, 1990-1993: A content analysis. (*ERIC Document Reproduction Service No. ED 379 831).

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.

Hulek, A. (1983). Changing approaches to methodology of research in special education. *International Journal of Rehabilitation Research,* 6(1), 3-9.

Journal of Special Education. (1984), 18(1).

Journal of Special Education. (1985-86), 19(4).

Kratochwill, T. R., & Levin, J. R. (1992). *Single-case research design and analysis: New directions for psychology and education*. Mahwah, New Jersey: Lawrence Erlbaum.

Maxwell, S. E., & Delaney, H. D (1999). *Designing experiments and analysing data* (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum.

McWilliam, R. A. (1991). *Mixed method research in special education*. (ERIC Document Reproduction Service No. ED 357 554).

Michael, E. (1989). Special education in Israel. *International Journal of Special Education,* 4(1), 59-64.

Stanovich, P. J., & Stanovich, K. E. (1997). Research into practice in special education. *Journal of Learning Disabilities,* 30(5), 477-481.

Swanson, H. L., & Alford, L. (1987). An analysis of the current status of special education research and journal outlets, *Remedial and Special Education,* 8(6), 8-18.

Tawney, J. W., & Gast, D. L. (1990). *Single-subject research in special education*. Paramus, New Jersey: Prentice Hall.

Todd, R. E., & Reece, C. C. (1990). *Graduate introductory educational research courses: A look at what is required.* (ERIC Document Reproduction Service No. ED 324 274).

Torgesen, J. K., & Dice, C. (1980). Characteristics of research on learning disabilities. *Journal of Learning Disabilities,* 13(9), 531-535.

Carol Joyce Blumberg
Department of Mathematics and Statistics
Winona State University
Winona MN 55987-5838, USA
E-mail: cblumberg@winona.edu

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 245-257.

Printed in Granada, Spain.

MICHAEL GLENCROSS AND ANDILE MJI

THE ROLE OF A RESEARCH RESOURCE CENTRE IN THE TRAINING OF SOCIAL SCIENCE RESEARCHERS

At the University of Transkei teaching and research are considered to be two sides of the same coin. Research is thus regarded as a fundamental and indispensable activity. With the University facing the challenge of becoming competitive within the new structure of higher education in South Africa, a strategic plan for research in the Humanities and Social Sciences has resulted in a flexible blueprint for research development. The Research Resource Centre supports this initiative by facilitating research capacity development and research excellence within the University. This paper discusses the role of the Research Resource Centre in the provision of research training for social science researchers and provides details of the various workshops and seminars used to develop skills in the use of statistics and computer-based statistical packages.

1.BACKGROUND

As in most universities, teaching and research at the University of Transkei are considered to be two sides of the same coin. Research is thus regarded as a fundamental and indispensable activity. With the University facing the challenge of becoming competitive within the new structure of higher education in South Africa, a comprehensive research planning exercise was undertaken in 1998 that involved the Faculties of Arts, Economic Sciences, Education, Law, Health Sciences, and Science. The objective was to develop a co-ordinated suite of strategic research plans that would form the basis for research activities in the short and medium term and provide a framework for assessing external funding from a variety of sources. The completion of the Strategic Plan for Humanities and Social Sciences for the period 1999-2001 (Dye, Glencross, Mijere, & Ntusi, 1998) has resulted in a flexible blueprint for research development that will take the University well into the new millennium.

The Research Resource Centre at the University was established in December 1997 with the aid of funding from the then Centre for Science Development, now the Division of Social Sciences and Humanities within the National Research Foundation. This followed the successful establishment of another Centre, the Govan Mbeki Research Resource Centre, at the University of Fort Hare. The Research Resource Centre provides and facilitates regular on-going research training and other related support to academic staff and postgraduate students. Links between these two Centres are maintained by email and telephone, information about seminars is shared, although no between-Centre visits have taken place yet. The National Research Foundation has also attempted to set up a third Research Resource Centre at the University of Zululand, but staffing and infra-structural difficulties have so far prevented this.

2. FUNCTIONS OF THE RESEARCH RESOURCE CENTRE

The functions of the Research Resource Centre must, of necessity, be viewed within the context of the University as a whole. The University of Transkei, as an institution of higher learning in a rural underdeveloped area of South Africa, has its major focus on empowering people to meet the needs of a society in transition. Thus, its policies and programmes are directed primarily towards redressing imbalances of the past while concurrently striving to meet challenges of the present and the future. The University,

> "*perceives itself as part of a family of African institutions of higher learning firmly committed to the scientific, technological, cultural and human advancement of the African continent and to that of the international community*" (University of Transkei, 1999, p.1).

The mission of the Centre is to facilitate research capacity development and research excellence within the University. Research capacity development is first and foremost about developing appropriate research skills at an individual level and promoting a culture of research at the institutional level which together lead ultimately to research excellence nationally. Broadly speaking, the major objectives of the Centre are consistent with those of the University, namely, to enable academic staff and students to acquire research knowledge and skills so that they are able to initiate quality research projects and participate effectively in ongoing research. Although the primary focus is on research in the human and social sciences, the Centre embraces the full range of disciplines within the University. The primary functions of the Centre are as follows:

- To provide assistance with project planning and writing of research proposals;
- To provide a statistical advisory service to facilitate data acquisition, capture and analysis within research projects;
- To provide information to researchers on research and research policy at other institutions and government agencies;
- To organise seminars, workshops and short courses related to all aspects of the research process;
- To provide physical resources in the form of computers and statistical software for the production of reports and other academic outputs, e.g., conference posters, conference papers and refereed journal articles.

In addition to these functions, the Centre strives:

- To promote the use of Information Technology in the human and social sciences and so assist in the creation of a significant mass of networked information that can enrich a sense of community, foster intellectual collaboration, preserve cultural information and ultimately improve the quality of teaching and learning within the University;
- To become a centre of expertise and excellence in quantitative and qualitative data analysis methods and relevant computer packages;
- To provide a link between the South African Data Archives and individual researchers;
- To provide a regular newsletter;
- To produce a number of research reference guides covering all aspects of the research process;

- To provide links with other research institutions and Internet sites through its own Internet web site (http://www.utr.ac.za/rrc/).

3. AN APPROACH TO THE TRAINING OF SOCIAL SCIENCE RESEARCHERS

Our approach to the training of social science researchers has involved a structured view of social research (Mouton, 1996) and has featured the use of workshops and short courses, supplemented by a variety of research seminars. At all times, the focus has been on the complete process of research as a coherent, integrated activity that involves the stages described below (Mouton, 1996).

Formulating a research problem

Two key tasks are involved. First, the 'what' of the research study, that is, the unit of analysis must be specified and second, the 'why' of the study, that is, the research objectives or purposes must be made clear. The unit of analysis, or 'case', requires the researcher to be clear about what kind of social entity is to be studied, what the variables are and what relationships may exist between them. The research objectives may be identified from a combination of the existing background knowledge and the interests, motives and preferences of the individual researcher.

Research design

It is a *sine qua non* that a well-defined research problem is needed for any research investigation. The all-important research design, which is basically a set of guidelines for addressing the research problem, follows logically from the research problem and enables the researcher to anticipate later research decisions and maximise the validity of the final results.

Conceptualisation

Conceptualisation is seen as the process of defining the key concepts in the statement of the problem. Since it is essential for the researcher to relate his/her work to an existing body of theoretical and empirical knowledge, conceptualisation also involves integrating the research study into a larger conceptual framework.

Operationalisation

This consists of providing links between the key concepts in the statement of the problem and the actual phenomena to be studied. Invariably this involves the construction of a measuring instrument such as a questionnaire, test or observational schedule, whose items serve to operationally define the variables in the study.

Sampling

The idea of sampling is familiar to most people. In social research, sampling means some form of random selection of elements from a target population to produce a representative selection of population elements.

Data collection

In social research the fact that human beings are the focus of inquiry and usually react to being studied and investigated, creates unique problems that are not experienced in the physical sciences. This reactivity is affected by both the kind of data source and

the control measures used by the researcher. In addition, errors associated with data collection arise from effects related to the researcher him/herself, context effects and effects originating in the research setting.

Controlling for all of these effects is practically impossible, but has given rise to a number of methods, such as triangulation, anonymity, confidentiality, experimental and control groups, that help to minimise threats to reliability and validity.

Data analysis and interpretation

The analysis of social research data involves firstly, the reduction of the wealth of data collected to manageable proportions, and secondly, the identification of patterns and themes in the data. This often involves a combination of both quantitative and qualitative data analysis.

Writing the research report

Like the research study itself, the research report is a function of a variety of factors: the purpose of the research, the interests of the researcher and the assorted practical constraints of resources. The research report thus represents a reconstruction of the research process and is written in the form of a logical and persuasive argument. There are clear differences between a master's dissertation, a doctoral thesis and a journal article for publication.

The workshop activities have been used: (1) to introduce beginning social science researchers to the theoretical and practical perspectives of research, (2) to develop understanding of quantitative statistical techniques and the associated computer packages used in data analysis; and (3) to develop writing skills and promote the publication of research activities. For example, the following topics have all been addressed in workshops and short courses over the past two years:

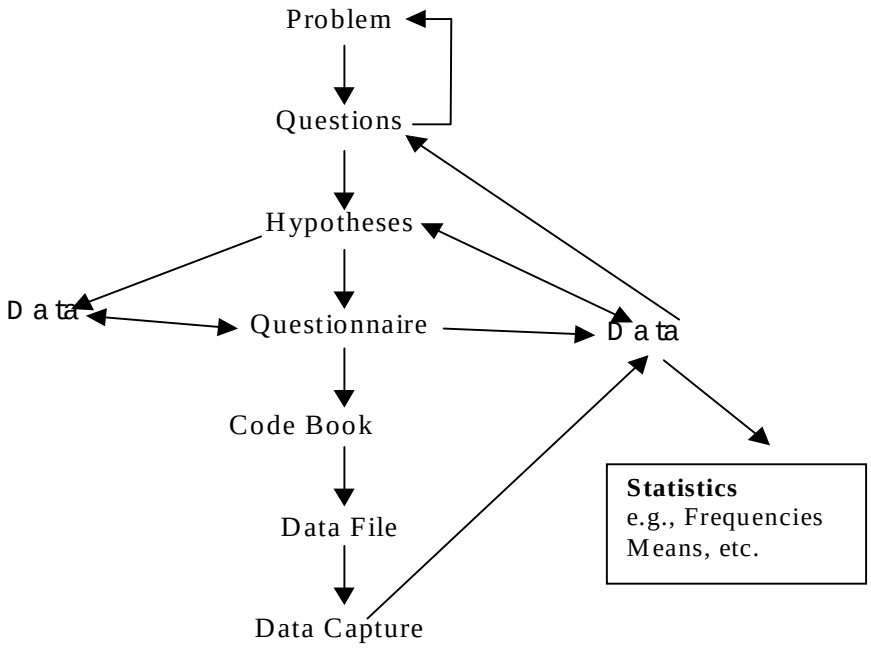
- Basic and advanced project design;
- Design of research instruments;
- Literature searching and data base access;
- Use of data archives;
- Secondary data analysis;
- Programme evaluation;
- Supervising research projects;
- Basic quantitative data analysis (frequencies, tables, means, correlation);
- Advanced quantitative data analysis (ANOVA, principal component analysis);
- Creating a codebook:
- Visualising empirical relationships (scatter plots, simple and multiple regression, correspondence analysis);
- Use of SPSS for data analysis:
- Proposal writing;
- Report writing;
- Writing for publication.

4. TOWARDS THE STATISTICAL EDUCATION OF SOCIAL SCIENCE RESEARCHERS

Collectively, the academic staff and postgraduate students who make use of the Research Resource Centre's facilities have a wide range of backgrounds (from anthropology to zoology and including education, nursing science, psychology and sociology), varied research skills (from beginners to well-experienced) and an uneven level of statistical knowledge (from frequencies and percentages to multivariate analysis). In addressing the statistical education needs of social science researchers, we have found it effective to relate all activities to the context of research and, wherever possible, to a specific research project. This inevitably involves using real rather than 'fake' data, an approach supported by a number of other statistics educators. For example, Hirotsu (2001), argues that it is essential to use actual problems to teach statistics, while Ospina and Ortiz (2001) support the use of real data to solve real problems as in the long run, this is important for statistics education. Similarly, Svensson (2001) encourages the use of the researchers' own research problems as a way of understanding methodological and statistical theories and has adopted an approach that focuses on statistical strategy rather than statistical technique. Such an approach puts statistics in context and serves to motivate researchers. In this way we are able to stress the conceptual understanding of statistical ideas, ensure the appropriate application of a technique and emphasise the correct interpretation of the results.

We have represented the use of statistics as a tool for research by means of a relational diagram of the elements of the research process (Fig. 1).

Figure 1: Elements of the Research Process



In Fig. 1, we see that the research process begins with a problem. This is articulated in the form of research questions that in turn are formulated as hypotheses. The hypotheses give rise to a questionnaire and subsequently the related codebook. From the codebook we are able to prepare an appropriate data file ready to receive the data. After administration of the questionnaire, the process of data capture takes place. The process

of data analysis may now be carried out. Both the hypotheses and the questionnaire have a direct influence on the actual data collected, while the hypotheses and the questionnaire influence the data analysis. Data analysis, which requires the use of statistical procedures, may now proceed and be used eventually to provide answers to the research questions posed earlier. Although this may be a simplified view of the research process, it serves to provide a perspective for the important step of data analysis and the use of statistical techniques

However, as all researchers are aware, the process of research is not linear as Fig. 1 appears to imply. Although the outcomes of data analysis provide answers, be they partial or complete, to the research questions, they invariably raise more questions and the whole process loops back to the problem and sets off another cycle of research. This cyclic approach to research has been used by a number of researchers and is clearly articulated by Bishop and Talbot (2001). They call the elements of the cycle Problem, Plan, Do, Study, Act (PPDSA). The five stages of an investigation are described as follows:

- *Problem:* Obtain a clear conception of learning goals;
- *Plan:* Plan the procedures to carry out a study;
- *Do:* Collect the data according to the plan;
- *Study*: Analyse and interpret the data;
- *Act:* Modify one's view of the problem in the light of the data.

The PPDSA cycle is a useful model for guiding novice researchers through the research process. The conventional forms of statistical analysis that we stress are summarised in Table 1 and Table 2. Descriptive statistics are used to describe the basic features of the research data. They provide simple summaries about the sample and the measures. Together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data:

> *With descriptive statistics you are simply describing what is, what the data shows"* (Trochim, 1999, p. 250). Inferential statistics *"investigate questions, models and hypotheses. In many cases, the conclusions … extend beyond the immediate data … Thus, we use inferential statistics to make inferences from our data to more general conditions …*" (Trochim, 1999, p. 250).

Table 1: Major Forms of Statistical Analysis

Descriptive Statistical Analysis	Inferential Statistical Analysis
This is used to enable the researcher to organise and summarise the data to render the results more comprehensible.	This is used to enable the researcher to establish whether the observed results represent true population values. It is used: 1. To estimate population parameters 2. For hypothesis testing, e.g., chi-square, t-tests, ANOVA.

In many instances, a major part of social science research is qualitative in nature and may for, example, involve interviews with volunteer respondents. The approaches of quantitative (statistical) and qualitative analysts are thus quite different. A quantitative researcher carries out the data analysis by examining individual elements,

> "*first in isolation (univariate statistics) and then in various combinations with other elements (bivariate and multivariate statistics)*" (Mouton, 1996, p. 169).

Table 2: Structure of Descriptive Statistical Analysis

Univariate Analysis	Bivariate Analysis	Multivariate Analysis
Used to identify properties of single variables	Used to identify relationships between two variables	Used to identify relationships among several variables
Nominal/ordinal data, e.g., race, gender, clas	Nominal/ordinal data, e.g., race, gender, class	Nominal/ordinal data, e.g., race, gender, class
Examples: percentages, mode, median, range	Example: Spearman's rank correlation	Example: correspondence analysis
Interval data, e.g., age, income, test score	Interval data, e.g., age, income, test score	Interval data, e.g., age, income, test score
Examples: mean, standard deviation.	Examples: Pearson product-moment correlation, regression.	Examples: principle component analysis, multiple regression.

It is important for the researcher to determine whether or not the results obtained from the sample data may be generalised to the population. This leads naturally to the use of inferential statistics to estimate population parameters or test hypotheses. In qualitative research, however,

> "*the investigator usually works with a wealth of rich descriptive data, collected through methods such as participant observation, in-depth interviewing and document analysis. The research strategy is usually of a contextual nature … (and makes) use of methods of data analysis that are more holistic, synthetic and interpretative*" (Mouton, 1996, p. 169).

For researchers who make use of questionnaires for collecting data, the twin issues of reliability and validity are crucial. Reliability deals with the accuracy and consistency of measurement and asks, 'Will the same methods used by different researchers and/or at different times produce the same, or similar, results?' The different forms of reliability and the appropriate measures, e.g., test-retest (correlation), equivalent form (correlation) split halves (correlation), internal consistency (Cronbach alpha), are discussed in terms of the contest in which they arise. Validity, which addresses the soundness or effectiveness of a measuring instrument, asks, 'Does the instrument measure what it is supposed to measure?' There are several forms of validity, e.g., face, criterion, content and construct, and all are discussed as and when needed.

5. WORKSHOPS

To clarify our approach, we now give examples from some of our workshops. The first was primarily intended to introduce participants to SPSS, but also served to help them develop a conceptual understanding of basic statistical ideas, learn how to use statistical techniques correctly and to interpret the resulting output sensibly. The others were aimed at introducing researchers to more sophisticated multivariate techniques. In each workshop, we adopted an experiential, hands-on approach, with all participants using a PC with SPSS and a suitable data set. Instruction was provided through the medium of a PC and data projector, supported where necessary by transparencies on an overhead projector.

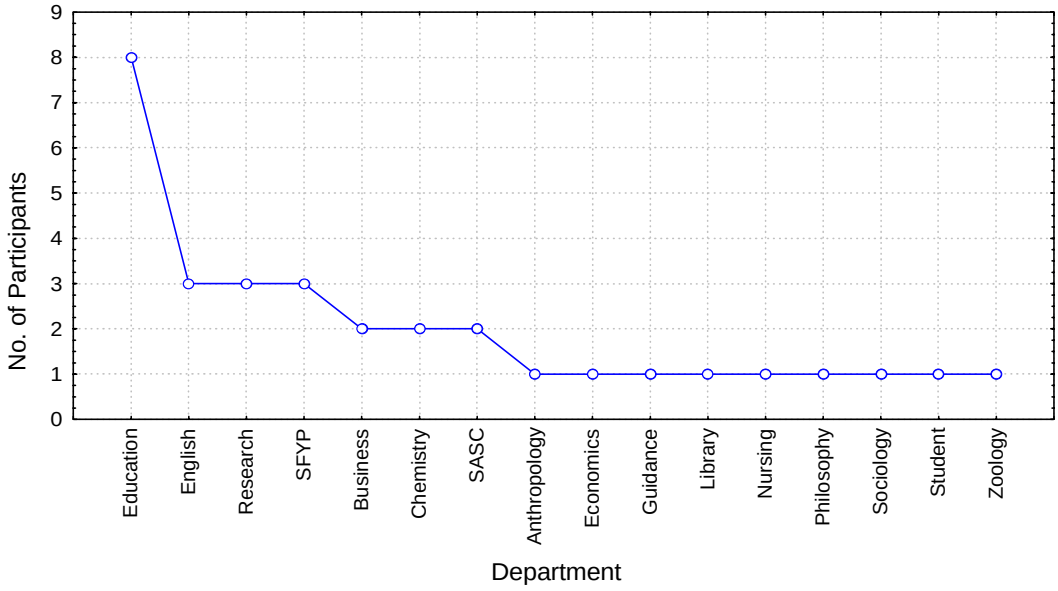
In the first workshop, after the introduction a general overview of SPSS was provided. Participants were helped to access SPSS and navigate their way through

several layers of windows. They were then introduced to the idea and need for creating a codebook and the related data file. This was done by providing participants with an extract from a survey questionnaire and copies of five completed questionnaires. Using an interactive question-and-answer session, participants worked in pairs to create their own copies of the codebook. Participants were then introduced to the procedure for preparing a data file and, using the newly created codebooks, proceeded to enter data from the questionnaires. After data capture the need for data checking and cleaning was dealt with, leading naturally into data analysis. For this purpose, participants were provided with copies of the full research data set, which was then used, for the introduction and use of frequencies, cross-tabulations, tables, graphs, descriptive statistics and correlation. The comparison of means using t-tests and one-way analysis of variance was also explained.

As much as possible, we avoided simply telling participants what to do, relying instead on asking questions about the data and obtaining ideas from them about what could or should be done. In this way, participants were lead naturally to suggest appropriate routes for the analysis and we were able to introduce the relevant statistical ideas within the context of the specific data set.

On other occasions, workshop participants were provided with access to larger data sets that were used as vehicles for driving the analysis and developing understanding of principal component analysis, correspondence analysis, multiple regression and CHAID. Briefly, for those unfamiliar with it, CHAID (Chi-squared Automatic Interaction Detector) is a sophisticated segmentation modelling method for analysing large quantities of categorical data (Kass, 1980). In the workshop on regression, for instance, participants were introduced to the concept of visualising relationships using scatter plots. After sufficient exploration of a number of relationships, simple and multiple regression ideas were developed, with participants being encouraged to provide suggestions for the analysis and challenged to give meaningful interpretations to the results. The workshops have not attracted large numbers of participants, but have drawn staff and students from a wide range of subject areas. This is shown in Figures 2, 3 and 4 in which details of attendance at three workshops are given.

Figure 2: Summer School (1998) - Writing for Publication (N=32)



As with any educational task, we have experienced problems in presenting workshops. An unexpected one for us has been the lack of computer literacy among staff and students alike. This has included unfamiliarity with the use of a mouse and a general lack of understanding of a windows environment. Fortunately, workshop participants have responded positively to most situations and progress has been good once the initial handicaps have been overcome. A more serious difficulty is the general lack of familiarity with and formal knowledge of basic statistical ideas. The background levels range from almost nil to one or two people with experience of multivariate techniques. We have found that introducing statistical ideas only within the context of a particular research project and its data set has been beneficial and less threatening to workshop participants.

Figure 3: Workshop 5 (1999) - Creating a Codebook (N=17)

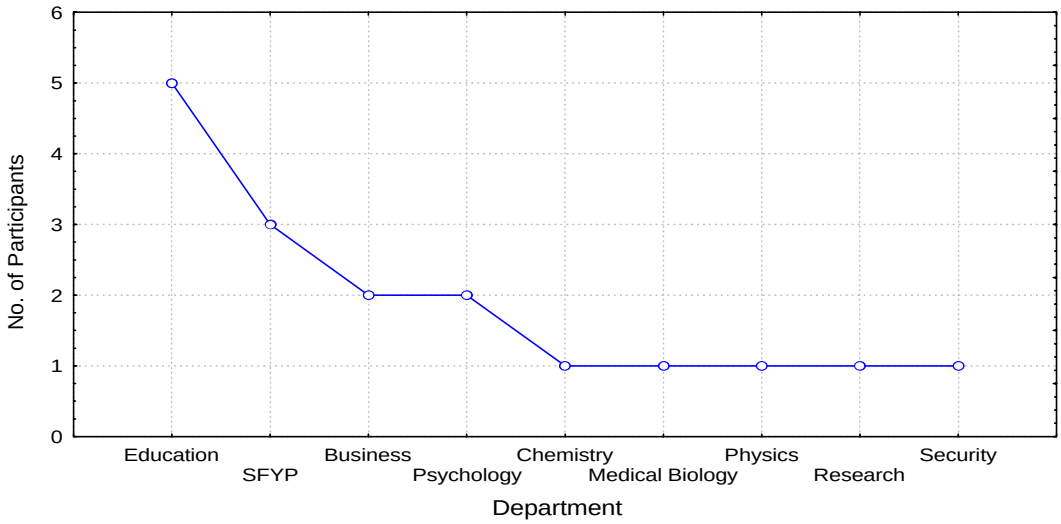
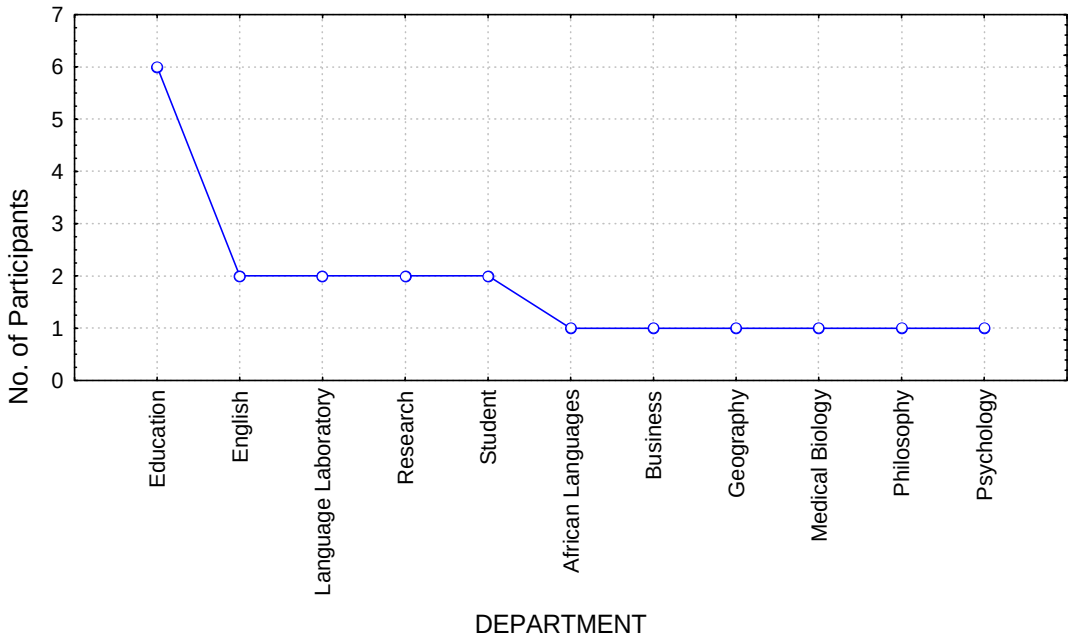


Figure 4: Workshop 5 (2000) - Research Instrument Design (N=20)



6. QUALITY ASSURANCE

South Africa today is facing a multitude of social and educational changes. With new academic policies being promoted and the recent developments in terms of the Minister of Education's "Size and Shape" initiative, universities are rapidly adjusting to the need to operate efficiently, productively and economically (Kotecha & Luckett, 2000). It is widely acknowledged that education offers

> *"the high road to economic productivity, some measure of social equality and democracy in the modern world"* (Singh, 2000, p. 5).

In order to achieve a competitive edge in the global economic and educational scenarios of the 21st century, quality checks are essential. Quality checks are also necessary to ensure the effective implementation of curricula and the concomitant assessments, not only to ensure competitiveness, but also to ensure the professional and personal development of academic staff in higher education (Mammen, 1999). There is a clear need

> *"to devise new ways of evaluating our educational programmes and accounting for what we do"* (Singh, 2000, p. 6).

The successful transformation of higher education requires, as a non-negotiable matter, the development of some form of quality assurance. The establishment of a considerable number of quality assurance units in recent years is a reflection of this concern (Mammen, 2000). Within the boundaries of South Africa, the South African Qualifications Authority (SAQA) Act (SAQA Act, 1995), the report of the National Commission on Higher Education (NCHE) (NCHE, 1996), the South African Higher Education Act of 1997 and the subsequent establishment of the national Higher Education Quality Committee (HEQC) have collectively laid the foundation for quality assurance in the post-apartheid South African higher education (Mammen, 2000). As noted by Singh (2000):

> *"This relatively new national and international phenomenon has mushroomed into a fast growing industry … (although) the complexities surrounding the development of quality assurance … cannot be overestimated."* (Singh, 2000, p.7).

The necessary new structures and their corresponding new demands are already having an impact on busy academics and administrators, with some regarding quality assurance as

> *"a threat to their academic freedom and autonomy"* (Singh, 2000, p.7).

It is thus both understandable and not surprising that attempts to implement quality assurance practices have met with strong resistance in many educational circles.

Singh (2000) makes the pertinent observation that quality assurance asks fundamentally important questions of all of us, questions that should be part of our normal activities. For example:

> *"What am I trying to do or achieve? Why am I doing it in that way? … What is the context in which I am doing it? How do I know that it is effective and that I am doing a good job? Is*

this the best possible way of doing it? Was it worth it, after all? And so on" (Singh, 2000, p. 7).

Quality assurance involves, *inter alia*, internal self-evaluation. We have done this by keeping detailed records of the daily use of the Centre and attendance at workshops and seminars. Quality assurance also involves some form of independent external evaluation. The Centre was evaluated by the National Research Foundation in November 1999. The report was generally positive and enthusiastic about the work being done and the effect of the Centre on the research community in the university, although weaknesses were identified and suggestions for improvement were made.

7. PROPOSED DEVELOPMENTS

In order to consolidate the position of the Centre within the University and to enable a controlled expansion of the services provided by the Centre, a three-year Strategic Plan has been drawn up and a proposal for additional funding submitted to the National Research Foundation. The Strategic Plan makes provision for the following developments:

- The establishment of two Junior Research Fellowships. They are intended for registered master's students who would provide additional support in the day-to-day running of the Centre. Junior Research Fellows will be required to work a specified number of hours per week in the Centre. This will involve assisting with a variety of research projects and being involved with the expanded consultancy services to be offered by the Centre in 2001.
- The establishment of two Research Fellowships for registered doctoral students. In addition to additional support in the day-to-day running of the Centre, Research Fellows will assist with the development of individual, departmental and university research projects. They will also be involved with the expanded consultancy services to be offered by the Centre in 2001.
- The establishment of one Senior Research Fellowship at post-doctoral level. This will be used to enable an experienced researcher from outside the University to spend three to six months here and contribute to the growth and development of research expertise and excellence within the University.
- The provision of additional research resource materials. These include additional computers as well as books and research-focused computer software. For example, there is a growing need for at least one computer package for qualitative data analysis, while a variety of specialist packages for quantitative data analysis would enhance the versatility of the Centre.

8. SUMMARY

In this paper, we have provided a view of the role of the Research Resource Centre at the University of Transkei in the training of social science researchers. We have attempted to provide a contextual perspective of a Centre that is vital for the facilitation and development of research capacity among researchers at a university in a rural underdeveloped area of South Africa. We have highlighted the objectives of the Centre

and described its functions, which are not restricted to the social sciences. However, we have focussed on the researchers in the social sciences and described an approach to the research and statistical education of this group. Our approach is one that encourages and promotes experiential, hands-on, active participation in all aspects of the research process. We especially espouse a conceptual approach to the learning of statistics (see for example, Weldon, 1986). Although we make use of standard statistical packages, e.g., SPSS and Statistica, not every possible research situation is accounted for in such packages. In these instances, we have found a valuable resource to be that of Silver and Hittner (1998), which is effectively a compendium of statistical software programs for the more adventurous researchers.

The role of the Research Resource Centre is essentially that of a facilitator. We have responded to people's needs as and when they arose. More importantly, we have endeavoured to anticipate such needs by providing a range of workshops, short courses and research seminars that collectively promote a positive environment for researchers. How successful we have been is reflected in a recent external evaluation of the Centre. The Centre,

> *"leans towards quantitative approaches to research and has a strong emphasis on developing quantitative research methods and skills. (It) has been the most effective RRC in disseminating research information, promoting and informing researchers about (research) programmes and grants, and providing on-site services and support"* (Simons & Subotzky, 2000, p. 19).

We were, nevertheless, criticised for an over-emphasis on quantitative research methods, something we are attempting to deal with.

No researcher who deals with social issues of any kind, whether at a theoretical or practical level, can do so effectively without reference to empirical information - the facts or the data. As is generally known,

> *"statistics generally has a bad reputation among social science students"* (Bless & Kathuria, 1998, p. v).

The reasons are not entirely clear. Statistical reasoning is not particularly difficult, while the ubiquitous computer and the accompanying statistical packages have made the tedium of lengthy calculations a thing of the past. It is our considered opinion that a Research Resource Centre such as ours has a pivotal role to play in facilitating research capacity development and research excellence within the University generally and particularly in the research training and statistical education of social science researchers.

REFERENCES

Bishop, G., & Talbot, R. (2001). Statistical thinking for novice researchers in the biological sciences. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 215-226). Granada: International Association for Statistical Education and International Statistical Institute.

Bless, C., & Kathuria, R. (1993). *Fundamentals of social statistics: An African perspective.* Cape Town: Juta.

Dye, A. H., Glencross, M., Mijere, N., & Ntusi, T. (1998). *Strategic research plan for humanities and social sciences at the University of Transkei 1999-2001*. Umtata: University of Transkei.

Hirotsu, H. (2001). Statistical training of researchers in total quality management: The Japanese experience. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 53-63). Granada: International Association for Statistical Education and International Statistical Institute.

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29(2), 119-127.

Kotecha, P., & Luckett, K (2000). A way forward: emerging perspectives on a national quality assurance system for South Africa. *South African Journal of Higher Education*, 14(2), 204-210.

Mammen, K. J (1999). Towards an academic transactions and assessment framework for university science educators and learners at the undergraduate level: a case study. Unpublished DEd proposal. University of Transkei, Umtata.

Mammen, K. J. (2000). Development of instruments employed to assess the quality of undergraduate student assessment within the context of TQM at a South African university. *Paper presented at the 12th International Conference on Assessing Quality in Higher Education, RMIT University, Melbourne, Australia, 28-30 June, 2000.*

Mouton, J. (1996). *Understanding social research.* Pretoria: J. L. van Schaik.

NCHE (1996). National Commission on Higher Education report: A framework for transformation. Pretoria: HSRC.

Ospina, D., & Ortiz, J. E. (2001). Statistical research and consulting in universities from developing countries: the Colombian case. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 295-310). Granada: International Association for Statistical Education and International Statistical Institute.

SAQA Act (1995). *South African Qualifications Authority (SAQA) Act, 1995.*

Silver, N. C., & Hittner, J. B. (1998). *A guidebook of statistical software for the social and behavioral sciences.* Boston: Allyn & Bacon.

Simons, R. & Subotzky, G. (2000). *Consolidated report on the evaluation of the research resource centres at the Universities of Zululand, Transkei and Fort Hare*. Cape Town: Education Policy Unit, University of the Western Cape.

Singh, P. (2000.) Quality assurance in higher education in South Africa. *South African Journal of Higher Education*, 14(2), 5-7.

South African Higher Education Act (1997).

Svensson, E. (2001). Important considerations for optimal communication between statisticians and medical researchers in consulting, teaching and collaborative research - with focus on the analysis of ordered categorical data. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 21-35). Granada: International Association for Statistical Education and International Statistical Institute.

Trochim, W. M. K. (1999). *The research methods knowledge base*, second edition. Ithaca, NY: Cornell University Custom Publishing.

University of Transkei (1999). *Strategic plan*. Umtata: University of Transkei.

Weldon, K. L. (1986). *Statistics: A conceptual approach*. Englewood Cliffs, NJ: Prentice-Hall.

Michael Glencross and Andile Mji
Research Resource Centre University of Transkei
Private Bag X1, Unitra, Umtata, South Africa 5117
E-mail: glencros @ getafix.utr.ac.za, mji @ getafix.utr.ac.za

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 259-262.

Printed in Granada, Spain.

BRIAN PHILLIPS

DISCUSSION

1. TRAINING REGULAR EDUCATION AND SPECIAL EDUCATION TEACHERS IN THE USE OF RESEARCH METHODOLOGY AND STATISTICS

Blumberg's paper gives a most comprehensive overview of the research skills researchers in any social science, and in particular education, should be familiar with before carrying out their own research. The author is to be congratulated for such an informative and well-researched paper.

In discussing the research methodology and statistics needs of primary and secondary level teachers seeking Master's for Special Education teachers, Blumberg outlined the goals and organisation, topics used, and ended with a discussion of some specific research tools of which she felt special education teachers should be aware.

Firstly I like the idea of classifying classroom teachers also as researchers, though I would not restrict this to teachers. I think it would also apply to people in an increasing number of positions such as doctors, police, social workers who all are continually collecting and analysing data on people's health, behaviour and so on. However I did wonder whether, as the author claims “they perform well-designed experiments in their classrooms.” How could I not agree with the major goals, that is to:

1. Help them critically read, correctly interpret, and decide the validity of conclusions in the published and unpublished literature;
2. Give them the tools necessary to complete research-based projects and/or theses required for their Master's degrees; and
3. For Special Education teachers, they should learn how to better collect and analyse data that will be used to help determine each student's IEP.

These admirable aims are fine, if not somewhat optimistic, for a 25-hour course in my view. As far as the author's aims go, some questions need to be answered first. For example: What base of statistics knowledge base were the students coming from? From the skills Blumberg suggested, as given by Todd and Reece (1990), such as "Can formulate a testable hypothesis" and "Has an understanding of the standard deviation" it is presumed that they had little or no previous statistical/research methods training, which makes programs such as the author's even more ambitious in such a time period.

Her comment that some 90% of Master's programs require at least one research or statistics courses is very encouraging. However, what they learn in these courses maybe subject to question. Is the program appropriate to these student needs and abilities? Should many have a general statistics literacy program rather than the typical Introductory Statistics course?

The author claims that an advantage of having research methodology and statistics combined into one course is that it is more efficient in delivery. This maybe true, but limits students to the one instructor, students learn different things from different instructors so I tend to favour having different instructors in different courses, even

within the same course (team teaching, visiting lecturers etc). However it is important, as the author points out, that the "faculty members teaching the different statistics courses (to the same students) communicate well with each other to minimise needless overlap".

I agree that descriptive statistics is extremely important (maybe 34% as reported) - I think it should be even higher, as too often more "high level" statistics are reported before the data is seriously looked at. However I am not in full agreement about the author's suggestion that "not much time should be spent training them in the use of a specific computer package" because they are always changing, Once one package is understood well, it can be built into the program, rather than be seen as something separate, it becomes a tool. Furthermore, the use of laptops makes learning with computers in the classroom much more friendly. Certainly understanding the output from statistics packages is most important.

The author's suggestions on the need for these teachers to understand sufficient about hypothesis testing to enable them to critically read and correctly interpret research-based articles and the difference between statistical significance and practical significance is very important. Also confidence intervals must be understood and written in language that third parties such as administrators and parents can appreciate.

A question in the author's mind was "how much these teachers need to know about specific hypothesis testing procedures?" For all of these procedures they should not have to compute anything by hand/calculator or by computer, but rather be able to know what to use and how to interpret the results.

Most of the projects and theses that Master's degree students complete have fairly simple designs. Hence I agree that time should not be spent teaching them a catalogue of complicated designs. but it is especially important that teachers doing research be urged to consult a statistician before they collect their data.

They should appreciate the distinction between correlational and causal studies so appropriate statements can be made. The use of surveys is very popular in many areas and principles of questionnaire design and sampling should be taught in any course on statistics especially in the social sciences. In fact I would say that good teaching in survey research methods may have avoided the huge problems which occurred in Florida during the 2000 Presidential elections!

I agree with the author's thoughts that researchers need to understand the standardised test procedures including reliability and validity so they can be better prepared to prepare their instruments before collecting their data. Also it is good to see that the author does not limit herself to traditional study methods which generally involve groups of subjects. Hence some introduction to other such as example meta-analysis and single-subject and related designs are seen as appropriate due to the fact that these researchers often only have access to small sample sizes.

Also an important component that is often assumed, but probably rarely taught, is Statistical Thinking. Blumberg proposes this needs to be incorporated in general statistics teaching, which I approve.

Overall this was a most useful and thought provoking paper and I recommend it to anyone teaching statistics to education researchers.

2. THE ROLE OF A RESEARCH RESOURCE CENTRE IN THE TRAINING OF SOCIAL SCIENCE RESEARCHERS

This paper showed how teaching and research are fundamental and indispensable

activities. The authors gave an excellent example of a Research Resource Centre, which provides and facilitates regular on-going research training and other related support to academic staff and postgraduate students.

I agree with their concept of a Research Resource Centre, which is viewed within the context of the University as a whole. As statisticians we need to see statistics and research methods not as our sole territory, rather we are probably in the best position to control joint schemes for research methods training across our institution. An advantage of such a centre is that researchers from many disciplines can make use of the same resources and expertise and share ideas and knowledge across disciplines.

The need for the type of workshop activities outlined is crucial for any researcher. As the authors state these include: formulating a research problem, research design, conceptualisation, operationalisation, sampling, data collection, data analysis, interpretation and writing the research report.

I am comfortable with students getting some in depth knowledge with a statistics package, even though packages change as time passes. As mentioned earlier a package can be integrated as a tool, rather than be seen as something separate, though in Africa it cannot be assumed that researchers have sufficient computer literacy to get straight into a package. In such cases some basic computer skills need to be incorporated into the program.

The knowledge of secondary data sources and how to access data is increasingly important enabling researchers to readily make use of the raw data from other studies along with their own research. I would not be so confident as the authors that researchers really understand issues of sampling, they may be familiar with the ideas, but do they really use them?

As in Blumberg's work, they see issues of reliability and validity as important as many of the researchers make use of questionnaires. The tactic of asking questions rather than simply telling people what to do is an excellent approach for researchers who may not like being treated as an undergraduate and may think they know the basic ideas. I agree with the authors' claim that "the best way to introduce statistical ideas is within the context of a particular research project and its data set has been beneficial and less threatening to workshop participants." Another advantage of a set-up like a Research Resource Centre is being able to try things on the spot.

I also agree with their opinion that such a Research Resource Centre has a pivotal role to play in facilitating research capacity development and research excellence within the University generally and particularly in the research training and statistical education of social science researchers.

These activities are to be applauded and the involvement of a young statistician at an international conference is an example others are encouraged to follow.

3. OVERVIEW

Summing up about both papers I strongly agree with the Blumberg's comments that "researchers should be strongly urged to consult with a statistician when they design their studies, before they begin data collection, as changes are being made during data collection, and before beginning their data analysis."

An important part of any training for researchers is for them to learn to know when they should ask questions and what questions to ask.

To give some recent examples from PhD students who approached me with problems, which although not uncommon, still surprises and depresses me:

> Student A: *"I want to find out what are the characteristics of successful business women. I have sent out questionnaires with about 300 questions. It has taken me about 3 years to get to this stage and I want to know what to do with the answers."*

I asked how many responses she had, and the answer was 14. But she said they were good respondents!

> Student B: *"Another even more extreme case, he had not yet collected the data but had about 3000 questions."*

I asked how long it would take for someone to answer the questions and he said about 6 hours, if they were quick!

In neither case were they prepared to change their questionnaires as their supervisors had been approved them, and they did not have time - they just wanted to know what to do with the data.

The statistical methods in both papers cover a lot of territory. What is appropriate for one group is not important or even of interest for another. For example we can often get hung up on hypothesis tests, but if they are mainly concerned with estimating rather than comparing, confidence intervals may suffice.

There were a number of other topics mentioned in each paper which could be included in courses for research students, trying to include too many may well overwhelm students. A variation would be to have the program, in which some topics are given a only a very brief introduction and which directed different students to specific courses such as short course modules relevant to their specific research.

Such modules could be available to researchers across the institution or even from neighbouring institutions to help share resources. My main point is that it is not necessary for a particular department to try to do everything themselves, rather make use of the diverse and rich resources which usually already exist within their institution structure, or which in some case can be easily found outside.

Both papers are give very useful ideas for people training researchers who use statistics in the social sciences and the authors are to be praised for their efforts.

REFERENCES

Blumberg, C. J. (2001).Training regular education and special education teachers in the use of research methodology and statistics. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 231-244). Granada: International Association for Statistical Education and International Statistical Institute.

Glencross, M., & Mji, A. (2001). The Role of a research resource centre in the training of social science researchers. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 245-257). Granada: International Association for Statistical Education and International Statistical Institute.

Todd, R. E., & Reece, C. C. (1990). *Graduate introductory educational research courses: A look at what is required.* ERIC Document Reproduction Service No. ED 324 274.

Brian Phillips
Swinburne University of Technology
School of Mathematical Sciences,
P.O. Box 218, Hawthorn 3122, Australia
E-mail: bphillips@groupwise.swin.edu.au

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 265-275.

Printed in Granada, Spain.

SHRIKANT I. BANGDIWALA AND SERGIO R. MUÑOZ

TRAINING OF STATISTICIANS AND CLINICAL RESEARCHERS WORLDWIDE TO COLLABORATE AS CO-INVESTIGATORS WITHIN COUNTRY CLINICAL EPIDEMIOLOGY UNITS: THE EXPERIENCE OF THE INTERNATIONAL CLINICAL EPIDEMIOLOGY NETWORK (INCLEN)

Clinical researchers rely on biostatisticians in order to design, conduct and analyse observational and experimental studies involving populations of subjects. In many countries, trained biostatisticians are not readily available. There are many possible approaches to this problem, including educating the medical or health professional to be a researcher with an understanding of statistical methodologies, as well as training statisticians to be biostatisticians with an understanding of clinical considerations. The International Clinical Epidemiology Network (INCLEN) embarked on such an endeavour by creating clinical epidemiology units that included both approaches, trained clinical epidemiologists as well as biostatisticians. The specific statistical training needs of both types of students are described.

1. INTRODUCTION

Physicians and other health professionals are increasingly aware of their need for biostatistical knowledge, not only if directly involved in research activities, but also if, as a clinical practitioner, one wishes to keep abreast of advances in the field.

One alternative for the clinical researcher is to completely rely on a trained biostatistician and to blindly accept the answers obtained from such collaboration. For the clinicians reading the latest scientific journals in their field, this is equivalent to ignoring the methods section of the research articles. This is evidently not a healthy approach, for the obvious reason that an uneducated researcher is not in a position to critically appraise the literature or to effectively collaborate in a research team.

In addition, this approach is undesirable because the biostatistician is placed on a pedestal, viewed as a necessary evil that must provide significant results, and there is no semblance of a collegial collaboration. The end result of this alternative is often poorly planned, poorly conducted, poorly analysed, or poorly presented medical research, which is as unethical as recommending inadequate treatment for their patients.

Evidence of such inadequate collaboration can be seen in reviews of the often poor quality of the methodology and statistics used in peer-reviewed medical journals (Altman, 1994; Coste, Fermanian, & Venot, 1995). Training of the health professional in proper understanding of statistical concepts is not easy. For example, Estepa and Sánchez-Cobo (2001) point out the difficulty in properly teaching the concept of ‘association,’ an important concept in the health professions. In addition, given the increasing complexity and new methods emerging in the statistical field, there is an

increasing clear need for statistical support in medical research at an early stage (Hand, 1994; Altman, 1998).

It is in the best interest of the health research professions and the biostatistical profession that medical professionals planning to be clinical researchers obtain relevant, targeted statistical education. This is not to argue that the biostatistically-trained health professional should be able to be their own biostatistician, but that if they have a basic understanding of the statistical issues present in their research, they will be able to better collaborate with a biostatistician and to therefore be better researchers (Bangdiwala, 1989). The central role of statistics in epidemiology and public health requires that all health professionals acquire competence in the use and interpretation of statistics.

The availability of a trained biostatistician is often not the case in many developing countries (Crivisqui & Abruzzini, 2001). There usually are many trained statisticians, especially in the areas of econometrics and official statistics, or more commonly, well-trained mathematical statisticians with little or no understanding of epidemiological research design, methodologies, or substantive issues of the fields of application (Ospina & Ortiz, 2001).

One alternative for these statisticians, if called to interact with medical researchers, is to completely rely on the physician and to blindly accept the answers obtained from such collaboration. Similarly as the situation described above, this is not a healthy approach. The statistician must also be trained and appreciate the nuances of the field of application if they are to collaborate efficiently. It is thus necessary that the statistician become a biostatistician.

Individuals trained in clinical epidemiology research or in biostatistics are often uncommon professionals in many countries. This can have advantages from a scarce commodity standpoint, but more often they face special challenges. In some countries their special skills may not be recognised as relevant to the country's health priorities. Clinical researchers face the challenge of the pressing needs for their clinical expertise, and often their research skills go under-utilised as they are pressured to treat patients by their institutions or government.

On the other hand, trained biostatisticians are sought by research institutions, international agencies and private enterprise, and may be difficult to retain in academic research medical schools, especially since a non-physician in a medical school may not be considered as an equivalent colleague by the medical doctors. In addition, they often face 'professional isolation' as they have few, if any, colleagues to discuss issues with, and may not be able to maintain themselves current with the advances in their field given the relative lack of resources such as journals and the competing demands for their time. The challenge is thus not only to train these individuals, but to provide them with a career path appropriate to their training, and continued nurturing during their early careers.

This paper presents the experience of the International Clinical Epidemiology Network (INCLEN) training program in training both types of professionals, physicians to be clinical epidemiologists, and statisticians to be biostatisticians. Section 2 describes the biostatistical training for physicians, while section 3 describes the biostatistical training for statisticians. The underlying assumption is that both types of scientists would be collaborating in medical research as part of a clinical epidemiology unit located in a country with not much additional expertise available. Finally, the success of this training is subjectively assessed in section 4.

2. INTERNATIONAL CLINICAL EPIDEMIOLOGY NETWORK TRAINING OF PHYSICIANS

The International Clinical Epidemiology Network (INCLEN) training program began in 1984, and was originally funded by the Rockefeller Foundation as a world-wide program that competitively selected physicians from developing countries for a one-year post-graduate training program in clinical research methods (Halstead et al 1991). The selected physicians were faculty from 27 specific medical schools that were participating in the network. These medical schools were initially selected because of their willingness to establish a Clinical Epidemiology Unit (CEU) as an independent organisation within the medical school, responsive only to the Dean of the School of Medicine.

The idea was that trained clinical researchers, after returning home from training, would have a 'second home' in which to pursue careers in clinical research, in addition to, but separate from their clinical discipline. There was protected time for research - 20% or one day a week – so that their skills in research would be used and the clinical demands would not overcome their entire time. The CEU was comprised of various clinical researchers from a variety of medical disciplines, plus a biostatistician, a health economist, and a health social scientist. The idea was to have a multidisciplinary team of researchers to collaborate effectively. Since its inception, the INCLEN program has trained over 500 health professionals world-wide.

The goals of the INCLEN program were to develop units of excellence in clinical epidemiology research at the participating medical schools in the developing countries. The ultimate hope is that with qualified researchers in a country, the country's pressing health priorities would be adequately researched. As such, the program not only provides training to the future investigators, but also time protection for conducting research activities and the necessary biostatistical and other support required upon their return to their home institutions.

For their training, physicians attended centres at research universities in Canada, Australia, or the United States, for 12-18 months of training. Most physicians did not have any statistical training prior to their participation in the program. Currently, INCLEN has regionalised the training by creating Clinical Epidemiology Research and Training Centres (CERTC) in those CEUs that have progressed in their infrastructure, in such diverse countries as Brazil, Chile, Colombia, Thailand, Philippines, and India. The training curriculum is basically the same as at the original training centres, only with a different venue and faculty. However, at the regional training centres, it was decided to provide a greater emphasis on statistical training for the health professional, since it was less likely that they would be able to obtain the necessary biostatistical support at their home institutions.

Specific aspects of the statistical training of the clinical researchers at the University of North Carolina CERTC are presented in Table 1. The curriculum varied somewhat among the original four training centres and the current regional training centres, but the core elements are quite comparable. The statistical training, aside from the topics of epidemiological research methods, aimed at providing basic training in statistical concepts and methods.

The physicians took the Supplemental Course concurrently with Course I at the beginning of their training, since it was considered fundamental that they have basic statistical software knowledge in order to perform the work required for the biostatistics course. The biostatistics course was offered also at the beginning since it was felt

necessary in order to perform well in the core courses of clinical epidemiology research methods as well as in other courses they took during the academic year.

Table 1: Curriculum Topics of the Statistical Training of Physicians

Topic	Details
Supplemental Course: Introduction to STATA	4 lectures
Creation of data files and data entry, create new variables, subset data files	
Produce graphs	
Perform simple statistical descriptive analyses	
Draw random samples	
Practice	Use software for Course I homework
Course I: Introduction to Biostatistics	12 lectures
Introduction to statistics	Dealing with uncertainty and variability
Types of data	
Elementary probability theory	
Samples and populations	Random sampling
Probability distribution functions	Binomial, Poisson, Gaussian
Descriptive and exploratory data analysis	Graphs, tables, summary statistics
Concepts of statistical inference	Hypothesis testing, confidence intervals, Central Limit Theorem
Methods for one sample	Parametric and non parametric
Methods for two samples	Parametric and non parametric
Analysis for contingency tables	Chi-squared tests
Assessing independence and correlation	Simple linear regression, correlation, diagnostics
Introduction to multiple regression and ANOVA	Issues of multiple comparisons
Course II: Regression models	10 lectures
Exploring relationships	
Logistic regression	
Survival analysis	
Course III: Various topics	
Issues in biostatistical critical appraisal of the medical literature	Articles selected by the clinicians
Study design	Discussion of alternatives and of implications of chosen design
Study conduct, quality assurance	Ensuring adequate data quality
Appropriateness of analyses methodologies	Discussion of alternatives and of possible other results; interpretation of statistical importance of findings
Practices: Relevant project in clinical field of student	Self-chosen project (2-3 weeks) with oral presentation

The biostatistics training took place intensively over a 4-week period prior to the academic semester, but it consisted of the complete topics of the regular 3-credit introductory course offered by the Department of Biostatistics (BIOS 150).

In the first and second semester of the academic year, the physicians took a Core Curriculum on clinical and epidemiological research methods (see Table 2 below). The statistical concepts were already covered in Course I and thus they were able to apply them to the various commonly used epidemiology study designs.

Table 2: Curriculum Topics for Clinical Epidemiology Training of Physicians and Statisticians. Course in Clinical Epidemiology

Topic	Details
Introduction to basic epidemiology	5 lectures per week; two semesters
Epidemiology study designs	Cross-sectional, case-control studies, cohort studies, clinical trials
Measures of disease frequency	Probabilities and odds; prevalence, incidence
Measures of association	Relative risk, odds ratio, incidence density ratio
Measures of impact	Attributable risk
Diagnosis	
Sensitivity and specificity	Test characteristics, predictive values
Likelihood ratios	ROC curves
Biases	Selection, measurement, ascertainment
Research structure/measurement	
Cause-effect evaluation	
Observer variability	
Quality of life	
Clinical measurement	Risk, prognosis, validity
Quantitative research methods	
Interaction	
Confounding, Matching	
Social science methods	
Scale development	Reliability, internal consistency
Measuring social structure	
Qualitative research methods	
Epidemiology of medical care	
Outcomes research	
Quality of care	
Practical research skills	
Data collection, data management	
Human subjects and medical ethics	
Health policy and health economics	
Introduction to health policy analysis	
Health economics	
Cost-benefit and cost-effectiveness analysis	
Miscellaneous topics	
Medicine and culture	
Disease in historical perspective	
Medical anthropology	
Clinical decision making	
Meta analysis	
Cost effectiveness	
Professional development	
Writing a grant proposal	
Writing an abstract, writing a paper	
Writing statistical sections for papers	

The intensive 2-week course in data management and statistical software at the beginning of their training prepared them for the computer analytic requirements of the methods and statistical courses. They were thus equipped to handle the assignments and use the computer for their class work.

The second course on statistical methods was also tailored to the needs of the physicians. Course II was taken in the second semester of the academic calendar, and it provided them with the necessary tools to understand statistical regression models commonly used in clinical epidemiology research. This course was given as an intensive 2-week course and consisted of models such as logistic regression and proportional hazards regression models. In addition, students could choose to supplement Course II with a standard semester-long course on multiple regression techniques (BIOS 163).

During the second semester, a biostatistical critical appraisal course was offered. The innovation of this course was that it differed from the standard critical appraisal of the medical literature in that the clinical importance of the article was not reviewed, and the discussion focused instead on the statistical methods chosen. Importance was stressed on the fact that for all situations, one is faced with alternative choices, and that the selected method needs to be justified statistically.

Physicians appreciated this approach, given the analogy with diagnosis and treatment decisions they must make when faced with a particular patient during their clinical work. The selected statistical method also implies specific results, and these were the focus of discussions. Another innovation is that this course was attended by both the physicians being trained to be clinical researchers as well as the statisticians being trained to be biostatisticians, so that the interactions and the discussions were also a learning experience for all involved. A list of the contents of the 'biostatistical critical appraisal' is given in Table 3.

As shown in Table 1, the biostatistical training provided to physicians at the CERTCs is somewhat limited in the number of in-depth biostatistical courses that the physicians can take.

Table 3: Contents of a Biostatistical Critical Appraisal of the Medical Literature

Topic	Discussion questions
Study design	Is the chosen study design the most appropriate (given practical limitations) to answer the research questions and hypotheses?
Sampling	Is the Population sampled reproducibly described?
	Is the Sampling method reproducibly described?
	Are the inclusion and exclusion criteria clearly defined?
Sample size	Is the sample size justified?
	Is it adequate to answer the research questions with sufficient power?
	Are there non-response, non-compliance, and loss to follow-up?
Significance level	Is it adequate, adjusted for multiple testing?
Quality of data	Are procedures standardised; is staff trained adequately?
	Are data entry and management appropriate?
	Is handling of missing data adequate?
	Are possible biases discussed?
Variables	Are they clearly defined and objectively measured?
	Are transformations / categorisations justified?
Statistical procedures	Are they adequately described?
	Is their choice justified; according to types of variables?
	What alternative procedures are possible?
	What implicit / explicit assumptions are needed, justified?
Results and interpretations	Are they supported by the statistical procedures?
	Is statistical terminology properly used?

The medical professional is thus not considered to be independent statistically, but the Network is designed so that the necessary biostatistical support is available at the home Clinical Epidemiology Unit (CEU). This approach enhances the functioning of the CEU by encouraging a close collaboration among members of a CEU.

Despite advances in communication and the vast resources available through the internet, most researchers in many developing countries continue to feel isolated from technological advances in the statistical field, and the efforts of providing continuing education to medical researchers and biostatisticians is an integral part of the program. The topics and types of continuing education courses offered to physicians and biostatistician are quite varied, but concentrate on methodological issues pertinent to their clinical research activities.

Thus, topics have ranged from epidemiological methods [meta analyses, diagnostic tests, issues in clinical trials], research methods [management of multi centre studies, scientific writing, grantsmanship], to statistical methods [interim analysis for early stopping, multiple imputation for missing data, modelling longitudinal repeated measures data].

The continuing education courses are essential to the continued nurturing of the trained professionals, and thus tend to be more inclusive rather than exclusive by discipline. Thus clinical researchers may opt to take a short course on a more statistical topic, and a biostatistician may choose to take an epidemiology methods short course.

The courses are typically 1-week long short courses offered by faculty from one of the original training centres in collaboration with faculty from a regional CERTC. In a given year, from 1-3 courses may be offered in different regions globally, hosted by a local CEU or CERTC. In addition, regions of the INCLEN network have annual professional meetings, and the opportunity for offering short courses to a larger audience with less travel expenses is often taken. The short courses offered in conjunction with these annual meetings will typically be only 1-2 day workshops, since it is difficult for the professionals to be away for an extra week in addition to the time away for the conference.

The continuing education courses are offered on a sporadic basis, usually upon request from a group of CEUs, and based on the availability and expertise of the faculty of the training centres. The topics are not selected through a structured process, although periodically the members of the network are polled as to possible future topics of interest.

As part of the continuing training in the future, web based distance learning courses similar to those presented by Lee (2001) and Stangl (2001), but using ‘case studies,’ are currently being developed by the regional training centres. Researchers will be able to post problems on a web site. Through an interactive ‘chat’ process, alternatives will be discussed, exploratory data analysis will be conducted and evaluated, strategies for statistical modelling will be discussed and evaluated, and analyses will be conducted and interpreted.

3. INTERNATIONAL CLINICAL EPIDEMIOLOGY NETWORK TRAINING OF BIOSTATISTICIANS

Originally, the training of statisticians into biostatisticians was naïvely thought by the network to be analogous to the training of physicians to be clinical epidemiologists.

Thus, the INCLEN program funded a single year of post-graduate training for statisticians as well.

However, most statisticians seeking further graduate-level training already had adequate master's level training in mathematical statistics, and thus a further master's degree was not useful to their personal careers. Furthermore, given that upon return to their countries, the statisticians were expected to be integrated members of the CEU, not having a doctoral degree was a clear professional handicap for the statistician. Administratively, a non-physician in a medical school faced institutional difficulties in even getting an appointment in the medical faculty.

Professionally, not having a doctoral degree made interactions in the 'close collaborating team' of the CEU not completely collegial. Eventually, the need for doctoral level training was recognised by the Network as many individuals opted for doctoral training after seeking additional support beyond the one-year provided by the Rockefeller Foundation.

The content and methods for the training of the statisticians are presented below. Many individuals opted for either a traditional master's degree in biostatistics, or for a doctoral degree in biostatistics. These individuals took the standard curriculum from the Department of Biostatistics as any other student in those degree programs. The details of these *curricula* are presented in Table 4 for the University of North Carolina at Chapel Hill. Also, students seeking the MS, DrPH, or PhD degree must meet the following additional requirements for intermediate or advanced biostatistics or statistics course-work:

- *Master of Science (MS):* Three hours of courses numbered above l65 in Biostatistics, or equivalent;
- *Doctor of Public Health (DrPH):* At least 12 hours of courses numbered 200 or above in Biostatistics, or equivalent;
- Doctor of Philosophy (PhD): At least 12 hours of courses numbered 200 or above in Biostatistics, or equivalent.

In addition, Table 2 above provides the details of the non-standard curriculum course taken by the participants as part of the INCLEN training. The statisticians took this Core Curriculum in clinical epidemiology research methodology, an intensive 9-hour credit course for 2 semesters. The physicians also took the same Core Curriculum, and thus all individuals had a uniform understanding of a common language for their future interactions and collaborations.

Although there is no perfect formula that works for training all statisticians, there is general agreement that statistical training for collaborators in applications of statistics should be driven by the specific considerations of the fields of application. In addition, special supplementary courses were imparted on critical appraisal of the medical literature from a biostatistical viewpoint (see Table 4 above).

Other innovative aspects of the training included the involvement of the biostatisticians in consultation and co-teaching of workshops, since the biostatisticians required such experiences in order to effectively provide such services at their respective CEUs.

Table 4: UNC Biostatistics Degree Requirements (in Effect During INCLEN Training at UNC, 1986-1995)

	Statistical Courses by Type	Degree for which Required			
Computing		MPH	MS	DrPH	PhD
BIOS 111	Introduction to statistical computing and data management	X	X	X	X
Probability and Statistical Inference					
BIOS l50	Elements of probability and statistical inference		X		
BIOS l60, l6l	Probability and statistical inference I and II			X	X
STAT 134, l35	Intermediate statistical theory I and II				X
Biostatistical Applications					
BIOS l45	Principles of experimental analysis	X[a]			
BIOS l62	Introductory applied statistics	X[b]	X	X	X
BIOS l63	Introduction to linear models		X	X	X
BIOS l64	Sample survey methodology	X	X	X	X
BIOS l65	Analysis of categorical data		X	X	X
BIOS l67	Applied stochastic processes				X
BIOS 263	Advanced linear model theory			X	X
BIOS 266	Advanced linear model methods			X	X
Professional Biostatistics					
BIOS l9l	Field observations in biostatistics	X	X	X	X
BIOS 341	Principles of statistical consulting (2)	X	X	X	X
BIOS 342	Practice in statistical consulting (2)	X	X	X	X
BIOS 350	Training in statistical teaching in the health sciences (3)			X[c]	X
Paper/Dissertation					
BIOS 392	Master's paper	X	X		
BIOS 394	Doctoral Dissertation			X	X

Notes:
a - Or BIOS l63
b - Or a BIOS course numbered above l64
c - Three extra credit hours of BIOS 342 may be substituted for this requirement.

4. DISCUSSION AND CONCLUSION

The success of this endeavour in training physicians to be clinical epidemiologists and in training statisticians to be biostatisticians has been mixed. One of the original intents of the INCLEN program was to establish these CEUs as centres for excellence in various countries so that local well-trained clinical researchers would study the health problems faced in those countries. Other training programs with such lofty goals have failed in the past, as trainees did not return to their countries, thus creating a ‘brain drain.’ The establishment of a CEU that provided a career path in research, coupled with a variety of monetary and travel incentives, was the planned answer. This has worked in some situations, but in others, given the value of these trained individuals, internal brain drain has meant that the qualified trained clinical researchers and biostatisticians have

left the CEUs for more lucrative positions within private enterprise and international organisations. This is especially true for the biostatisticians within the team, primarily since biostatisticians are a rare commodity in many countries, but also because they do not have high salaries within the medical faculty since they are not clinicians.

The goal of educating the medical professional to be a researcher with an understanding of statistical methodologies and issues is a lofty one, but attainable. There are many possible training methods (Bangdiwala 1993), and the specific methods of the INCLEN training program have achieved reasonable success in many countries. Training health and medical professionals in statistical methodology is not a threat to the statistical profession, but actually enhances our profession. Given the potential abuses of statistical methods made easy by the proliferation of statistical software readily available, the potential danger to the profession from the ill-trained casual user is great. Thus, an educated consumer is the best client for our statistical profession.

The training of statisticians to work effectively as team collaborators with the clinicians has been a harder battle to win. Challenges exist in many institutions from an administrative standpoint, such as in reduced possibility of an academic appointment in a medical school, and limited ability to get promoted within the medical faculty as a biostatistician, coupled with reduced salaries because of no clinical practice. Biostatisticians are thus more likely to be lured away by private enterprise, and thus the CEU suffers. From a training standpoint, this INCLEN model requires 3-4 years for training a doctoral statistician contrasted with just one year for a clinical epidemiologist. This expense is not covered by the original funding agency and has meant that not all statisticians desiring a doctoral degree have been able to attain it. Finally, from a professional standpoint, the career development of the biostatistician requires continued contact and interactions with other biostatisticians and especially continued education. In addition to the training received during the degree program, given the professional isolation of the biostatisticians, continuing education training is still an important part of the program. Biostatisticians participate in the same continuing education opportunities as described above for clinical epidemiologists. However, there is still a gap in updating the training with the recent advances in computer intensive methodologies in the statistical field, by the fact that trainers within the Network may not be in a position to provide such continuing education, while outsiders command high honoraria.

The challenges in training professionals to research the pressing health priorities of developing countries are numerous. The INCLEN model includes training physicians in statistics, and training statisticians in clinical epidemiology methods. The complementary roles are institutionalised into clinical epidemiology units, and despite some pitfalls encountered, has been an effective way to increase the clinical research capability of these countries. Until a critical mass of biostatisticians are available within a country, continued education courses for the statisticians on advances in the field are essential for their career development.

REFERENCES

Altman, D. G. (1994). The scandal of poor medical research. *British Medical Journal*, 308, 283-284.

Altman, D. G. (1998). Statistical reviewing for medical journals. *Statistics in Medicine*, 17, 2661-2674.

Bangdiwala, S. I. (1989). The teaching of the concepts of statistical tests of hypotheses to non-statisticians. *Journal of Applied Statistics*, 16, 355-361.

Bangdiwala, S. I. (1993). Teaching biostatistical concepts and methodology to medical professionals: Various alternatives. In L. Brunelli, & G. Cicchitelli (Eds.), *Proceedings of the First International Association for Statistical Education* (pp. 315-320). Perugia, Italy: International Association for Statistical Education.

Coste, J., Fermanian, J., & Venot, A. (1995). Methodological and statistical problems in the construction of composite measurement scales: A survey of six medical and epidemiological journals. *Statistics in Medicine,* 14, 331-345.

Crivisqui, E., Abruzzini S., & Batista, C. M. (2001) How to bridge the gap between the statistical methods available and their effective and potential use by researchers in social sciences. A few thoughts about the experience of the PRESTA programme. In C. Batanero (Ed.), *Training Researchers in theUse of Statistics* (pp. 277-293). Granada: International Association for Statistical Education and International Statistical Institute.

Estepa, A., & Sánchez Cobo, F. T. (2001). Empirical research on the understanding of association and implications for the training of researchers. In C. Batanero (Ed.), *Training Researchers in theUse of Statistics* (pp. 35-51). Granada: International Association for Statistical Education and International Statistical Institute.

Halstead, S. B., Tugwell, P., & Bennett, K. (1991). The International Clinical Epidemiology Network (INCLEN): A progress report. *Journal of Clinical Epidemiology,* 44, 579-589.

Hand, D. J. (1994). Deconstructing statistical questions. *Journal of the Royal Statistical Society A*, 157, 317-356.

Lee, T. R. (2001). Teaching biostatistics to medical personnel with computer based supplement. In C. Batanero (Ed.), *Training Researchers in theUse of Statistics* (pp. 139-145). Granada: International Association for Statistical Education and International Statistical Institute.

Ospina, D., & Ortiz, J. (2001). Statistical research and consulting in universities from developing countries: The Colombian case. In C. Batanero (Ed.), *Training Researchers in theUse of Statistics* (pp. 295-310). Granada: International Association for Statistical Education and International Statistical Institute.

Stangl D. (2001). Design of an internet course for training medical researchers in Bayesian statistical methods. In C. Batanero (Ed.), *Training Researchers in theUse of Statistics* (pp. 147-157). Granada: International Association for Statistical Education and International Statistical Institute.

Shrikant I Bangdiwala
University of North Carolina, School of Public Health,
Department of Biostatistics, CB# 7400
Chapel Hill, NC 27599-7400, USA
E-mail: kbangdiw@bios.unc.edu

Sergio R. Muñoz
Facultad de Medicina, Universidad de La Frontera,
Temuco, Chile
E-mail: munozs@ufro.cl

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 277-293.

Printed in Granada, Spain.

EDUARDO CRIVISQUI, STEFANO ABRUZZINI & CARLOS MARCOS BATISTA

HOW TO OVERCOME THE GAP BETWEEN THE AVAILABLE STATISTICAL METHODS AND THEIR EFFECTIVE USE BY RESEARCHERS IN SOCIAL SCIENCES. A FEW THOUGHTS ABOUT THE EXPERIENCE IN THE PRESTA PROGRAMME

PRESTA is a training programme in applied statistics for teachers and researchers in South American universities, sponsored by the European Union, which started in 1994. It was intended to make a convergent presentation of "exploratory data analysis" (in the Anglo-Saxon tradition) and " data analysis" (in the French tradition) to a "critical mass" of users. The seminars organised in its first quinquennium were attended by 2,500 researchers and teachers from about 300 universities. Working in the field, we soon realised the importance of the distance between the currently available statistical methods and their effective and potential uses. Therefore, the research programme is focused on the pedagogical contents and the institutional components of this "gap". In this paper, we summarise our experience with regard to the main factors that led to this situation and the strategies developed to solve it.

1. INTRODUCTION

Since the eighties, the Data Processing Methodology Laboratory of the Université Libre de Bruxelles, Belgium, in collaboration with other European universities, has been holding statistical methods seminars at the main public universities in Argentina, Bolivia, Chile and Paraguay. These seminars have basically covered statistical methodology and associated data processing technology, applied to sampling surveys in socio-economic studies.

In 1993, this experience led us to propose a different type of international co-operation with South American universities in the field of training researchers in statistical methods: The "Programme de Récherche et d'Enseignement en Statistique Appliquée" (PRESTA – Programme of Research and Training in Applied Statistics). This initiative got immediate support from other European higher education institutions and the sponsorship of the European Union.

We first present the outline of the "diagnosis" on which the PRESTA programme was based (Section 2) and then we summarise its method of operation and main achievements (Section 3). In the last two sections, we present a critical reflection about the gap observed between statistical methods currently available and their effective use in social research. Section 4 deals with the course's content and the didactic approaches experienced in the seminars organised. Since the experience was carried out on a large regional scale, we took into consideration the social and institutional dynamics that determined the persistence of such gap, which is analysed in Section 5.

2. SOCIO-ECONOMIC RESEARCH IN SOUTH AMERICA SINCE THE EARLY NINETIES

The co-operation experience with Latin America made it possible to observe, throughout more than a decade, the gradual development of a certain mismatch between the growing number of studies that request the use of statistical methods and the availability of professionals to carry them out.

2.1. EXPANSION OF THE DEMAND FOR STUDIES

The increasing demand for socio-economic studies in Latin America during the Eighties should be examined in the context of the economical and social transition that the majority of the Latin American countries were undergoing. In fact, the return to constitutional regimes was accompanied by governmental budgetary adjustment policies. Imposed by the burden of their foreign debts, these policies had considerable social, economic and political effects, since they mainly relied on large-scale privatisation plans for national companies and decentralisation of the power, without decentralising their financial resources, and preparing the human resources to take over their new responsibilities.

In this context, Latin American countries had to reorganise simultaneously their local government, regional government and the government of large cities. These countries once again established regional economic agreements, which intended to create new economic areas (Mercosur Agreement, Pacto Andino). These new forms of government naturally needed new studies to understand and define their policy. As a result, there was a widespread demand for statistical information, which could not be fulfilled by the former official statistical systems, and help from universities and the private sector was requested. This caused a proliferation of local and regional sources of statistical information, without any apparent effort to agree on methodology. Consequently, statistical data lost credibility, which lead to an increasing demand for additional studies.

Such expansion of demand would not have been possible without the great number of microcomputers available, which reduced considerably the costs of data storing and processing. Therefore, public administrations were able to create a large number of databases at low cost, sometimes without taking into account ethical considerations. The commercial distribution of statistics programs made it seem easy to exploit the information stored in these databases. New software and computer equipment was acquired and the budget that should have been partially devoted to training professionals was limited to "black box" training in the use of statistical software.

The confusion, maintained by commercial companies between (lack of) statistical knowledge and knowledge of the *modus operandi* of statistical software, enabled a false alternative to be strengthened, or even legitimised, in academic areas: "Pure statistics" versus "applied statistics". In addition, the statistics training of the "users" was dependent on the statistical software available, which was created for a different socio-cultural context and to meet the requirements of the dominant economic markets.

2.2. RESEARCHERS' PROBLEMS IN HANDLING THE NEW REQUESTS

The training of senior staff responsible for carrying out these studies highlighted the deficiencies of higher education structures, which were unable to steer the process of technological innovation needed in all fields related to the systematic observation of collective phenomena [(1)]. Local higher education groups could not participate in this

process, due, especially, to the deteriorated condition of the main public universities, after a long period of authoritarian regimes. Public universities had to adapt to constitutional regimes while becoming "mass" universities, and being subjected to a reduction in funding. This justified the opening of higher education to the market, and the attempts to privatise the existing public institutions, which lead to a quick proliferation of higher education institutions, in a general context of scarcity of qualified lecturers. This shortage, combined with the mechanisms for evaluating the institutions, led to the creation of many specialised courses, which did not meet the need for training teaching staff. The classical choices of training abroad were not appropriate to meet the needs on this scale.

The pedagogical content and forms for the statistics training of professionals in social sciences were poorly suited to the labour market requirements and changes. In sections 4 and 5, we analyse the conditions for these transitions.

2.3. THE INITIAL COOPERATIVE STRATEGY IN THE PRESTA PROGRAMME

The aforementioned facts made us consider that the co-operation between European and Latin American universities should not be restricted to sporadic organisation of post-graduate courses. Therefore, a co-operative strategy in five stages was devised to avoid the main factors that prevented the local supply of trained senior staff, while creating a lasting and autonomous local system for statistical training. In 1994 a five-year programme of inter-university co-operation sponsored by the European Union was started with the following aims:

a) To promote co-operation between European and South American universities in the field of statistical methods applied to the social sciences, by developing:

- The training of South American university lecturers on recent statistical methods useful for social sciences;
- The technical and scientific contribution to "distance statistics training" and the European specialists support to stimulate the dissemination of this activity in other universities in the region;
- The dissemination of scientific reference documentation in Spanish and Portuguese, accessible to all Latin American researchers;
- A network of European and South American university laboratories, joining forces on applied and/or basic research projects, including quantitative research methods in social sciences.

b) To promote co-operation among various South American universities, and among universities and the national public institutions responsible for socio-economic research or action, by encouraging:

- The organisation of regional seminars for training researchers and senior staff in the field of recent statistical methods used in the social sciences;
- The development of regional "distance education" services of applied statistics in social sciences;
- The development of joint research with the methodological support of European laboratories.

3. ACTIVITIES CARRIED OUT BY THE PRESTA PROGRAMME: OPERATING METHODS AND RESULTS

The PRESTA Programme was conceived over a central question: How the demand for updating the knowledge of statistical techniques and methods for South American "users" (real or potential) could be adequately met?

3.1. TRAINING TRAINERS

We had to devise a process that was sustainable and extensive enough for training local lecturers who could gradually take over the teaching responsibility in the planned training of a "critical mass" of South American professionals [(2)]. This was necessary since the voluntary participation of European lecturers was limited by their institutional responsibilities.

During the first four stages, eight "Training Trainers" sessions and eight "Pedagogical Workshops"[(3)] were carried out. We also ensured the participation of local lecturers in the various "thematic seminars" held. The *Latin American Applied Statistical Workshop* was an effective complement to the process of training local lecturers. 344 people took part in these training activities, which corresponded to 250 lecturers from 10 South American countries. In the first stage of the programme, 90 of these lecturers took part in the teaching team and were responsible for training researchers, at least once. There are many testimonies from colleagues and academic staff of participant institutions about the considerable impact from the "Training Trainers" sessions. Many lecturers have incorporated the content acquired during the training into their teaching. Consequently, a great number of Social Science Faculties have been able to adapt their curricula to reflect new content, due to the availability of trained lecturers.

Content of training courses

The innovative character of the content of the "Training Trainers" activities was widely recognised. However, the conventional training of the Latin American teaching staff imposed certain limitations that will be discussed in section 3.2 and that determined the orientation of the course’s content, which was based on the educational requirements presented in section 4.

Training modalities

The local lecturers liked many aspects of the training offered, which were different from their conventional training. They appreciated the organisation of intensive training cycles, which were organised in two full-time 10-day sessions with an inter-session period where participants had to carry out a personal, guided application. They also valued the scientific documentation they were able to access throughout this training.

The organisation of "pedagogical workshops" was also highly appreciated. In Latin America, no previous similar experiences, where lecturers collectively reflect about the best way of teaching statistical methods to professionals, had been organised.

The PRESTA network and "horizontal" co-operation tools

The programme gradually developed tools for "horizontal" co-operation and training. The participation in the "Training Trainers" cycles was deliberately limited to two lecturers from each institution. This served to establish genuine inter-institutional training groups, and to encourage representation from a large number of institutions in

all the training courses. In this way, we created favourable conditions for the collaborative work of lecturers from different countries, and for the co-operation of lecturers from large universities and other not so well known institutions. This was one of the key factors for the operation of the PRESTA network.

In time, the PRESTA electronic list(4) became a forum for exchanging and communicating scientific and technical information between lecturers and Latin American institutions. Organising the PRESTA network in four zones, which cover the 10 countries taking part in this programme, facilitated the development of these "horizontal" co-operative links. These zones were established according to pre-existing inter-university networks, so that the programme action could complement or reinforce existing links between institutions in the same or different countries.

The establishment in 1998 of the PRESTA Programme International Co-ordination Committee, with representatives from each zone, fostered the operation of a co-ordinating structure closer to the beneficiary institutions and individuals. Consequently, the committee was better suited to manage the needs, and to co-ordinate the human resources involved in the programme.

3.2. TRAINING A "CRITICAL MASS" OF PROFESSIONALS

The demand formulated by the beneficiaries of the co-operative programme and their geographical and institutional dispersion required a structure for continuing education, at the post-graduate level, which included all the higher education institutions, was adapted to local needs and conditions, and, at the same time was flexible enough to change in form and content as the needs and conditions evolved.

The training of the final beneficiaries of the programme was mainly carried out by 66 training modules or "local seminars"(5) and 19 "thematic seminars". This action continued through continuing training seminars on research methods, set up by the programme reference centres. The training of local researchers was intensified through technical and methodological assistance to their work and research projects.

Carrying out such a high volume of courses by teams of two or three lecturers required the organisation of 10,600 lecturer/hours, 67% of which took place during the last two stages of the programme, mainly by local lecturers. The 2,550 participants in these activities corresponded to about 2,300 South American professionals, including a great number of sociologists, economists, demographers, psychosociologists, historians, geographers and statisticians(6).

The final aims of the programme also required the training of groups with a large inter-institutional membership, which lead to a rapid growth in the number of beneficiary institutions. Over 800 institutions were represented in these training seminars, half of which were research units from the main Universities of ten South American countries. The other half included public research institutions, national or regional government research departments, large municipal administrations, official statistics offices, and NGOs (Non Govermental Organisations) research centres or similar organisations.

All these seminars were aimed to facilitate training to the entire target group of the programme in each country concerned. The achievement of this coverage objective was due to the constant efforts made to launch the training activities, by moving them as far as possible from the major urban centres to locations in the interior of the beneficiary countries(7).

Content of training courses for professionals

The first aspect that attracted the participants' and managers' of local institutions attention was the fitting of training content to the local needs and reality.

We had observed that the training of statisticians and professional "users" was almost exclusively limited to an approach to statistical instruments oriented towards "confirmation" or modelling of the observed phenomena. Particularly in the training of "users", an inappropriate separation between the "quantitative" and "qualitative" methodology was maintained. This purely academic hiatus even acquired a certain reality quantum among students, as it was combined with a radical separation between training in statistical methods and the introduction to research methodology. As a consequence, the training provided was inadequately oriented towards the analysis and solution of specific problems. Our courses were, therefore, deliberately focused on an exploratory approach to statistical methods, by presenting a series of tools directly linked to research.

Traditional teaching of statistics emphasised the algebraic presentation of methods, without explaining the conceptual and practical links between them and their practical application in research. This training, incorrectly considered as "theoretical" basic training in statistics, was moreover received at very different levels by the professionals of the various countries concerned. There were considerable differences in statistical expertise between professionals of different disciplines, and between researchers in the same discipline trained in different institutions, even within each country. The training groups were therefore very heterogeneous from the viewpoint of their formal statistics knowledge and this heterogeneity could have been an insurmountable handicap, had it not been incorporated as a *de facto* reality into the conception of the pedagogical process adopted in the training. This is the reason why we refined the instruments for distributing these seminars, and we adjusted the criteria for the selection of the candidates, giving a greater weight to their experience in using quantitative research methods as compared to their previous training in statistics.

This approach avoided the pitfall of a theoretical training with contents unsuited to the beneficiaries' expectations, and requiring the use of their deficient knowledge base. At the same time, we also avoided reducing the professionals' training in applied statistics to the popularisation of specific techniques and software.

Training professionals modalities

The post-graduate training available in these countries usually extended the conventional curricula for training professionals to specialised master degrees or doctorates. Post-graduate training had been organised into long courses, with contents unsuited to the requirements and development in science and technology. Consequently, the professionals' demand for updating their statistics knowledge was not covered.

Our strategy for permanent training in statistics proved to be well suited to the local conditions. The option of creating local teaching teams within the PRESTA network zones, while seeking the participation of local institutions in the preparation of each seminar, enabled to gather and mobilise the human and material resources available to serve a sector of the target population of the programme who lived in places rarely reached by international scientific co-operation programmes. This also served to experiment, on a vast regional scale, the integration of new educational technologies and the specialised human resources locally available for improving the training of local professionals. The training system created was well adapted to the beneficiaries' expectations. Focusing on a pedagogical process at a higher education level, it benefited from the multidisciplinary nature of groups, and even from their heterogeneity, to

favour peer training and to lead to a collective acquisition of the skills.

An initial period devoted to balance the participant's knowledge, served to review the basic concepts, and to create a "common language" in these multidisciplinary groups. Local seminars were extended by [(8)] short consolidation courses, devised to help the participants to gradually move towards using the methods taught on real data from their own research. The lecturers provided methodological assistance parallel to the seminars, and at the end of the course, a session was devoted to presenting and criticising some participants' research work. These sessions also served to make this ongoing work more widely known, and to establish collaborative links between local researchers.

The training system was equally effective in encouraging local development of other training courses. It was evident that to provide specialised training to South American professionals, local and thematic seminars had to be linked, and the creation of permanent research methods seminars with consolidation courses had to be encouraged. The rapid development of training activities in the beneficiary countries showed the enthusiastic reception to the initiative, and justified the co-operative strategy adopted, which would not have been possible without the voluntary unpaid contribution of many European and South American lecturers, who devoted a great deal of time and hard work to preparing, organising and performing the programme activities.

4. MAIN PEDAGOGICAL RESEARCH CARRIED OUT IN THE PRESTA PROGRAMME

Working in the field soon made us realise the importance of the "distance" between the statistical methods available and their use in social research in Latin America. Therefore, appropriate instruments were gradually devised to reduce the effects of the determinant factors for this situation, at least those that could be affected by inter-university co-operation. The programme research activity took a critical look at "statistics teaching" offered to "users" in socio-economic research. The main components of this analysis are presented in this section.

From certain attitudes towards statistics among professionals, a lack of knowledge in the programme beneficiaries about statistics as a set of instruments to observe collective phenomena was noticed. The statistical training of these professionals maintained a parallelism between the teaching of statistical methods and the teaching of the research methods specific to each discipline. It was expected that future professionals have sufficient maturity and wisdom to produce, individually and spontaneously, an operational synthesis of the content of these two areas of learning, while the lecturers themselves only very rarely tried to produce such a synthesis.

The training of "users" was based on too strong a use of the *notion of measurement* neglecting the *concept of classification*. This meant that the professionals acquired a misconception of data, which was an obstacle to their understanding of the recoding of variables. They were not conscious that the measurement scale of an observed phenomenon is not intrinsic to the data, but a "language" chosen by the researcher to describe a collective phenomenon. By forgetting the concept of comparison, the training provided to future researchers produced a loss of sense of the concept of information produced by the observation of collective phenomena.

The training of professionals devoted little attention to those statistical-methodological aspects that enable us to evaluate the coherence between the research theoretical framework and the consequent technical choices. Some examples are the

strategies for managing uncertainty and error, the choice of an appropriate representation space based on the creative recoding of the observational data; the choice of weightings and even of substitution rates implicitly adopted by the majority of multivariate methods; or the choice of properties in any aggregate function.

These deficiencies in training resulted in the users' great difficulties in choosing from the vast array of tools available, and in justifying the appropriate statistical method to deal with a research problem. It was often observed that researchers were inclined to "delegate" these operational choices to the statistician, or they even were forced to "blindly" use the default options of statistical software. The observation of these deficiencies in the researchers' *modus operandi* led to concentrate the didactic effort on the two areas that will be presented below.

Making the use of descriptive methods a crucial feature of statistical knowledge

The content of training courses for trainers and the majority of local and thematic seminars was based on the convergent presentation of methods specific to the Anglo-american "exploratory data analysis" and to the "data analysis" French school. This group of methods formed the "toolkit" that should be available to researchers and lecturers working in non-experimental statistics (human and social sciences, economics, public health, environmental sciences, agronomy, etc.) to be able to meet the requirements of local demand for socio-economic studies.

The methods, which were unfamiliar to the South American scientific community, were presented as "tools" for building representations of complex phenomena, which demanded, first of all, an exploratory approach to reach a comprehension of the same, before any attempt to "freeze" them into a pseudo-explanation model. Unidimensional and multidimensional exploratory statistics are often perceived, in academic circles, as a mere group of foundation concepts for statistical modelling and inference. However, by guiding the training provided by the programme towards local professionals' greater ability to solve specific research problems, this conventional academic hierarchy could be contradicted.

In presenting these methods, the emphasis was placed on the analysis of a group of observations rather than on the relationship of a group of variables. This shift in emphasis contributed to the development of a great capacity among the researchers to use statistical methods for building synthetic representations of their subject of study, and to consider these representations as "tools" for understanding the phenomena observed.

Didactic strategy: giving priority to comprehension of the methodology

The second key area of specific didactic research was based on the purpose of the seminars organised by the programme to train social science researchers, i.e. to train "users" of statistical methods, rather than statistical researchers, who are supposed to develop new methods. That is why the presentation of the rational basis and operating principles of the methods should be oriented towards understanding the instrument methodology. A deductive presentation, with a pronounced mathematical formalisation, would have been an obstacle rather than an aid to understanding the research instruments, and we decided to rely, as much as possible, on the language and concepts of Euclidean geometry. This strategy made it possible to base teaching on an intuitive perception of the properties of representation spaces, which facilitated the understanding of essential concepts to social science researchers, without sacrificing rigor and operational understanding. This also contradicted the belief that "users" cannot acquire sufficient understanding of the rational basis of these methods, due to their

ignorance of the formal language of linear algebra.

Computing developments, which have facilitated graphical representation, reinforced this didactic choice by incorporating several dynamic geometrical representations, data and methods into the courses. This didactic approach also offered a communication code that transcended the disciplines and specialisation of the researchers.

5. INSTITUTIONAL ASPECTS OF THE DISTANCE BETWEEN STATISTICAL METHODS AND THEIR EFFECTIVE USE

The scale of the task of disseminating knowledge enabled the identification of some institutional dynamics that contributed to maintain, or even to reinforce a certain "statistical cultural vacuum" among the professionals responsible for carrying out research studies in a broad sector of Latin America. It was impossible to develop a training structure while ignoring the institutional constrains that inhibited the dissemination of knowledge, and which encouraged an inappropriate "praxis". In fact, four main dynamics were identified.

5.1. DIFFUSSION OF SCIENTIFIC KNOWLEDGE

The university publishers, who assured the functioning of traditional ways to disseminate scientific and technical knowledge, were reduced to adapting to market forces, which cannot assure an adequate dissemination of updated documentation to the scientific community. When ad hoc mechanisms for access to documentation disappear, two undesirable and complementary effects occur: On one hand, certain scientific traditions (or schools) get a dominant position due to the mere fact of having access to the publishers controlling the distribution channels and the scientific publishing market; On the other hand, when the dissemination of scientific documentation is limited, the scientific community develops a dependent behaviour with regard to knowledge (which had become rare in the local environment), leading to strategies of individual and passive appropriation of this knowledge. These symptoms were observed in the Latin American University milieu, although the effects of this disappearance were even more harmful in non-university environments.

That is why all the programme activities encouraged the dissemination of updated scientific and pedagogical documentation among South American professionals. It was mostly translated into Spanish and Portuguese, and consisted of lecture notes, guides to solving the application tasks, user manuals for computer programs, books and scientific papers. In addition, "student kits" were also distributed for the main computer programs used in the courses. An essential part of the lecture and guided study notes and the reference bibliography in the local seminars, were made available to the scientific community through the Programme's web site, in a series of Acrobat® documents.

A minimum document base was created in the scientific libraries of each local institution that hosted a seminar and a larger one in the reference centres responsible for facilitating the access to documentation to the PRESTA network members. The journal *"Metodológica"* was another distribution channel for specialised scientific documentation. This journal publishes papers on methods and techniques for quantitative research in the social science from European and Latin American researchers in a variety of fields.

5.2. THE DIVISION OF TASKS ADOPTED IN THE RESEARCH TEAMS

The weak penetration of a "statistics culture" and the absence of university extension programmes that could relate the “users” and the university statisticians encouraged the adoption of an absurd division of tasks between the various professionals involved in the elaboration of the socio-economic studies. In this work, the "field staff" were the least-experienced professionals, responsible for the observation stage, consisting not only of gathering and preparing the observational data, but also of designing the survey questionnaire.

The "statistician" was not responsible for bringing the tools that would objectively process the information produced by the "field staff" into these research systems, but his job was just to "demonstrate" the scientific validity of the system, by contributing to the definition of the representative sample size, and by ensuring the "processing" of the "responses" obtained. The statistical data analysis was carried out without any reference to the conditions and objectives of the observation. and the "statistician" was often asked to identify "significant" relationships and "interesting models" based on the observational data. Finally, at the end of these disconnected operations, a third participant, the "specialist", was brought into operation to interpret the statistical results and write the final report. This breakdown of tasks between three participants in the research leads to the denaturing of the collective and multidisciplinary practice of research, and disqualified all the professionals involved.

We could not envisage organising the training of researchers without addressing this unhealthy traditional division of tasks within research, and therefore all the programme activities aimed to transfer the experience and knowledge required to organise multidisciplinary research teams. We collaborated directly on the development of local teams of joint research projects in various fields, facilitated meetings between research teams, and contributed with methodological and organisational experience to prepare joint and multidisciplinary research projects in various fields.

5.3. POST-GRADUATE TRAINING (MASTERS AND DOCTORATE)

Faced with the shortage of human resources specialised in applied statistics, the Latin American universities had developed a number of post-graduate courses "in statistics". This considerable effort had a number of disadvantages.

First, this specialised training offer was not included in a human resources policy of this sector, in agreement with each country's specific requirements. The courses were often isolated initiatives of higher education institutions, which led to very local recruitment of students who continued their studies in this way. It is interesting to note that the demand for such training courses was well below that for PRESTA training activities. The mismatch between offer and demand for this training is produced by the lack of sufficient integration of the specialised human resources available at a regional level.

In fact, these post-graduate courses are often set up with the academic staff available at the institution, and according to criteria of excellence which generally lead to reproduce the equivalent curricula offered by European or North American institutions. Thus, an abstract strategy of promoting high-quality scientific research absorbs the few resources available, in detriment of any complementary strategy based on the continuing education of working professionals. Therefore, the supply of training generated is qualitatively and quantitatively out of line with the requirements.

The passive repetition of foreign courses produces an impoverishment in the content

of these post-graduate courses. In addition, their content reproduced graduation courses at a more advanced level, which led to a lack of interest towards post-graduate courses among working professionals. Despite all, there is an interesting potential of specialised staff in Latin America, which would allow a programme of continuing education for active professionals. This would lay the foundation on which high-level scientific research could be built, and which would be managed by university laboratories, local research institutions or official statistics offices.

5.4. UNIVERSITY EXTENSION PROGRAMMES

With the disappearance of university extension policies and the impoverishment of specialised training in applied statistics, the university statistical laboratories have reduced their contacts with "users" and diverted the aims of the essential contacts between academic statisticians and the organisations performing socio-economic research. In fact, the university extension program has been reduced to "consultancy" assignments to a few specialists, ignoring, therefore, the importance of the transference of knowledge that the University is supposed to provide. That is the reason to attempt to revive the university extension tradition, by developing, through a specific experiment, the forms and content of a large two-way transfer of knowledge between a few groups of University statisticians and broad sectors of statistical methods users in research.

5.5. AN ACTIVE INSTITUTIONAL POLICY FOR PROMOTING HORIZONTAL RELATIONS

Being aware of these institutional dynamics, it was indispensable to carry out a practical experience of the training potential that could arise from strengthening the horizontal relations between local and regional higher education institutions. In fact, the local institutions responsible for organising the "Training Users" seminars had to set up contacts with all the study and research centres in their surrounding area. The training itself fostered the practice and development of exchanges within inter-institutional and multidisciplinary groups. This led to a greater collaboration between university and non-university centres, which went beyond the framework and even the content of the PRESTA Programme.

However, it was not enough to create meeting opportunities. It was still necessary to identify the concrete spheres of action and to provide active support to the participants willing to engage in new institutional practices capable of effectively neutralising the effects of the dynamics described above. Therefore, five active institutional policies[(10)] were carried out: promotion of "common academic areas"; use of new teaching technologies for continuing education of active professionals; development of reference centres as research and education units working for the benefit of various sectors of applied research; reinforcement of links between producers of official statistics and university statistics and promotion of the applied statistician professional profile. In the implementation of these policies, the programme attempted to act as a federative body, so as to support awareness-raising and active commitment by all the individuals and institutions concerned.

Promotion of "common academic areas"

The organisation of the PRESTA network enabled us to put into practice "common academic areas", a concept that was suggested by Rector Brovetto of the Universidad de la República Oriental del Uruguay, Secretary General of the "Grupo Montevideo".

Seminars were organised for training professionals several times in the network zones, where specialised human resources of various public universities in the zone (trained by the programme) were offered to train statistics "users" in universities, public administrations or research centres in a given town or region.

To reach this goal, scientific, technical and pedagogical training had to be provided to a sufficient number of local lecturers, and moreover, it was necessary to prepare appropriate didactic materials and to teach the appropriate forms of organised multidisciplinary and inter-institutional seminars. The establishment of the PRESTA International Co-ordination Committee served to disseminate the necessary organisational knowledge among local teaching staff. Many successful experiments of public higher education establishments in the region working in "common academic areas", aroused genuine interest among the local academic authorities, which realised that this experiment could be extended beyond the field of applied statistics.

New technologies in the training of professionals

From the very beginning of the programme, we considered the possibility of developing a "distance education" line of action for "training users" of statistics. However, when completing the "Training of Trainers" we realised the limitations and disadvantages of this technique in relation to our objective of large-scale dissemination of knowledge throughout the scientific community in the region.

In fact, this was an arduous task without "training effect" on the local teaching staff, apart from the small group of local colleagues in charge of the project. In addition, the training structure associated with "distance education" was not a proper network, but a series of "vertical relationships" between the training centre and the individual locations where participants were studying. We felt that this structure was inappropriate to reinforce the integrated strategy of "Training Trainers" that we adopted in parallel with "User Training". Finally, it seemed that the access to continuing training by local professionals was not due to their isolation, which would have justified the use of distance education. Apart from a few exceptional cases, the whole target population for the co-operative programme was working in institutions located in urban areas usually close to the university facilities.

It was, then, considered more appropriate to concentrate on the development of a training structure organised into networks and capable of being spread in the large urban centres, as well as in smaller university towns, so as to lead to the setting-up of continuing education institutions. From the second stage, priority was given to the development of simultaneous local seminars using videoconferencing systems that were installed since 1995-96 in a number of South American universities. In others universities, a partial access was made available through systems installed by large national or multinational companies operating in South American countries. In co-operation with the Universidade Federal de Santa Catarina, Brazil, we promoted the attractive educational properties of this new medium for the benefit of training.

In fact, the videoconferencing system brought the potential of full interactivity with simultaneous transmission of sound and pictures. It permitted simultaneous local seminars, connected over a network, with specific teaching support, which had unquestionable organisational and pedagogical advantages.

The organisation of simultaneous local seminars enabled the organisation of a higher modality of common academic areas. This could bring together the human resources of a large number of public institutions into the PRESTA network zone to assist a training activity that could accommodate several hundreds of local executives at the same time, organised into small groups scattered over a vast geographical area (for example, at the

level of the Brazilian states). Each group undergoing training attended a classic "local seminar", with theoretical courses provided by the videoconference network and guided work sessions, managed by the teaching teams, with a pooling of resources through the videoconference network. The interactivity allowed by this system enabled the "operation" of a single virtual lecture theatre, which vastly increased the number of exchanges between local researchers and teaching staff, while being geographically spread, and working in very remote institutions as regards the main university centres in the region.

From the pedagogical viewpoint, this medium enabled the use of dynamic images, and colour, which greatly improved the didactic capacity to present the rational basis for statistical methods, linked to their use in various research fields.

Development of Reference Centres

The gradual establishment of reference centres was one key feature for the "Training Trainers" network. The starting up of reference centres was subject to the dynamism of the sub-regional networks established, as the annual training plans were implemented. First, we tried to identify and train a sufficient number of motivated teaching staff to assist in the dissemination of knowledge proposed. We had in this way guaranteed the optimum convergence between these personal aims and the level of commitment by their university institutions. To achieve this goal, we carried out a broad dissemination of the didactical material for "Training Trainers" among local teaching and management staff. The selection of participants (maximum three candidates per institution) was carried out with personal criteria (candidate's qualification) and institutional criteria (reasons to participate, commitment to support the participation of lecturers in other activities proposed by the programme, etc.).

Through initiatives taken by small groups of motivated lecturers, some institutions were gradually involved in a wealth of exchanges with university and non-university institutions within the PRESTA network. We fostered the establishment of bilateral and multilateral agreements between these institutions. Where such agreements already existed, we contributed to increasing their dynamism by means of programme activities. In parallel, we increased the availability of scientific documentation in these groups, and wherever necessary, their computer and teaching equipment. We encouraged the involvement of other colleagues, and closely involved local lecturers in the preparation, organisation and co-ordination of the programme training activities.

This enabled us to meet a large number of lecturers, and we succeeded in identifying the local institutions that were prepared to fully participate in the programme co-operation by contributing with available human and material resources. This dynamic led to the establishment of reference centres with different institutional forms [(11)]. This strategy of development of reference centres allowed highly decentralised management of the majority of training seminars carried out, particularly during the fifth stage of the programme. It facilitated the mobilisation of initiatives and local resources, while making our training system more suitable to the requirements, via structures that were close to the beneficiaries, and better qualified to identify their training needs and interpret their requests.

Strengthening of links between "producers of statistics" and "statisticians"

The intensification of these links is an undeniable necessity, and not just in the Latin American context. We tried to achieve this goal throughout the five years of working on the programme. First of all, a small quota for participation in all the "Training Trainers" activities was reserved to the official statistics offices senior technical staff [(12)].

However, the training demand from this sector was mainly for local and thematic seminars in the four zones of the PRESTA network. The training departments of the main national statistical institutes requested the programme to help in the technical and methodological updating of their scientific personnel. However, we considered that this was beyond the limit of the co-operation programme and preferred to create the local conditions for the rapprochement between the staff members and university statisticians.

The training system spread out was able to bring about this rapprochement. That explains the many exchanges that have occurred among these research teams, and the enthusiastic participation of "producers" of official statistics in the training activities, as well as at the *First Latin American Applied Statistics Conference*.

Reinforcing the "applied" statistician's professional profile

One consequence of the dynamics described above, which are making the dissemination of statistics operational use difficult is the weak academic position of those statisticians interested in the ways of using these tools. To change this trend, we organised the *Annual Regional Conference on Applied Statistics* [(13)], according to the scientific colloquia standards, which has become a very effective way to stimulate the work of university and non-university research teams.

The preparation of this conference takes a year, through the monitoring of the reference centre research teams. This dynamised the exchanges within the four PRESTA network zones, and the local researchers have submitted over forty research projects to the organising committee, twelve of which were selected and presented. In organising this conference, we pursued a pedagogical objective while encouraging the collective reflection about the role of the "participants" and the structure of statistics applied to socio-economic studies in Latin America.

Through the contacts with the research groups in all the PRESTA network zones, we identified complementary methodological contents to be presented in short courses. In addition, the work by local researchers was selected and grouped into common methodological spheres, which encouraged debate and collective criticism of these applications. The composition of the audience, as well as the choice of themes in the courses and in a closing Round Table discussion, stimulated important debates. A recurring theme in the reflections was the concern to identify, the main difficulties that are hampering Latin America in the development of synergies between university statisticians and statistical "users", while recognising the technical trends and current policies for change, with a strong regional integration of the national statistical systems.

6. CONCLUSION

The activity carried out by the PRESTA Programme laid the foundations for a training system of teaching staff in higher education establishments. Therefore, the main features of this system deserve to be highlighted. This system appropriately integrated two dimensions that are very often separated in the training of university lecturers. It combined the necessary scientific and technical updating of the teaching staff with essential critical reflections about the specific pedagogical requirements for scientific training of professionals. In addition, it was leaned towards the theoretical and practical teaching of statistical methods, giving a central role to the linkage between the rational basis of methods with the conditions and limitations of use of these scientific instruments throughout research in social sciences.

This system was structured as a network, covering all the beneficiary countries. This

allowed teams of teaching staff to be mobilised, to constitute "common academic areas", through which the "horizontal" co-operation links could be created and strengthened between higher education institutions in the region. Finally, this training structure brings together the conditions for sustainable growth. In fact, it mobilises local lecturers through the autonomous operation of reference centres, which enables the training initiated in the courses under the programme to be continued through a continuing education structure.

The experiment developed revived the university extension tradition in the major South American universities, and allowed them to rediscover the social and scientific importance of this university activity. It showed in very specific terms, in an important field and on a vast scale, the most appropriate way of meeting the demands addressed to the university to update knowledge. Our work provided the organisational and pedagogical knowledge necessary to produce and disseminate university extension work in applied statistics. Furthermore, the interest aroused by this initiative has also inspired the hope that it will generate similar initiatives in other scientific fields.

Summing up, the experience with the PRESTA Programme has led to a rediscovery of the importance of high-level, non-academic training, capable of linking updated scientific and technical knowledge to the solution of the major social and economic issues on which the local scientific community is working on.

NOTES

1. It is difficult to summarise the main reasons for this widespread deficiency in the training of Latin American professionals, without mentioning the substantial differences in the situations found (and which are still found) in the different countries of the region. We are aware that the "local analysis" of these reasons has to take into account the obvious differences between, for example, the statistics higher education structures in Brazil or Colombia and those in Paraguay or Bolivia (just to mention extreme situations).
2. We invite the reader to consult the detailed presentation of each "Training Trainers" activity on the PRESTA Programme's web site. (http://www.ulb.ac.be/assoc/PRESTA).
3. This corresponds to 944 hours of theoretical and practical courses by teams of six or more teaching staff. This volume of hours of courses required about 5,400 lecturer/hours, with European colleagues providing 2,970 hours, and local lecturers providing the remaining 2,430 hours.
4. The PRESTA list has been administered, since 1995, by the Universidad de Concepción, Chile.
5. On the PRESTA web site, we give a detailed presentation of the content and specific arrangements for "local seminars". In the first stage, we held 2 local seminars and 1 thematic seminar, while in the fifth stage of the programme, 26 local seminars and 7 thematic seminars were organised. This illustrates the way in which the multiplier effect anticipated from the "Training Trainers" activity was spread.
6. This group also included, lesser numbers of agricultural engineers, veterinary surgeons, town planners, public health specialists, lawyers, engineers, mathematicians, computer scientists, surveyors, food industry specialists, etc.
7. It was unknown in Latin America for a training programme to involve such a wide spectrum of the whole scientific community in the countries of the region.
8. Consolidation courses took 6 to 8 four-hour sessions at which participants carried out a series of specifically prepared and guided tasks in small working groups with the support of a local lecturer. That facilitated the emergence of continuing training groups, among peers, where researchers could raise methodological questions from their own research work.
9. The contractual arrangements for the research explicitly refer to the theme of the study and the number of surveys that must be carried out. Therefore, the sample size and the very

notion of representativeness are established before and regardless of any operational definition of the purpose of the study.

10. These policies, and the basic options of the PRESTA Programme were debated for the first time at the International Conference "Experiments and Prospects for the Teaching of Statistics: Challenges for the 21st Century". This conference was held in November 1999 in Florianópolis, Brazil, in co-operation with the Universidade Federal de Santa Catarina and the International Association for Statistical Education.
11. Some universities have created a multi-disciplinary Committee for applied statistics; others have set up an Applied Statistics Laboratory; and yet other institutions have entrusted these functions to a part of the scientific personnel in their statistics departments.
12. This has enabled statisticians who are working on the production of statistics relating to education, health, the environment, agricultural production, employment and other economic information to take part in the training courses.
13. The First Latin American Applied Statistics Conference was organised by the Universidade Federal de São Carlos, in November 1999 in Sao Carlos, Brazil.

APPENDIX: PRESENTATIONS OF PRESTA PROGRAMME

Crivisqui E. (1997). *Conference on the PRESTA programme. Result and Perspectives from an Interuniversity Co-operation Experience in Latino America*. Université Libre de Bruxelles.

Crivisqui E. (1998). Escuela de Estadística, Facultad de Ciencias Económicas, Universidad Central de Ecuador, Quito. Ecuador.

Droesbeke J. J., & Crivisqui E. (1998). *Conference AOSC-98.* Budapest, Romaine.

Crivisqui E. (1998). Universidad de la Frontera, Temuco, Chile.

Droesbeke J. J., & Crivisqui E. (1999). Académie Royale des Sciences d'Outre-mer, Bruxelles, Belgium.

Crivisqui E. (1999). *Conferência International «Expêriencias y Perspectivas do Ensino da Estatística – Desafíos para o século XXI»*, Universidade Federal de Santa Catarina, Florianópolis, Brazil.

Crivisqui E. (1999). *Conferencia Internacional «Métodos estadísticos aplicados a los estudios medioambientales»*, Universidad Nacional de La Plata, Argentina.

Crivisqui E. (1999). Facultad de Ciencias Aplicadas a la Industria, Universidad Nacional de Cuyo, San Rafael, Mendoza, Argentina.

Crivisqui E. (1999). *Primeras Jornadas Latinoamericanas de Estadística Aplicada*, Universidade Federal de São Carlos, Brazil.

Crivisqui E. (2000). Facultad de Ciencias Sociales de la Universidad Nacional de Colombia, Bogotá, Colombia.

REFERENCES

Batista C., Hochztain E., & Aguirre, C. (1999). Experiências metodológicas no ensino de análise exploratória de dados na América Latina (Methodological experience in teaching exploratory data analysis in Latino America). In *Conferência Internacional: Experiências e Perspectivas do Ensino da Estatística.* Florianópolis, Brasil: PRESTA. http://www.inf.ufsc.br/cee/frame2.htm

Crivisqui, E. (1998). An experiment in the teaching of statistical methods using a videoconference network, *ISI Newsletter*, 22 (3), 66.

Crivisqui, E., & Droesbeke J. J. (1999). An experience of international co-operation: The PRESTA Program. In *Academic and Official Statistics Co-operation* (pp. 135-141). Luxembourg: Eurostat.

Droesbeke J. J., & Crivisqui E., (1999). Bilan d'une coopération quinquennale entre l'Europe et l'Amérique latine dans le domaine de la statistique appliqué. *Bulletin des Séances de*

l'Académie Royale des Sciences d'Outre-mer, 45 (3), 403-417.
Crivisqui, E, & Droesbeke, J.J. (2000). *Rapport d'évaluation de la cinquième phase et rapport final du programme quinquennal de coopération inter-universitaire PRESTA*. Brussels: PRESTA.

Eduardo Crivisqui
Laboratoire de Méthodologie du Traitement des Données,
Université Libre de Bruxelles, Bruxelles, Belgique
E-mail: ecrivisq@ulb.ac.be

Stephano Abruzzini
Commission Européenne DG Emploi - G/5
Rue de la Loi 200 - B-1049 Bruxelles Tél. +32 02 2994470
E-mail:stefano.abruzzini@cec.eu.int

Carlos Marcos Batista
Departamento de Ciencias Sociais, Universidade Federal de Goias,
Goiânia, Brasil
E-mail: carlosmb@fchf.ufg.br

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 295-310.

Printed in Granada, Spain.

DAVID OSPINA & JORGE E. ORTIZ

STATISTICAL RESEARCH AND CONSULTING IN UNIVERSITIES FROM DEVELOPING COUNTRIES: THE COLOMBIAN CASE

This paper focuses on describing the experience of statistical researchers and consultants of the Department of Mathematics and Statistics, faculty and students, and of researchers who are not statisticians, at the Universidad Nacional de Colombia. An analysis is made of the statistics programs of study including consulting activities, and of three surveys that were designed, distributed and analysed: one for the newly graduated and senior year students concerning their training in research and consulting during their studies; a second one for statistics staff at the department and a third for non-statistician researchers at the University who usually have to deal with statistical methods.

A study is made of the main problems in the formation and training of students as statistical researchers and consultants, as well as the difficulties of the experienced researchers and consultants. Recommendations to improve the actual situation are proposed.

1. INTRODUCTION

The Universidad Nacional de Colombia (UNC) is the only institution of higher education in Colombia which offers undergraduate and graduate programs (Specialisation, Master and Ph.D.) in statistics. The undergraduate program started in the second half of the 50's; the specialisation and master programs at the beginning of the 80's and the Ph.D. program will start this year. The undergraduate program has been modified on several occasions through the years in order to keep abreast of scientific and technological advances.

The latest and most important modifications occurred in 1995. The master program was reformed most recently two years ago. An international committee suggested by the ASA (American Statistical Association), made an evaluation of the program for the purpose of obtaining international accreditation.

The main conclusions of this evaluation are mentioned later but they have not been implemented completely. The specialisation program, an equivalent to the Master of Arts at the U.S. universities, has had relative success and has served as a bridge to get in touch with industry and some public and private agencies and institutions. The Ph.D. program is just starting, and therefore, no evaluation should be made of it.

In this paper the curricula of the statistical programs, with the exception of the Ph.D., are examined to determine their main weaknesses. To find out the opinion of the students (newly graduated and senior year statistical students), of the Statistics Department staff and of some non-statistician researchers and consultants of the University, three different surveys were taken and analysed. Through the analysis of the

curricula and these surveys, the main reasons that usually prevent the statistician graduates from being successful in consulting and research were detected. As a consequence, important changes in the structure of the programs, modifications to several subjects' syllabi and a drastic change of the consulting process offered by the department are suggested.

The actual curricula are studied in the first part of the paper. This also includes a description of the statistical demand at the department and how it is satisfied. A second section is devoted to analysing the data provided by the surveys. There is also a section of general conclusions on the limitations of the curricula and a last one giving some recommendations to improve the actual situation.

2. STATISTICS PROGRAMS

2.1. UNDERGRADUATE PROGRAM (COMITÉ PROGRAMAS CURRICULARES DE ESTADÍSTICA, 1999)

The undergraduate program in statistics at the UNC is the only one in Colombia, which has been functioning without interruption since its creation, on 17th February, 1956. The number of graduates has totalled 400 in 40 years. Programs in statistics have started at some other universities but they have failed to survive because of low student population. This has mainly been due to the poor tradition of statistical studies. Before the 80's the high school students did not have a clear idea what statistics programs were about.

The people responsible for the statistical work have been usually trained at the same institutions through seminars and short courses given by experts. A few of them are sent abroad to obtain a more formal education. Eventually, the UNC contributes to this training but this is not the rule.

In contrast to other developing countries, there is not an important association caring for statistical education. In China, for example, according to Wei (2001), there is a National Statistical Education Association, which, among other activities, is in charge of the official statisticians' training. Different from Colombia, they have a well-programmed national statistics officers' training program. Colombia, with a population over 40 million and more than 100 colleges and universities is in need of a strong statistics association and an improved co-ordination among the institutions that have to do with statistics work, whether public or private.

The program at UNC has overcome several crises, for the same reason, but now seems stronger than ever. At the beginning it was a six-semester program (on the technological level). In 1970 a new plan of study of eight semesters was introduced and finally, in 1979, the program became a professional one with a duration of ten semesters.

The curriculum has had major and minor changes in an effort to maintain itself on a par with the scientific and technological advancements. The goals of the program are to form professionals who contribute to the development of statistics as a discipline and who possess abilities and capacities to apply statistical methods to the analysis and solution of problems and to formulate models in technical and scientific areas.

This goal has not been completely fulfilled because opportunities for interaction with scientists and professionals, leading to research and consulting, have not been as

successful as has been desired.

The main reason for that must be that the program has been traditionally dominated by theoretical knowledge. Only during the past years (second half of the nineties), has a Consulting course become part of the curriculum. Furthermore, the office of extension of the department has allowed some of the students to get in touch with real data when they collaborate with the projects under the supervision of faculty members. A description and analysis of the actual program, trying to detect its weaknesses, follows.

Basic phase

The basic phase of study corresponds to the first four semesters of the curriculum and is common with the mathematics program. This phase provides the student with the fundamental elements of the areas of mathematics and statistics, and tries to give a general vision of the profession. The subjects of this cycle are: Foundations of Mathematics, Calculus I, II, and III, Differential Equations, Elementary Geometry, Linear Algebra I, Linear Algebra II, Logic of Programming, Linear Programming, Professional Introduction, General Statistics, Probability and Statistical Inference.

Basic statistics phase

The phase devoted to fundamental knowledge of statistics extends from the fifth to the eighth semester and consists of nine subjects that constitute the culmination of the disciplinary nucleus in the professional formation of the students. The acquired formation establishes the difference from other disciplines and brings the student to understand the different theoretic and practical orientations in statistics. The following subjects are included in this phase: Statistical Theory, Sampling Theory, Statistical Methodology, Statistical Computing, Linear Models, Multivariate Analysis, Experimental Design I, Non-parametric Statistics and Time Series I.

Statistics specialisation phase

This phase constitutes the flexible part of the curriculum. It usually starts in the sixth or seventh semester and goes to the ninth. The student must choose an area of specialisation and take six subjects corresponding to that area. The objective is to study the theory and statistical methods related to the area in depth. The areas are Data Analysis, Mathematical Statistics, Time Series, Experimental Statistics and Statistics in Populations.

Other requirements

In addition to all the subjects above described, the student must take and approve a comprehensive exam of scientific texts in a foreign language (English, French or German), before the fifth semester. Two humanities courses and two more subjects, called “context courses”, related to the analysis of topics of general academic or national interest are also required.

The Consulting course is given during the last semester and it is expected that, during this course, the student will analyse real data from problems where statistical methodology and theory are needed. As a culmination of his studies the student may choose between writing a monograph or doing a semester of practicals with some interested industry or institution.

Discussion

According to the document prepared by the Statistics Curriculum Committee the new statistical demands show the existence of an important difference between what is expected from the statistician and what is given by his university formation. These demands are a product of the global human evolution, especially in relation to technological development, socio-economic evolution and the new politics of international openness and globalisation.

The technological evolution implies a new conception of information systems, a change in the way data are produced and collected and also a change in their treatment. It is important to know and use modern information techniques, to be able to analyse and interpret the results from statistical analysis more than simply to use the tools, and to formalise the complete process in experimental research, among others.

Opinion and studies from specialists help in making decisions about an essential modification of the program of study. Jammalamadeka (1994) holds that there must be equilibrium among mathematics, statistics, computation and areas of application (biology, economy, medicine, etc.).

This means less reliance on mathematics (some of the mathematics courses could be modified and a couple of them suppressed), more and better use of computers (which also demands the purchase of more and better computers) and more interaction with substantive disciplines. The program itself apparently fulfils the needs for statisticians in industry. However, there are some applied areas that have not been considered. Among them, psychometrics and biometrics seem to be the most important. Kettenring (1995) suggests that:

> *"Basic building blocks of any program should continue to be well-rounded statistical knowledge including subjects such as data analysis, statistical computing, sampling, linear models, experimental design, time series, multivariate analysis, and so forth"* (p. 3).

The program covers all of these topics as basic courses with the exception of Data Analysis, which is a specialised subject. The problem is not the number of courses of the program because they are numerous and cover the main areas, possibly more extensive than most of the North American and European statistical programs. According to Scheaffer (1991), the important issue is that the application of statistics courses to real data has not been sufficient.

> *"Engaging in real research provides students with a feeling of ownership of the data and the excitement of discovery that is absent from textbook exercises." "The best way to know statistics is not only to teach but to be a consultant in it"(*Brook, 1994, p. 257).

The actual contact with real data often occurs during the last two semesters, mainly through the Consulting office. Schuenemayer (1991) also refers to this situation when he states that, first, statistics refers to data and, second, every statistician is a consultant. Therefore, students' training for consulting must start earlier in the program.

Finally, in an article presented at the 1993 Joint Statistical Meetings in San Francisco, Garfield (1995) mentions some important aspects needed in statistical education. They are teamwork and collaboration, communication skills (oral and written), solution of real problems with real data, internships and real-world experiences in analysing data. All the recommendations should be given keeping in mind what lies ahead. Moore (1997) warns about the changes due to the advance of technology:

> *"The quantization of society is in its turn driven by the implacable advance of technology. Changes in computing, communications, and multimedia come so rapidly that comments in a printed journal are out of data before publication. Technology changes how we teach as well as creating demands for teaching new content. My thesis is that the most effective learning takes place when content (what we want students to learn), pedagogy (what we do to help them learn), and technology reinforce each other in a balanced manner.*" (Moore, 1997, p. 124)

Essentially reforms of the curriculum are important but more than that there must be a commitment among staff members concerning the changes needed in the teaching and the contents of the different subjects.

To summarise, the program must be substantially modified in the following aspects: A reform of some of the first mathematics courses is needed, among them, Elementary Geometry, Professional Introduction, Linear Algebra (I and II) and Calculus (including Differential Equations). The main geometric concepts can be included in one of the Linear Algebra courses.

The four calculus courses can be reduced to three including the main concepts of differentiation, integration, series and differential equations. The number of courses should be diminished, but the level should remain the same.

> "*It is nonetheless true that statistics makes heavy and essential use of mathematics, that advanced training in statistics requires considerable exposure to mathematics, and that elaborate mathematical theories underlie some part of statistics.*" (Moore, 1997, p. 125).

The Professional Introduction course should be moved to a more advanced semester, maybe the fifth or the sixth, when the students can take better advantage of it. Garfield and Gal (1999) give some important challenges for statistics teachers (or professors). Some of these challenges are equally valid for the statistics students and should be taken into account as part of the Professional Introduction course in case the curriculum is modified. These are:

1. Understand the purpose and logic of statistical investigations;
2. Understand the process of statistical investigations;
3. Master important procedural skills;
4. Develop interpretative skills and statistical literacy;
5. Develop ability to communicate statistically.

There must be an emphasis on communication skills. This could be attained by demanding well-written reports from the students in those courses that involve home assignments or through a specific course. What Glencross and Mji (2001) say about the need for social researchers to learn how to write research reports, is also valid for statistics students because they will have to face this issue very often once they finish their studies. While statistical and methodological knowledge is certainly important, it is not effective if the consultant cannot adequately communicate it or cannot manage a consulting session properly" (Belli, 2001).

Better education and training in computational statistics and information systems in general is required. Familiarity with computers is an essential requirement for each statistician. One of the most important challenges in the actual world has to do with the

handling of large amounts of data and this is possible only with good computational training and knowledge.

Moore (1997) is right when he states that heavy use of computing technology is essential for realistic learning of practical statistics. Additional benefits are also obtained. Advances in computing have made areas like re-sampling and non-parametric regression of practical relevance and have opened up many application areas for research in data analysis and statistics (Cameron, 1997).

One or two more courses on information systems or database handling are necessary. The main problem, also shared by Chinese programs (Wei, 2001), to fulfil this requirement is that still some of the main applied courses are taught separately from computer courses.

The specialisation cycle has not given the expected results. The general feeling is that a more integral formation in basic statistics areas is preferable to an early specialisation in one of them that could be obtained during graduate studies.

The specialisation in one particular area for the students, often with no real experience at all, prevents them from taking some other courses in different statistical areas that would allow them to qualify better in the statistical job market. Besides, the reduced number of teachers sometimes makes it difficult to offer some of the specialised courses, which are demanded by the students. An increase in contact with the external world is urgent:

> *"Statistical training should be truly holistic because that is the way statisticians are expected to operate in the real world..... To reach this state in university statistics education, closer working ties with industry would be worthwhile."* (Kettenring, 1995, p. 3).

> "*It is recommended that university courses on quantitative research include a section on the use of official statistics and the pros and cons of using large micro data sets, e.g. using hierarchical data sets, dealing with missing values and non-response, with imputed and edited data, appropriate use of weights, analysis of complex design, and managing disclosure control*" (McDonald, 2001, p. 127).

> "*…with the large amount of data available via the Internet, there can be no excuse for not getting students to work on real data*" (Jolliffe, 2001, p. 357).

The university has a shortage in providing these tools, not fulfilling the challenge of both, globalisation and technology. The seminar, the consulting course and the monograph must be reoriented to get all the benefit that can be obtained from them. A recent graduate should be capable of participating in interdisciplinary groups.

It seems urgent to create what is called '*Statistics* Lab' in many universities in other countries where several members of the Statistics Department could share the whole responsibility of the consulting process. The problem that newly graduated statisticians have when they leave the university and have to cope with the reality has been already mentioned:

> *"The shock for newly graduated statisticians is to find that many of the practical techniques used on a weekly basis seem to start near to where their university courses ended. Moreover the customer appears so knowledgeable. The trust can be built through the teamwork."* (Pike, 1994, p. 358).

As Godino et al. (2001) recommends consultancy courses should have an essentially practical orientation and be based on the philosophy of workshops and seminars, with the support of a network of centres where practices would be carried out. Smith (1994) considers that a consulting course must be based on three items:

(1) Applied statistical projects;
(2) Rreport writing; and
(3) Classes given by experienced students.

The first one has been the raw material while the second one has been neglected and only during the last semester was considered seriously. No class by experienced students has been considered but some members of the staff have been eventually invited to give some lectures. It is important for the three items to be satisfied in the regular course.

A semester devoted to practical training must be guaranteed through regular agreements with industry and private and public institutions. This requirement, an alternative to the monograph has failed in the past and is being implemented once again. This seems to be a difficult task due, in part, to the poor statistical knowledge that people in charge of industries, factories and institutes have. Statistical work has been usually done by economists, business administrators or accountants, who often do a poor job. The program has come up short in this sense. As Scheaffer (1997) recommends, when talking about new pedagogy and content in statistics education the learning of statistics must move from passive to active.

Students' consulting activities

A statistician must be able to effectively communicate with researchers and practitioners and be conversant in their functional areas. Consulting skills should therefore be an important aspect of statistical training (Belli, 2001). In Colombia, like in Spain, there is neither a culture that favours statistical consultancy, nor a deserved scientific or economical recognition of this work (Godino et al., 2001).

The Statistics Department has a consulting service offered to the University community through their students. The two main reasons to keep the consulting service are the same presented by Belli (2001) in her survey to 106 USA departments, that is, to provide statistical assistance to faculty and to serve as a training ground for consulting students.

The service charges no fees and is usually provided to students from other disciplines, especially for theses or final monographs. In a few cases researchers from other departments also ask for assistance of this kind. The consulting activities are part of the undergraduate program (registered as a consulting course to be taken in the ninth semester) and are required by all students.

Every student should stay in the consulting office during a two-hour period twice a week. Those interested in the service must register in advance for a 20 minutes session each time. They must take with them all the pertinent documentation, including the research proposal, the most important references on the theme, the measurement instruments, and, eventually, the data collected. The initial discussion is focused on the description of the problem and the methodological aspects of the task, especially on the objectives.

As a general rule, there is no discussion of the statistical analysis strategies during the first session. These will be discussed when the problem has been sufficiently clarified. There are as many sessions as required. In some cases methodology and data collection are followed closely; in others discussion of related concepts seems to be more important in order to adequately orient the statistical task. An experienced member of the department staff, responsible for the course, supervises the process.

As part of the training, every week there is a session among the members of the consulting group where the problems and solutions for the different situations are discussed. If there are difficulties with some specific areas, consultation with specialised teachers in the respective topics is proposed.

When the student gets the data collected, he/she analyses it and writes an initial report for the customers. This report must contain a reading of the obtained data and some recommendations on the tables and graphs presentations that the customers must complete. Very often the data have already been collected using instruments of dubious quality or inadequately applied and there may also be doubt cast on the observation methodology (experimental or from sample). From the students' point of view the experience is surprising. Their first impression is that they have not been prepared to face this kind of problem.

Experimental design and sampling are the main areas where people consult. This means that most of the student activities as consultants are referred to specific problems on estimation and hypothesis testing.

The absence of group work during the program of studies, with a few exceptions as well as the limited interdisciplinary work, constitutes a major weakness of the curriculum. This prevents the students from participating as colleagues from the beginning of the projects, sharing responsibilities in some of the research phases.

2.2. GRADUATE PROGRAMS

2.2.1. SPECIALISATION

This program has been designed for professionals in different areas involved mainly with the implementation of statistical methods. It is a three-semester program for full-time students, which has run for almost two decades. However, since most of the students have to work, the classes have been offered at the weekend making it into a part-time program, which lasts five semesters instead of three.

The program has been taken to some provincial universities with good results, especially for professionals with a major in mathematics. During the last few years participation of students from other disciplines has increased, a situation that has helped the statistics program because it has become well known to many people and institutions, producing a larger number of applications.

North American or European universities do not offer this type of program. However, in the United States a similar graduate program called Master of Arts in which the candidate does not have to write a thesis but rather a monograph is offered. The main difference is that at the UNC the courses are not the same as those offered in the masters program, whereas in the United States they usually are. The curriculum includes both basic and special subjects as well as a seminar, whose description follows:

Basic phase

The basic subjects provide the students with a general formation in statistics and mathematics, especially those with a deficient background in these areas. The subjects are Basic Probability Theory, Linear Algebra, Basic Mathematics, Statistical Methods and Regression Methods.

Special phase

The special subjects are related to specific topics in a given statistical area. Usually the students choose those most related to their work experience. The subjects are Econometrics, Sampling, Qualitative Data Analysis, Non-Parametric Statistics, Experimental Design, Time Series, Multivariate Analysis, Data Analysis and Special Topics in Statistics.

Seminar and other requirements

The Seminar is a course designed to present and discuss several themes and experiences at different levels (applied, theoretic or educational) related to the statistical interests of the students. The idea is to familiarise the students with the application of statistical methods to specific situations.

Besides the above courses, the students must take and approve a comprehensive test in a foreign language (English, German or French) before the first year. As a final requirement they must elaborate, under a professor's guide, and defend, a monograph, generally related to their jobs.

Discussion

The program in general is working well. However there are some important aspects missing. The first one has to do with the relatively poor emphasis in computation techniques, especially large data base handling. This is not a difficult problem to solve. Nevertheless, the statistics professor should be very careful when teaching these topics because computer programs in some situations give the false sense that the statistical professional assistance is not needed (Gowler and Diggle, 1987). Shimada (2001) has already pointed it out when he says:

> "*…there are many statistical packages available that make it easy to perform stochastic procedures. Therefore, today's students and researchers may think they can handle their data processing needs, and obtain stochastic results simply by clicking a PC button. However without being aware of it, they can make mistakes and treat their data incorrectly*". (Shimada, 2001, p. 127).

In the same article Shimada presents a real situation where it is very easy to make an important mistake. Jolliffe (2001) also refers to this situation when she suggest that the statistician has to some extent been replaced by the computer, so it is particularly important that, researchers are made aware of the dangers of misusing statistical packages and of the errors inherent in some routines.

Probably this is one of the reasons why an increasing number of people are using these tools. Muller (1989) had already pointed out how statistical computing has started to become a common practice. The second aspect is the lack of consulting. Even though most of the students work in an industry or at a private or public institution, they are not familiar with statistical consulting practice. A course similar to the one in the

undergraduate program should be established.

2.2.2. MASTER OF SCIENCE

The Master of Science (MSc) program is directed to professionals with intermediate knowledge in statistics, probability and mathematics (including linear algebra). It was created officially in October 1979 and has been reformed on various occasions since then.

The most significant is possibly the last reform in 1998. In twenty years less than forty students have been able to get their degree. According to the 1998 reform, the aim of the program is to form statistical researchers able to formulate a statistics problem, to modify or to innovate existing solutions and to use the new results in the analysis and interpretation of the statistical data pertaining to the problem.

The curriculum has been designed to last four semesters for a full-time student. It consists of four basic courses and at least two special ones, and three seminars destined to orientate the students properly in their area of interest.

Basic phase

The basic subjects provide the students with a solid formation in statistics in order to enable them to face a statistics problem with the required academic seriousness. The subjects are Advanced Linear Algebra, Linear Models, Mathematical Statistics and Statistical Methods.

Advanced phase

This phase includes the special subjects, related directly to the specific area to which the student's research problem belongs, as well as the seminars. The candidate must write a thesis referred to his research problem as culmination of the program. The originality of the work is considered to be proof of his ability to research in statistics.

International evaluation

An Accreditation Team appointed by Dr. Kettering, then President of the ASA reviewed this Master program. This Team was formed by: Professor Alicia Carriquiry (Iowa State University), Professor Wayne Fuller (Iowa State University) and Professor Edward George (University of Texas at Austin).

The Committee visited the Department of Mathematics and Statistics of the Universidad Nacional de Colombia (UNC) in August 1998. The Committee found that the Master of Science program meets requirements comparable to those satisfied by programs in countries such as the United States and Canada.

However, the following comments and recommendations were made (Carriquiry, Fuller and George, 1998):

- The Department has an well-educated, diverse faculty, which displays strength in both theoretical and methodological areas. The research component in the Department would be improved if faculty members have the opportunity to obtain advanced degrees abroad.
- The student body, while enthusiastic and competent, could be larger given the number of faculty in the Department. One way to increase the number of graduate students in Statistics is to recruit students from other disciplines. The benefits of a

student body with different educational backgrounds are not fully realised.

- Admission requirements contribute to the lack of students from disciplines other than Statistics. Rather than requiring that potential students pass an admissions exam, admissions should be based on more general criteria. Instead, students should be required to pass a qualifying exam approximately one year into the program.
- Graduation rates appear to be unacceptably low. This is likely due to the emphasis put on the thesis component of the program, as most students seem to successfully complete the course component. The new program, to begin this semester, must include at least an additional semester of course work and de-emphasise the thesis component.
- Research problems undertaken by the students seem to be dominated by theory. Good methodological work is challenging and so is good research. Most students work outside the university and have real problems to work on. Furthermore, the Extension Service in the Department is a source of interesting problems. There have been 34 graduate students who have obtained their Master's degree since its creation. It is interesting to find out that only 5 among all the theses were applied. The theoretical approach has been the general rule, especially in the areas of Non-parametric Statistics (5), Time Series(6) and Experimental Design and Linear Models (5). This has been partially the reason for such low graduation rates.
- The use of computers should be emphasised in both course work and research. To facilitate the inclusion of computing as an integral part of modern statistics, the computational facilities in the Department need to be extended and modernised.
- Even though the UNC has another graduate program (Specialisation) with less theoretical requirements and more emphasis on methods, the recommendations remain appropriate for a masters program designed to serve both students planning to proceed to an advanced degree and those who will function as professional statisticians outside the University.

Additional discussion

Completing the recommendations given by the Evaluation Committee, training in consulting is also needed, especially for those students interested more in statistical methods than in statistical theory. Graduate students should be encouraged to participate in interdisciplinary projects and work more closely with other professionals. Hunter (1981) pointed out that more than helpers or leaders, statisticians must play the role of colleagues:

> *"The two roles of helper and leader are characterised by one-way communication, the helper receives, the leader transmits. The role of the colleague, on the other hand, necessarily involves two-way communication, and hence makes possible what, for me, is the joy of being a statistician: working on and learning about many different problems, and sharing with clients the excitement of solving those problems*"(Hunter, 1981, pg. 73).

3. SURVEYS ANALYSIS

Three different surveys were designed to learn the opinion of the staff of the Department of Mathematics and Statistics, and non-statistician researchers who have become familiar with statistics methodology through their projects. The analysis of these

surveys is a second source of information concerning the statistics formation given by the University, the main problems that people as users of statistical techniques have to face, and the statistical areas of interest in everyday work

3.1. STUDENTS' SURVEY

A survey of 11 newly graduated and senior year students was analysed. 41% of the statistical applications used corresponded to the area of experimental design. Almost all the rest of them were on non-parametric statistics and sampling designs (approximately 25% for each area).

The use of other statistical techniques was much less frequent and in these cases no big difficulties were encountered in the solution of the problems. This means that most of the consulting tasks carried out by the students was concentrated on hypothesis testing and estimation problems.

There were pathologic cases where the student was asked to analyse some inconsistent and arbitrary data, collected with no scientific methodology at all. At the beginning the student required help from a faculty member, especially to define the analysis strategy to identify and carry out the methodological conditions of the designs. As the student gains experience, he/she also gains criterion independence to solve the cases. He/she learns to identify the errors in the design and in the information, provided by the consultant users, and to propose alternative solutions.

What is clear from the students' survey is that to teach statistical consulting skills seems to be a necessity. Belli (2000) mentions two essential aspects to be taken into account, both related to the experience which must be based on real data and is obtained only through practice.

3.2. RESEARCHERS' SURVEY

Fifteen researchers from different disciplines at the university (psychology, meteorology, chemistry, physics, engineering, medicine, anthropology and biology), familiar with statistical methodology and methods were surveyed to learn their opinion with respect to three main topics. These are: Statistical areas more often used, general preparation in statistical methodology given by the University in the different programs, and the main difficulties related to statistics encountered in their research work. All the researchers have graduate studies at masters or Ph.D. levels.

The utilisation of statistical techniques referred to the different areas is quite diverse. Most of the researchers use two or more tools of this type. However, the most frequently used are Linear Models and Experimental Design. Despite the lack of specialised software, Non-parametric techniques, Sampling and Multivariate Analysis are also required.

The general preparation in statistical methodology given by the university is considered satisfactory by most of the interviewed professionals. This means that the academic level of the statistical courses is good, allowing the students, by themselves or with a consultant's aid, to develop successfully the statistical part of their research work.

With respect to the statistical difficulties, it seems that the researchers usually look for the statisticians' help once the research has advanced, especially when the work demands the definition of analysis strategies or the use of statistical computation. Very often the difficulties persist to the end of the research. This situation is similar to that

one presented by Belli (2001) where she states that most clients look for help once the data have been collected. The ideal situation seems to occur when the consultants are asked to participate from the beginning rather than when asked to salvage an experiment:

> *"In any investigation the planning, what data are collected, and the analysis depend on the objectives of the study, and the statistician needs to be aware of what these are, and to be involved as a member of the research team from the start"* (Jolliffe, 2001, p. 364).

3.3. DEPARTMENT OF MATHEMATICS AND STATISTICS' STAFF SURVEY

Most of the department of mathematics and statistics' members interested in research and/or consulting were surveyed. The main conclusions are drawn from the twelve questionnaires that were returned. More than half have to deal with consulting rather than research itself. Experimental design is the most consulted area, with almost 40% of the requirements.

The researchers, who ask for help, have a proper conceptualisation of the problem and correct handling of methodological aspects. The difficulties arise with data analysis and interpretation, even if the responsible groups are, supposedly, well prepared. One third of the studies were described as very limited in analytic strategies, computing technique usage and interpretation of results.

Designs of surveys and sampling have less demand, however they involve more difficulties throughout the different research phases. These have to do with the definition and operational aspects of the variables, analytic strategies, management and choice of the methodology, and processing and interpretation of final results. Almost all these problems are present in more than half of the consulting cases.

Most of the errors described by Godino et al.(2001), related to doctoral theses on mathematical education, are also present in the research work carried out by professionals in our country. Those associated to the experimental design, variance analysis and sampling techniques are most frequent.

It is also important to mention that, in general, difficulties appear more often with social type research than with research devoted to technical or pure scientific problems.

4. CONCLUSIONS

The students are exposed to a lot of theoretical concepts but they are not involved with real data as often as desirable. Additionally, there is no culture of teamwork. Only in a few courses are students asked to carry out statistical tasks in a group. Besides, there are not sufficient interdisciplinary activities that allow students to share responsibilities concerning the thesis or monograph of their academic programs, under advisors' supervision. For all of the above reasons the students usually feel insecure when they have to face a real problem during their senior year or as recent graduate.

On the other hand, the government education investment on Computing hardware and software has not been sufficient. Usually the acquisition of new machines and programs takes longer than expected. This often makes statistical applications go behind statistical theory.

Some of the existing limitations make some of the best students feel frustrated because they are not motivated towards research. This is especially true for those students who want to continue on a graduate program

Finally, many of the users of consultant services do not have statistical "culture". This leads to consider the statisticians as "helpers" for the statistical analysis rather than as colleagues.

5. RECOMMENDATIONS

The first thing to do is to create a Statistics Lab to provide a sufficient amount of real data for all the applied courses, not only for the undergraduate but also for the graduate programs. This requires to improve the computing hardware and software.

It is necessary for the Department of Mathematics and Statistics to establish closer contacts not only with other departments within the University but with off campus institutions and industries. It is important to undertake joint research and publishing. It is only through this kind of work that the statistical profession gets credit (Ospina, 1991). This task has already begun with the former chairman of the statistics program and now the new chairman has given it a fresh impulse.

This will increase the interdisciplinary work and will give the opportunity to work with real data. Additionally, some research projects could be developed, thus fulfilling, some of the needs of undergraduate and specially, graduate students. To reach this, a serious commitment of some faculty members towards the students is needed in order to form researchers.

Although there is an increasing demand in statistics, it is satisfied usually in a somewhat poor way, by some commercial rather than professional firms. It is important, then, to make the information directed to all the services provided by the Department available to as many people as possible. This will allow the consulting service to strengthen and provide a host of important data with which to work. Notwithstanding, a policy to collect payment for this service should be established.

The consulting service must be offered to more people outside the University. Also, a cycle of seminars or lectures on statistical methodology should be offered to non-statistician professionals, those interested in research, to prevent these people from making some well known and serious mistakes (to collect data in an improper way, to ask for statistical conclusions when the statistical analysis has been wrong, etc.). The untimely statistical assessment must be avoided. Otherwise, there is a risk of considering the statistician as an "expert to correct methodological mistakes", rather than a professional in the discipline.

REFERENCES

Belli, G. (2001). The teaching-learning process in university statistical consulting labs in the United States. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 325-338). Granada: International Association for Statistical Education and International Statistical Institute.

Brook R. J. (1994). Questions to ask a statistical consultant. In *Proceedings of the Fourth International Conference on Teaching Statistics* (Vol.1, pp. 257-264). Rabat: The National Organising Committee.

Cameron, M. (1997). Discussion: Current influences of computing on statistics. *International Statistical Review,* 65(3), 279-280.

Carriquiry, A., Fuller, W., & George, E. (1998). *The graduate program in statistics at the Universidad Nacional de Colombia: report of the accreditation team.* Unpublished document. Bogotá: Departamento de Estadística, Universidad Nacional de Colombia.

Comité de Programas Curriculares de Estadística (1999). *Propuesta de reforma de la carrera de Estadística* [A proposal to reform the graduate program in Statistics]. Unpublished document. Departamento de Estadística, Universidad Nacional de Colombia.

Garfield, J. B. (1995). Respondent, special section on CATS symposium on modern interdisciplinary university statistics education. *The American Statistician,* 49(1), 18-20.

Garfield, J. B., & Gal, I. (1999). Assessment and statistics education: current challenges and directions. *International Statistical Review*, 67(1), 1-48.

Glencross, M., & Mji, A. (2001). The role of a research centre in the training of social science researchers In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 245-257). Granada: International Association for Statistical Education and International Statistical Institute.

Godino, J. D., Batanero, C., & Jaimez, R. G. (2001). The statistical consultancy workshop as a pedagogical tool. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 339-353). Granada: International Association for Statistical Education and International Statistical Institute.

Gower, J. C., & Diggle, P (1987). *Reporte de un Comité de ISI, 49 Session, Florencia, Italia, 1993* [Report from an ISI 49 Session Committee]. México: Instituto Nacional de Estadística, Geografía e Informática.

Hunter, W. G. (1981). The practice of statistics: The real world is an idea whose time has come. *The American Statistician*, 35(2), 72-76.

Jammalamadaka, J. R. (1994). Future statistical training - some observations. In *Proceedings of the Fourth International Conference on Teaching Statistics* (vol.1, pp. 202-203). Rabat: The National Organising Committee.

Jolliffe, F. (2001). Learning from experience. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 355-370). Granada: International Association for Statistical Education and International Statistical Institute.

Kettenring, J. R. (1995). What industry needs. *The American Statistician,* 49(1), 3.

McDonald, S. (2001). Practical and educational problems in sharing official microdata with researchers. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 119-128). Granada: International Association for Statistical Education and International Statistical Institute.

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review,* 65(2), 123-137.

Muller, M. E. (1989). International relations and the ASA. *Amstat News*, 1989-2, 19-24.

Ospina, D. (1991). Teaching statistics in Colombia: Problems and potential solutions. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (Vol. 2, pp.115). Dunedin, New Zealand: International Statistical Institute.

Pike, D. J. (1994). Statistics in practice: Why is reality so different from the expectation? In *Proceedings of the Fourth International Conference on Teaching Statistics,* (Vol. 2, pp. 352-359). Rabat, Morocco: The National Organising Committee.

Scheaffer, R. L. (1991). Training students for statistical consulting. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics.* (Vol. 2, pp. 429-430). Dunedin: New Zealand: International Statistical Institute.

Scheaffer, R. L. (1997). Discussion, *International Statistical Review*, 65 (2), 157-158.

Schuenemeyer, J. H. (1991). Training statisticians to be consultants. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (Vol. 2, pp. 439-444). Dunedin: New Zealand: International Statistical Institute.

Shimada, T. (2001). Precaution against errors in using stochastic software. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 129-137). Granada: International Association for Statistical Education and International Statistical Institute.

Smith, J. T. (1994). Can statistical consulting be taught?. In *Proceedings of the Fourth International Conference on Teaching Statistics*. (Vol. 2, pp. 403-410). Rabat, Morocco: The National Organising Committee.

Wei, Y. (2001). The training of researchers in the use of statistics in China. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 311-317). Granada: International Association for Statistical Education and International Statistical Institute.

David Ospina and Jorge E. Ortiz
Department of Mathematics and Statistics,
Universidad Nacional de Colombia
Apartado Postal 75230,
Santa Fe de Bogotá, Colombia
E-mail: dospina@matematicas.unal.edu.co
E-mail: jortiz@matematicas.unal.edu.co

Batanero, C. (Ed..), Training Researchers in the Use of Statistics, 311-317.

YUAN WEI

THE TRAINING OF RESEARCHERS IN THE USE OF STATISTICS IN CHINA

The responsibility of training researchers in the use of statistics in China belongs to colleges, universities and research institutes. There is a National Statistical Education Association. Under the Association, the Higher Education Branch is an organisation of colleges, universities and research institutes who have a major in statistics or a statistics faculty. Since China has a population of 1.25 billion and more than 100 thousand official statisticians in the whole country, statistics training is a huge task.

There are degree and non-degree training programs. In the degree program, there are undergraduate programs (colleges and universities) and graduate programs (colleges, universities and research institutes). In the non-degree training, different training programs have been used.

Statistical methods are widely used in almost all the fields. The most important application areas are: official statistical work including sampling survey and data processing, micro-economic analysis, management and quality improvement, medical application, agriculture and industry experiment, etc. Most researchers in the above fields need to be trained. Many patterns have been used in training. Class teaching, group discussion, field training, TV, broadcasting programs and Internet are the main patterns.

1.BACKGROUND

China is in the stage of transferring from planning economy to market economy. During the planning economy period 20 years back, China had a huge employee system of making plans from township, county, prefecture, city, province to central government. The government bodies in each level had their own planning commissions to arrange development plans. Since the governments in each level have to make their plan on the basis of the local statistical data, they established statistical bureaus to carry out official statistical works.

Right now, the whole figure of employees is over 2 million, among them about 80 thousands are full time employees. Therefore, the training of official statisticians is the biggest task not only in the past planning economy period but also in the current transformation period for the Chinese government, the Chinese Statistical Society as well as for the Chinese statistical educators.

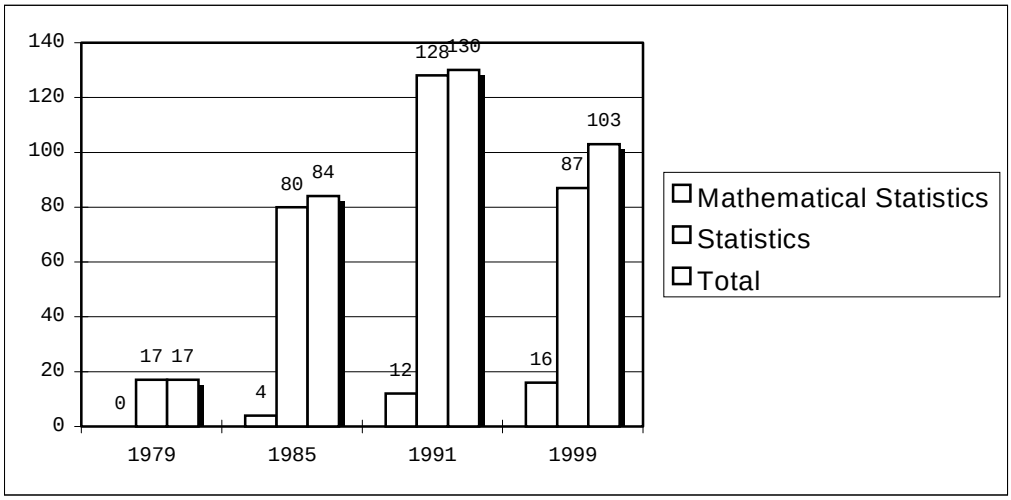
In the future, many other kinds of professional statisticians, such as bio-statisticians, medical statisticians, actuaries, statistical consultants, etc., will be needed. Therefore, the training programs will be more and more diversified. Because of this unique characteristic, I will discuss mainly the training of official statisticians in my report.

2.THE ORGANIZATIONS AND THEIR RESPONSIBILITY

Chinese universities, colleges, research institutes and many statistical training centres are responsible for training the researchers. Among 1000 higher education institutions, there are about 100 universities and colleges which have the statistical programs. That means they either established the statistics department or the statistics major. Since the official statistical system had a big demand in the past, most universities and colleges offered an economic statistics major.

With the transformation from planning economy to market economy, the statistical bureaus have had some reforms to meet this change and adopted more and more sampling surveys to substitute the overall reporting system. Therefore the statistical bureaus not only need fewer new employees than before but also need fewer graduates from economic statistics major (see figure 1). Now the higher education institutions are facing the challenge of this big change and adjusting their curricula. The number of universities and colleges with economic statistics major will decrease and the number with mathematical statistics (that is statistics) will increase. The official statistics bureaus in different levels have a big task to train their employees.

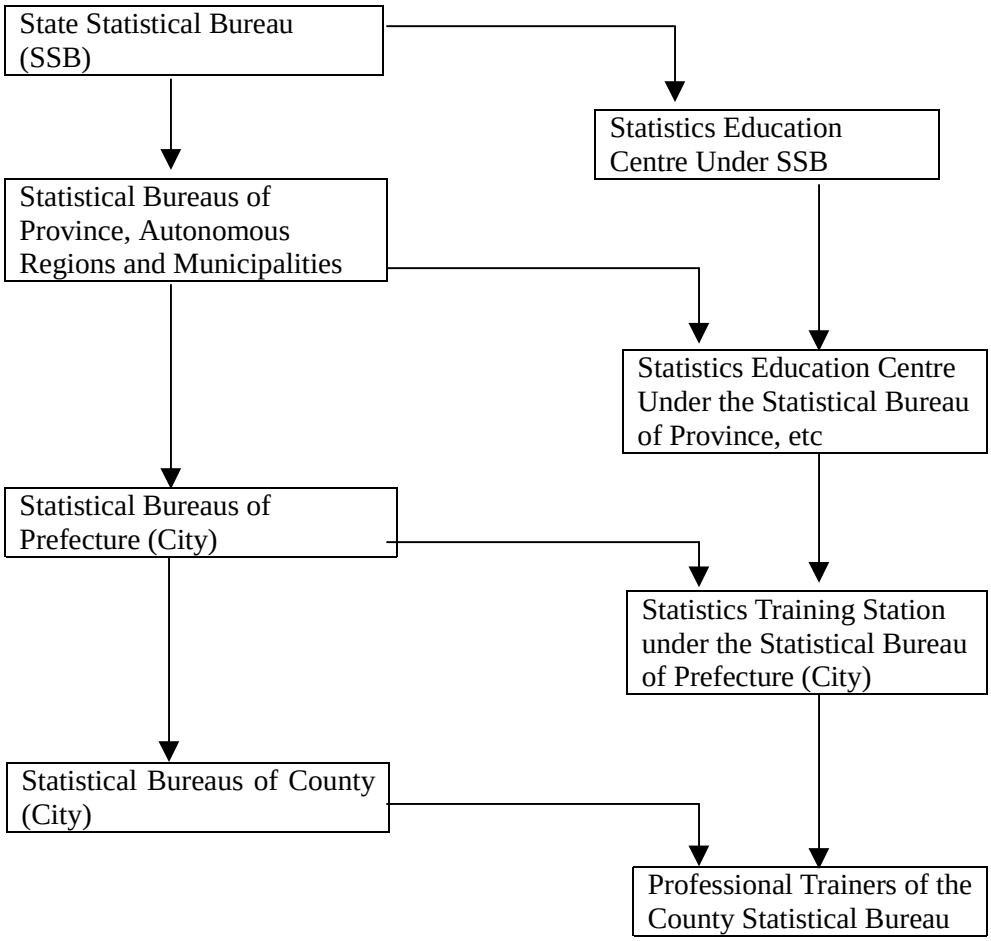
Figure 1. Frequency of Universities and Colleges with Statistics Major



In order to improve the education quality and to raise the training level, the universities, colleges with statistics programs and State Statistical Bureau (SSB) unified to establish the National Statistical Education Association (NSEA). Under the association, there are three branches: University and College Branch, Technical Secondary School Branch and Training Centre Branch. The university and college branch has 60 members which are departments with statistics majors or programs.

Each year the university and college branch organises at least one national meeting to discuss the statistics education reform and to enhance the education and training level. The technical secondary school branch has 15 members which are schools with the statistics major. With the high school more and more universal, the recruitment number from the statistics major in the technical secondary schools will be reduced in the future. The training Centre branch has 31 members which are provincial statistics training centres in each province, autonomous regions and municipalities directly under the central government. Figure 2 shows us the official statistics training system.

Figure 2. The National Official Statistics Training System



3.THE TYPE OF TRAINING PROGRAM

3.1 DEGREE PROGRAM

There are degree programs and non-degree (diploma) training programs. For the degree program, there are 12 academic degrees and 12 research fields stipulated by the Degree Committee of State Council, which are Philosophy, Economics, Law, Education, Literature, History, Management, Science, Engineering, Agriculture, Medical Science and Military Science. A statistics major for undergraduate study is under the degrees of Economics and Science. A statistics major for graduate study is under the degrees of Economics, Science and Medical Science. Most degrees have offered the statistics courses and have their own statistics faculties.

3.2 NON-DEGREE PROGRAM

Distance education

For the non-degree training programs, the most popular program is the Certificate of Economics and Management Statistics offered by Central Broadcasting & TV University and Statistics Education Centre of SSB. For the past 15 years, about 470

thousand students have been enrolled in this field and 250 thousand students graduated. On average, this kind of program needs 2.5 years. The universities have begun to offer statistics training programs and certificate programs through the Internet.

Job training

There are two kinds of job training: job knowledge training and job certificate training. Since 1990, in accordance with the reality that the level of statistical knowledge was not satisfied with the need of statistical work, a standard had been prepared for the requirement of knowledge and skill for statistical clerks, assistant statisticians, statisticians, and senior statisticians.

These standards and requirements have been subsequently implemented. By the year 1995, 800 thousand statistical personnel nation wide had participated in the training. In the job certificate training, according to the relevant laws and regulations of the government, official statistical personnel must obtain a certificate .

To accomplish this goal, two kinds of situations must be separated: Firstly, the majority of those who are already working had a previous certificate and they should be issued a certificate immediately after registration. Those who are not qualified to do this work should be re-trained before allocation to position. Secondly, there are a considerable number of new employees each year, and some of them do not possess the necessary qualification. They should be trained first and over 120 thousand employees have been trained and issued certificates since 1995.

Lectures, seminars and workshops

Except for the degree and non-degree programs, the university and college branch, statistics education centres as well as universities and colleges have organised many different short lectures, seminars and workshops for the statistical researchers, statisticians and employees.

There are two kinds of lectures and workshops. One kind of training is for special position staff. For instance, each year the Statistics Education Centre of SSB and universities organise the training workshops for the directors or deputy directors of provincial statistical bureaus. The provincial statistics education centres also organise the training workshops for county statistics bureau directors or staffs. Another training is for special topics. For example, the sampling survey technique and the system of national accounts are the most welcomed topics.

4.THE MAIN APPLICATION AREAS AND THE MAIN METHODS TO BE TAUGHT

Statistical methods are widely used in almost all fields. But right now in China the important application fields are as follows: official statistics, agriculture and industry experiment and management, economic situation analysis, medical hypothesis testing, etc. The researchers in the above fields especially need to be trained. We understand that the more developed the economy and society, the more widely used are statistics methods. Although at the current stage many fields like education, environment, risk management are less using statistics, and some other fields are short of statistics application, we, as educator and trainer, should do the training for the future in advance.

The topics to be taught should meet the needs of trainees. For the official statistician training, data collecting, tabulation, basic statistic analysis, regression analysis, time series analysis, sampling survey methods, econometric analysis are frequently involved.

For enterprise and company managers, tabulation, simple statistical analysis, cost analysis, forecasting and decision making are usually included.

5. THE MAIN ABILITIES TO BE EMPHASIZED

The aim of statistical training is not only to teach some new methods, but also to increase the trainee's abilities in their practice. The main abilities contain survey designing, data collecting, describing, computing, analysing and explaining. In order to achieve this purpose, we try to make the trainees understand the inside ideas, such as: Why statistical methods can explore the inside quantitative laws through the mass data; How statistical methods can solve the practical problem; What kind of method can be used under What kind of condition; How the statistical method can be correctly used, etc.

To understand, to emphasise, to answer these questions is the most important thing in our training. We should put the above ideas into our training materials as well as our textbooks, and train the trainers first.

6. TO TRAIN THE TRAINERS AND TRAINING MATERIALS

In order to achieve the high quality training, and to enable the trainees to master the capacities of applying the statistical methods correctly, we have to train the trainers first. One way to do this is to organise some seminars for faculties and for professional researchers. The University and College Branch of NSEA have organised this kind of seminar once a year. The participants from universities, colleges, research institutes, training centres and statistical bureaus get together to discuss how to organise effectively the training courses. Another way is to organise experienced experts to edit a series of training textbooks as well as the training materials. These experts are from the National Statistics Textbook Editing Committee, which includes three groups of experts:

- The first group is responsible for the textbooks and training materials of basic statistics methods, mathematical statistics and its applications in many fields;
- The second group is responsible for the textbooks and training materials of official statistics, economic statistics and some application fields;
- The third group is responsible for the textbooks and training materials used at technical secondary school.

Some famous statisticians like George Tiao, Jeff Wu, Lai K Chan and Ben-Chang Shia are the supervisors of this committee. Apart from editing textbooks, this committee also translated the good foreign textbooks into Chinese. Now we have already translated 15 books into Chinese and they are:

1. *Statistics* by David Freedman, Robert Pisani, Roger Purves and Ani Adhikari, 2nd Edition, 1991;
2. *Survey sampling* by L. Kish, 1965;
3. *Nonsampling error in surveys* by Judith T. Lessier and William D. Kalsbeek, 1992;
4. *Stochastic processes* by Sheldon M. Ross, 1983;
5. *Design and analysis of experiments* by Douglas C. Montgomery, 3rd Edition, 1991;

6. *Understanding robust and exploratory data* by David C. Hoaglin, Frederick Mosteller and John W. Tukey, 1983;
7. *Nonlinear regression analysis and its applications* by Douglas M. Bates and Donald G. Watts, 1988;
8. *Applied linear regression* by S. Weisberg, 1985;
9. *Statistical models and methods for lifetime data* by J. F. Lawless, 1982;
10. *Statistical methods for survival data analysis* by Elisa T. Lee, 2nd Edition, 1992;
11. *Forecasting financial and economic cycles* by Michael P. Niemira and Philip A. Klein, 1994;
12. *Statistical decision theory and Bayesian ana*lysis by James O. Berger, 2nd Edition, 1985;
13. *Discrete multivariate analysis theory and practice* by Yvonne M. M. Bishop, Stephen E. Fienberg and Paul W. Holland, 1975;
14. *Time series analysis forecasting and control* by George E. P. Box, Gwilym M. Jenkins and Gregory C. Reinsel, 1994;
15. *Introduction to variance estimation* by Kirk M. Wolter, 1985.

7. CASE STUDY AND TEAMWORK

Case study and teamwork are good measures in training. Like the course of business administration, it is important to find out good examples in statistics teaching. In 1996 and 1997, The City University in Hong Kong, Renmin University of China, Northeastern Finance and Economic University and Shanghai Finance and Economic University jointly organised 4 symposiums on Statistics Teaching and Training. On the symposiums, each university introduced their teaching experience as well as their cases. After the symposiums, a textbook of case study was edited and published. Now more and more universities, colleges and training centres choose these cases as their teaching material. There are 15 cases in the textbook and all of them are the real and practical problems in the current situation of China. Each of them has its training aim and some of them need teamwork. Now we introduce these cases:

- Case 1: The Questionnaire Design and Data Analysis for the Stock Buyers in China;
- Case 2: The Short-term Forecasting and Analysis of Chinese Industry;
- Case 3: The Correlation and Forecasting Analysis of the Two Stocks, Shenzhen Development Bank and Changhong Corporation;
- Case 4: The Analysis of Consumption Structure and the Estimation of Demand Function for the City Residents in China;
- Case 5: The Gini Coefficient Calculation using Survey Data;
- Case 6: The Effect Analysis of Age, Education Level, Marriage Status and Profession towards Mortality;
- Case 7: Can We Earn Profit from Pig Raising;
- Case 8: The Multi-Linear-Regression Analysis for Fiscal Income;
- Case 9: The Effect Factor Analysis of Women Birth Rate in China;
- Case 10: The Evaluation and Analysis for the Companies on the Stock Market;
- Case 11: The Quantitative Analysis of the City Differences and Urbanization;
- Case 12: The Comprehensive Evaluation on the Development Level of Economy and Society for the Cities in Liaoning Province;
- Case 13: The Time Series Model of the Stock Market in Hong Kong;

- Case 14: The Editing and Analyzing of the Index Number of the State Security;
- Case 15: The Probability Distribution of Individual Investor's Share Holdings.

8. STATISTICAL SOFTWARE AND INTERNET TRAINING

In recent years, computer aided teaching has been popularised in most universities, colleges and training centres. Some authors start to use Excel, SPSS, SAS in their textbooks. Many key universities, colleges and training centres delivered the training course for SAS. However, Excel is the most popular software.

Since China is a developing country and government education investment is not enough to meet the needs of the higher education development, computer systems and teaching facilities in most universities and training centres lag relatively behind. Therefore the students and trainees have not had enough time and chance to practice what they learned in the class.

The main problem for computer skill training is the combination of statistical method and software. In most statistical majors and training programs, the statistical courses and computer courses are offered separately. The textbooks are separately edited and published. The students and trainees can not solve the problems and questions using computer skilfully.

REFERENCES

Journal of Statistical Education, 1996-1999.

Statistical Education Centre (1997). D*istance education. An effective approach to improve the professional quality of the Chinese statistical personnel.* Beijing: State Statistical Bureau of China.

Wang, Q., & Lu, X. (1999) (Eds.), *Statistical case study.* Beijing: Northeastern Finance and Economics University Press.

Wei Yuan
Renmin University of China
175 Haidian Road, Beijing, 100872, China.
E-mail wyuan@mail.ruc.edu.cn

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 319-322.

Printed in Granada, Spain.

DISCUSSION

MARIA GABRIELLA OTTAVIANI

Going through the four papers of this session has been very interesting for me as I have found many points to reflect on.

1.TEACHING AND SOCIO-ECONOMIC CONTEX

I would like to begin by putting in evidence how statistics and its teaching are connected with the socio-economic and political situation of a country**.** This may disconcert some of us, but we cannot deny that this is once more testified by some of the papers presented in this session.

Wei's paper, after remembering us that China is in the stage of transferring from planning economy to market economy, shows how this evolution is also influencing statistics. In fact he observes that: the more developed the economy and society become, the more statistics methods are widely used. In practice this means that if in the past the focus was on the training of the official statisticians for government planning with economic statistics majors offered by most of Chinese Universities and colleges, in the future the official statisticians have to be trained also in sampling surveys techniques and the system of national accountancy, while statisticians have to be trained for enterprises and companies too.

A big change in statistics education and training has taken place also in Latin America. There, in the 80's, the end of the military régimes has lead to the global reorganisation of the countries of the area, consequently to the necessity of socio-economic quantitative information and, by this, to the increase in the number and quality of applied statisticians to support the changes.

When the development of a country becomes more favourable towards society at large and its economy, this requires a better knowledge of statistical methods which allows us to look into huge systems of observations on collective socio-economic and health phenomena that new technology allows to collect and memorise easily. Wei's paper about China, Batista's paper about Latino-America and Ospina's paper about Colombia are all in this line.

2.TRAINING OF PROFESSIONALS

The four papers also claim consequently for a process of education/training of professional statisticians able to handle data, from their collection to their analysis and interpretation. Wei underlines that the aim of statistical training is not only to teach some new methods, but also to increase the trainee's abilities in their practice. Batista's paper notes that the teaching of statistical methods is disruptive when it is separated

from those particular research methods belonging to the substantive discipline statistics to which it has to be applied

This is the consequence of the fact that the utilisation of statistical techniques referred to different areas is quite diverse as Ospina notes, and that interactions and discussions of applied statisticians and researchers of the substantive fields are a learning experience for all the people involved in quantitative research, as Bangdiwala suggests.

3. TEACHING AND PRACTICE OF STATISTICS

However, what seems to be very difficult is to create a correct "equilibrium" among the teaching and the practice of statistics. Wei observes at the end of his paper *that the student and the trainee do not have enough time and chance to practice what they have learned in the class.* Ospina, putting in evidence the needs of changing in the undergraduate Statistics programs in Colombia, says that what is urgent is *an increase in contact with the external world.*

Batista's paper may in some way help us to grasp the core of the problem. The authors note that statistical methods and methodological research are usually taught separately, while the future neo-professionals are asked to do the synthesis of the two fields. In my opinion the reasons for this are many. Scientists in some way are conservative. In the choice among training a user of statistical methods or a researcher in statistics – particularly devoted to develop new methods –, often without being aware of this, university teachers prefer to think that they are training a researcher in statistics.

On the contrary, society needs good users of statistics, able to join statistical methods and the research methods of the substantive discipline requiring their intervention, and thus able to choose the correct method to solve the correct problem. As Bangdiwala suggests*: The statisticians must also appreciate the nuances of the field of application if they are to collaborate efficiently.*

All of this is also put in evidence in Ospina's claims for the necessity of a good equilibrium among mathematics, statistics, computation and area of application (by the way, referring to the situation of Colombia, I wonder how and why an undergraduate student has to change his/her mind and become a passionate statistician after 4 semesters of full immersion in mathematics, according to the curriculum referred to in the paper).

4. TECHNOLOGY

Other problems for the education/training of statisticians come from the computer and the big changes due to the advance of technology. If a student or a trainee has the possibility to utilise or to be exposed only to a specific software, able to develop only particular kinds of statistical methods, this is obviously a limit "per sè" to the possibility of other methods to be presented and utilised (Batista).

Again if statistical courses and computer courses are separately offered and the textbooks are edited and published separately, the students and trainees cannot solve the problems and questions using the computer skilfully, as is noticed by Wei. But something worse may occur. Ospina remembers that computer programs in some situations give the false sense that statistical professional assistance is not needed. And

more clearly according to Bangdiwala: Given the potential abuses of statistical methods made easy by the proliferation of statistical software readily available, the potential danger to the profession from the ill-trained casual user is great. Thus, an educated consumer is the best client for our statistical profession (Bangdiwalla, 2001). In these words we not only see what the problem is but also its possible solution.

5. THE STATISTICAL PROFESSION

In fact, let me say that while, obviously, Bangdiwala says *our* statistical profession, referring to the activity of biostatisticians and clinical epidemiology, I think it is possible to endorse a little change in his phrase and agree with the idea that "an educated consumer is the best client for *the* statistical profession". In fact, another peculiar point emerging from these papers is the difficulty that statisticians have to face to have their profession and skills recognised in the university institutions, in the society at large and in the workplace. In my view the problem has at least two corners: the feeble position in the academy of the statisticians who are interested in the application of statistics to substantive disciplines and the poor statistical knowledge that people in charge in industries, factories and institutions have.

6. OVERVIEW OF PAPERS

I personally think that all the points I was struck by when reading these very interesting papers are problems we share in every country, no matter if it is a developed or a developing one. No doubt in an international perspective it is also important to keep in mind this kind of "classification", as there may be more than one way by which different countries may help each other to promote statistical education and training. Also from this point of view the four papers presented are very rich in proposals and examples.

In China, what seems the core of the problem is the vastness of the country that disconcerts everyone's is thinking on it. Obviously this gives a stronger sense to find and test possible alternative patterns to be used in training the large number of statisticians and applied statisticians needed by the country.

In Colombia, an international committee suggested by the ASA acted as an evaluator of the Colombian program in the view of giving international accreditation of the local undergraduate and graduate programs. This Committee has helped to examine the situation and from the paper it is as if, in Colombia, statistics teaching and training suffer from problems that are common to many Universities in the world: too much importance is given to mathematical statistics and to academy goals in comparison with the attention that should be given to a more instrumental use of statistics needed by the country's situation.

The International Clinical Epidemiology Network project, originally funded by the Rockfeller Foundation, was aimed to establish Clinical Epidemiology Units as centres for excellence in various countries so that local well-trained clinical researchers would study the health problems faced in those countries. The main problem the project seems to have to face is the "brain drain", also at national level, due to the fact that biostatisticians are very much in demand but there are far too few of them. The

consequence is that well-trained biostatisticians are inclined to leave the Centre where they have been trained, thus impoverishing the institution itself.

In Latino-America, the PRESTA project offers an example of how some European Universities with the sponsorship of the European Community have tried to fill the gap between the available statistical methods and their effective and potential users. The idea has been to utilise a co-operating strategy aiming at creating links among European and Latino America universities, to enhance co-operation both within Latino American universities and between Latino-American universities and their public national institutions. All of this has produced a local horizontal co-operative net of 10 countries in the area, insuring different geographical and institutional presence. This project enlightens the importance of the local contributions and the capability to produce synergy among local participants to the project.

What I have learned from these papers, on this matter, is the importance of the comprehension and attention which must be given to the local situations and the necessity to enhance statistics, its concepts and methods, making a correct and intelligent use of local human resources, tools and equipment. All of this is a challenge that needs time and energy to be faced because, as some authors have noticed, whatever applied field is on the ground, the creation of a “critical mass” of well-trained statisticians is required to obtain results that remain and grow in time.

REFERENCES

Bangdiwalla, S. I., & Muñoz, S. (2001).Training of statisticians and clinical researchers world-wide to collaborate as co-investigators within country clinical epidemiology units: The experience of the International Clinical Epidemiology Network (INCLEN). In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 265-275). Granada: International Association for Statistical Education and International Statistical Institute.

Crivisqui, E., Abruzzini, S., & Batista, C. M. (2001). How to overcome the gap between the available statistical methods and their effective and potential uses. A few thoughts about the experience in the PRESTA Programme. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 277-293). Granada: International Association for Statistical Education and International Statistical Institute.

Ospina, D., & Ortiz, J. (2001). Statistical research and consulting in universities from developing countries: The Colombian case. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 295-310). Granada: International Association for Statistical Education and International Statistical Institute-

Wei, Y. (2001). The training of researchers in the use of statistics in China. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 311-317). Granada: International Association for Statistical Education and International Statistical Institute.

Maria-Gabriella Ottaviani
Universita' di Roma "La Sapienza"
Dipartimento di Statistica, Probabilita' e Statistiche Applicate, Roma, Italy
E-mail: mariagabriella.ottaviani@uniroma1.it

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 325-338.

Printed in Granada, Spain.

GABRIELLA BELLI

THE TEACHING/LEARNING PROCESS IN UNIVERSITY STATISTICAL CONSULTING LABS IN THE UNITED STATES

The main focus of this paper is on how statistics students are trained in consultancy skills, as well as on how faculty in charge of university statistical consulting units perceive the consultant's role in training researchers. An electronic survey of 106 USA departments was conducted. Results indicate a wide range of practices in how students consult and how they are trained in consulting skills, but much greater consistency in the belief that such training is essential to a statistician's education and about the problems faced by both student consultants and student clients. The consulting service was seen as a useful way for researchers to learn because they would be working on a problem or data set of interest to them. Respondents discussed the importance of collaboration, particularly as a goal for consulting relationships that would benefit both parties, with reciprocal teaching and learning.

1. INTRODUCTION

Consultancy is an integral part of a statistician's life, whether in academe, government, or industry. Such consultation may involve brief interactions to respond to very specific questions or long term associations as team members on on-going projects. If we agree that:

> "*the function of statistics is to solve real problems (across all subjects from agriculture, through medicine to zoology)"* (Barnett, 1993, p.285),

then it follows that a statistician must be able to effectively communicate with researchers and practitioners and be conversant in their functional areas. Consultancy skills should therefore be an important aspect of statistical training. Such skills can be developed on the job (Ruberg, 1998), through statistical consulting courses (Khamis & Mann, 1994; Rangecroft & Wallace, 1998; van Belle, 1982), or via student consulting services in college or university statistics labs (Calvin, 1982; Halavin & Mathiason, 1994; Meyers-Tate, 1999).

Not only do statisticians need to understand enough about their colleagues' disciplines to be effective consultants, but also researchers should have enough appreciation of statistical concepts so that the association may be more productive. Bentley, Schneider, and Bentley (1998) showed how an introductory course in the combined area of archaeology and statistical reasoning allowed

> "*statistics students to learn the basics of consulting*" and *"humanities students to learn the benefits of a statistical perspective"* (Bentley, Schneider and Bentley,1998, p.347).

They provided a paradigm of how a course could be used to train both future consultants and consultees. In detailing how a pharmaceutical company trains statisticians to be truly effective team members, Ruberg (1998) stated that:

> "S*uch training impacts other scientists with whom the statistician inte*racts" (Ruberg, 1998, p.365).

He went on to say that while opportunities such as conferences exist for the statistician to become fluent in pharmaceutical issues, much of the learning comes from working with clinical scientists or on project teams. Likewise, much of the statistical training for clinical scientists comes from the interaction with statisticians.

Many universities in the United States, particularly those with doctoral programs in statistics, have statistical consulting laboratories or centres. Similar units also exist in other departments such as educational psychology. It is reasonable to assume that they provide a useful service for faculty, staff, and students as well as function as a training ground for students in both statistical applications and consultancy. The need for such services and training extends even to small liberal arts colleges without graduate programs (Herring & Jersky, 1995). Regardless of how they function, such units are also a forum for interaction between statistician and researcher and therefore should be a fertile ground for learning on both sides.

One aspect that characterises some articles dealing with statistical consulting labs is the explication of what is done or what has occurred in a specific university or department. In essence, these are case studies of particular centres or labs that provide, implicitly or explicitly, a model for others to follow (see e.g., Meyers-Tate, 1999). Authors often propose a paradigm for how consulting labs can be used to train statistical consultants, provide guidelines for the interaction between consultant and client, or for what the client needs to do for a successful interaction. In their often cited book, Boen and Zahn (1982; reprinted 1997) state that:

> "*It is well known, perfectly understandable, and widely accepted that colleges and universities do not fully prepare their graduates for 'success on the job' as judged by non-academic employers*" (Zahnm 1982, p. xi).

They therefore wrote their book based on their combined experiences in order to help fill that gap. Given the need for training in the non-statistical aspects of consultancy and in how to interact with researchers from diverse fields, as well as the many recommendations about what should or could be done, what is being done in college and university statistics consulting units? To what extent and in what ways are students trained in consultancy skills? What are consultants' perceptions of their role in training researchers from other fields? Are any guidelines provided for clients? In order to answer these and other questions about consultancy units, a survey of faculty in select university statistics labs within the United States was conducted. For this study, the main interest was in situations that employed student consultants.

2. PROCEDURES

The original sampling frame for this study was constructed by a number of electronic searches. Several search engines were used to search for statistical consulting labs, centres, units, or services. Then a few web pages with links to statistics and

mathematics departments were used to search for information about statistics consulting services. The final working list consisted of 248 potential university contacts. If no information was apparent from the web site, an e-mail was sent requesting if there was a statistical consulting unit or a consulting course in the department. The initial list was reduced based on replies that only faculty consulted or that a consulting unit no longer existed.

A closer inspection of 85 web pages for departments who did not respond to this request indicated that only seven were likely candidates for this survey. The rest did not appear to have student consultants, and most included academic grants and contract services, computer support labs, or departmental faculty who provided consulting services.

The remaining contact persons were then sent a pre-notification by e-mail about participating in the survey as well as a request for another contact for statistical consulting services within their university. In addition, they were asked if a consulting course was taught either by their department or elsewhere in their university. Changes were again made to the sample frame based on replies with new potential contacts and additional responses that no such service was available. This left a final usable sample of 106 contacts. This group was sent a second e-mail notification that included the URL for a questionnaire.

The questionnaire, along with an informed consent form approved by Virginia Tech's Institutional Review Board, was located at the web page http://www.nvgc.vt.edu/gbelli/StatLabSurvey. Responses to the questionnaire went directly to an Access database file. To insure anonymity, the responses were received without any identification of the respondent or any way to track from which computer the response came. Closed numerical responses were transferred to an SPSS version 10 data file for processing. Open-ended responses were formatted and printed out for content analysis.

Due to the nature of the items, only descriptive statistics were used on the numerical data: Medians, modes, frequencies, and percents. For a few items that asked respondents to "check all that apply," a multiple response procedure was used. This enabled computing the percentages based on the total sample size as well as on the total number of checkmarks. Open-ended responses were content analysed and organised into sets. In this way, common themes emerged and could be summarised. Individual quotes are used in the discussion if they capture the essence of a set of comments.

3. DESCRIPTION OF THE SAMPLE

A total of 43 questionnaires were returned (a 41% response rate). Although a number of items were general in nature, many were specific to student consultants and student clients. It is likely that at least part of the non-response was due to individuals seeing the items about students and not reading further. Another, more detailed, review of the final 106 web pages revealed at least 15 where this may have been the case. Web page content about consulting services varies dramatically, from less than a page with only cursory information to detailed hyper-linked pages that may include sample projects, forms for clients, and lists of both services provided and names of providers. It is often difficult to tell if faculty, hired consulting staff, or students are doing the consulting. Web pages with similar sounding mission statements can represent

dramatically different consulting units in terms of size and type of services provided. Many responded immediately to initial e-mail requests while others never replied.

While some simply provided the requested information, a number were quite enthusiastic about the need for information about students in consulting services. Certainly some departments included in the sampling frame did not have students consulting and, just as certainly, others that did so were missed. This is due to the way the sample was constructed and the imperfect information available. The non-response adds to the problem. However, even though it is impossible to tell definitively, the diversity of responses seem to reflect the diversity about statistical consulting that is evident in university web pages. While the responding sample is neither large nor random, the responses are informative.

About half of the respondents were directors of consulting units (53%) and almost one quarter were managers or supervisors, several of whom also seem to function as student mentors. The rest of the respondent profile included four consultants, two faculties who teach a consulting course, and two department heads. One respondent said they do not have a lab per se, and two provided no information. Although only four individuals listed consulting as their only role, a number of others listed consulting along with other roles such as focusing on coaching student consultants or teaching a consulting course. Three individuals said they did not use students as consultants.

Over half of the consulting units were housed in statistics departments (56%), with the remainder spread over a number of departments. These included six in mathematics departments and a few in joint departments such as mathematics and statistics or biostatistics and statistics. Additional departmental affiliations were Applied Statistics & Operations Research, Educational Psychology, Forest Science, and an Oncology department. One centre was housed in Learning Resources and Technology, another in the Computer Center, and one had no departmental affiliation. Over half the respondents said that statistical or research consultation was provided elsewhere in their university (58%), however in many cases that represented internal departmental “experts” who provided statistical consultation within their own department. A biostatistics department was mentioned the most (seven times), followed by education (five times).

With respect to the number of faculty lines allocated to consulting units, 12 said they had less than one full time equivalent (FTE), 11 had between one and two FTE, and nine had between 2.5 and 3.5 FTE. The rest responded in terms of hours per week or number of students allocated to the lab or said they had none. The consulting units also varied in terms of hours of operation. Four were open less than 10 hours per week, 15 between 10 and 30 hours, and another 14 either 35 or 40 hours. The remainder provided a variety of answers, such as by appointment only, by e-mail, all the time and no official hours.

4. THE CONSULTING LAB AND STUDENT CONSULTANTS

The first set of questions asked respondents to rank five reasons in terms of their importance for maintaining their consulting lab. Two reasons with a modal importance rating of one and median ratings of two were: (1) to provide statistical assistance to faculty and (2) to serve as a training ground for student consultants (see Table 1).

They were rated as most important reasons for maintaining consulting units by 47% and 44% of the respondents, respectively. Providing statistical assistance to students had both a median and modal rank of two, but was rated first by only 24% of the

respondents. Although two-thirds said that at least some of their services are fee based, particularly for clients who are external to the university, it is interesting that only 7% of the respondents felt that a primary reason for the statistical consulting service was to produce revenue or to serve external clients. Beyond the items listed, the following types of reasons were given for maintaining a lab. (Items are summary combinations of several responses):

1. Keeping faculty up to date by working on real problems, thereby enhancing classroom examples;
2. Initiating collaborative links across campus thereby promoting interdisciplinary research;
3. Provide statistical assistance to administration and staff;
4. Provide financial support to students.

Table 1. Reasons for Maintaining a Statistical Consulting Lab or Service

	Median rank	Mode	% rating item 1st	2nd
Providing statistical assistance to faculty clients	2.0	1.0	46.5	27.9
Serving as training ground for student consultants	2.0	1.0	44.2	14.0
Providing statistical assistance to student clients	2.0	2.0	23.3	48.8
Providing statistical assistance to clients external to the university	4.0	4.0	7.0	14.0
Generating revenue for the department or college	5.0	5.0	7.0	4.7

When asked to describe who provides consulting services, faculty were listed by 74% of the respondents and graduate students by 67%, with each representing roughly one-third of all the responses (see Table 2).

About 40% said students consulted as a requirement in either a consulting course (23%) or a statistics course (16%). One-quarter hired non-student consultants and 12% used undergraduates. The most common services, provided by over four-fifths of the consulting units, are: (1) interpreting statistical results (95%), (2) implementing statistical analyses (88%), (3) problem formulation/translating hypotheses into statistical terms (88%), (4) sample size determination (88%), (5) planning statistical analyses (83%), and (6) research design (83%).

Each of these represented about 10% of 341 checkmarks on a list of 14 possible services. The two least performed services dealt with consultations on measurement theory, scale construction (13%) and providing short courses or workshops (5%).

Table 2. Profile of Consultants

	N	% of respondents[a]	% of responses[b]
Faculty on staff	32	74.4%	34.0%
Graduate students, in assistantship or internship	29	67.4%	30.9%
Hired non-student consultants	11	25.6%	11.7%
Students in consulting courses, as requirement	10	23.3%	10.6%
Students in statistics courses, as requirement	7	16.3%	7.4%
Undergraduates, in assistantship or internship	5	11.6%	5.3%

[a] Percent based on number of respondents (n_a=43)
[b] Percent based on total number of items checked across the six options (n_b=94)

Respondents were then asked to check any of eight types of consulting skills for which they provided training to student consultants. Three of the respondents (7%) said they did not provide training for students. According to the rest (see Table 3), the main types of consulting skills training deal with applied statistics problem solving (87%), problem formulation (82%), general problem solving (74%), and oral and written communication (74%). These four items represented from 15% to 17% of the 199 total selections made by 39 respondents. Only 11 people (28%) indicated training in how to teach within a consulting session as a skill that was taught to student consultants, and less than half trained students in session management. These results seem counter-intuitive, given the importance placed on such skills.

Table 3. Consulting Skills Training Provided for Students

	N	% of respondents[a]	% of responses[b]
Applied statistics problem solving (e.g., about appropriate advice to give clients)	34	87.1%	17.1%
Problem formulation (e.g., translating a client's problem formulation to a statistical formulation)	32	82.1%	16.1%
General problem solving	29	74.4%	14.6%
Oral and written communication	29	74.4%	14.6%
Interpersonal relations	25	64.1%	12.6%
Ethical standards in statistical consulting	20	51.3%	10.1%
Session management	19	48.7%	9.5%
How to teach within a consulting session	11	28.2%	5.5%

[a] Percent based on number of respondents answering this question (n_a=39)
[b] Percent based on total number of items checked (n_b=199)

Open-ended responses following this question indicate the extent of variability in statistical consulting in universities. On one end, one respondent said that all of the skills given in Table 3 were provided in a class discussion mode. Another indicated that students take a six-week course in consulting before meeting with clients and that their first client meeting is with an experienced consultant. At the other extreme, one respondent indicated no formal training and that the students learn by assisting regular consultants and through experience. This diversity can be seen in Table 4, which provides a checklist of how respondents said that students learn consulting skills.

Table 4. How Student Consultants Learn Consulting Skills

	N	% of respondents	% of responses
Through informal mentorships	25	59.5%	22.7%
Formally via course work on consulting	23	54.8%	20.9%
Picked up through their own experiences	22	52.4%	20.0%
Through reading materials provided or suggested	19	45.2%	17.3%
Formally via scheduled meetings or workshops	15	35.7%	13.6%
Through other means	6	14.3%	5.5%

[a] Percent based on number of responses with student consultants (n_a=42)
[b] Percent based on total number of items checked (n_b=110)

Over half of the respondents, as well as less than a quarter of the total items checked, indicated that this was done primarily via informal mentorships or formal coursework (only seven indicated both ways) and/or picked up through their own experiences. Less than half credited reading materials or regular meetings and workshops. Additional comments indicated that students pick up these skills by working jointly with faculty consultants and by informal training and observation.

When problematic research or statistical issues arise, the most frequent way they are handled is via consultation with faculty (72%). In very few cases are issues resolved openly at regular staff meetings (12%) or jointly among the consultants (9%). In one case, they use videotapes of the sessions.

With respect to formal technical skills, respondents think that student consultants bring only a moderate level of knowledge to their consulting. Based on a five-point scale from limited knowledge to proficient, four areas achieved a median rank and at least one modal rank of three: applied statistics, statistical software, theoretical statistics, and research methods (see Table 5).

Measurement had the lowest median ratings: two for applied measurement and one for theoretical measurement. Over one-third of the respondents could not rate this area. Overall, though, this seems a reasonable profile for beginning student consultants, some of whom are doing this as part of a course. Given that improvement in statistical and research understand is at least part of the purpose of having students consult, it would be interesting to evaluate their knowledge gain as a result of this practice.

Table 5. Beginning Student Consultants Level of Technical Knowledge

	Median	Percent[a]					
		1	2	3	4	5	DN
Applied statistics	3	4.7	11.6	32.6	32.6	14.0	4.7
Statistical software	3	4.7	20.9	37.2	23.3	7.0	7.0
Theoretical statistics	3	16.3	16.3	32.6	14.0	11.6	9.3
Research methods	3	14.0	25.6	25.6	14.0	4.7	16.3
Applied measurement	2	27.9	14.6	14.0	7.0	2.3	34.9
Theoretical measurement	1	34.9	16.3	4.7	9.3	0.0	34.9

[a] Responses from 1 = basic coursework or limited to 5 = advanced coursework or proficient.
DN = don't know.

5. DEALING WITH CLIENTS

Considering the clients who come to consulting sessions, 44% of the respondents provide guidelines for clients before they interact with a consultant and 23% have time limits on the number of sessions clients may have.

The type of guidelines vary from formal written policies, information in a brochure and on the web, or new client questionnaires to signed contracts or waiver forms. In a number of cases, the main specification was that only consultation and no data analysis was provided.

Limits on the number of consulting sessions were placed in only a very few instances. A couple of web pages made reference to the client guide available from the Statistical Consulting and Research Group at Northern Arizona University. Over two pages long, this guide is available at the following URL:

http://odin.math.nau.edu/~scrc/client_guide.html. It contains answers to such questions as: "How much statistics do I need to know?" and "What will my first session be like?"

Over two-thirds (67%) have some type of fee for services. The most common seem to be for clients who are external to the university, for funded projects, or for routine tasks of data entry, program writing, and data analysis. Some additional responses were that work on grant preparation for faculty was free if a faculty consultant is included in the grant; that there was a sliding scale differentiating among industrial projects, funded campus projects, and unfounded projects, where the first three hours were gratis. In only four instances were all clients charged.

Almost half the respondents (47%) said that consultants are selected based on a match between client need and consultant skill, while 30% do it on a first come, first serve basis. Additional clarification on this item indicated instances where a combination of these two is used, where the client picks the consultant, or where the students volunteer after the applications are read. One Consulting Center director indicated that clients are matched to student consultants according to appropriate background if the student is in a consulting class or if the director takes the client call, but that students in the Consulting Center get clients depending on availability.

As shown in Table 6, most clients approach consultants after the data are collected (reactive), particularly so for student clients (64%). About one quarter of both student and faculty clients are said to be proactive, engaging a consultant early on in the research process. Only 14% of both faculty and external clients are felt to be true collaborators, where the client and consultant work as a team throughout project. It should be noted, however, that one quarter of the respondents either didn't know about or did not have external clients.

Table 6. Client Characteristics by Type of Client

	Student clients	Faculty clients	External clients
Reactive	62.8%	58.1%	46.5%
Proactive	25.6%	25.6%	14.0%
Collaborative	4.7%	14.0%	14.0%
Don't know or N/A	7.0%	2.3%	25.6%

The next set of items asked for a rating of student clients. The profile of the perceptions of areas with most to least expertise for student clients (Table 7) and student consultants (Table 5) does not match.

Table 7. Typical Student Client Level of Technical Knowledge

	Median	Percent[a]					
		1	2	3	4	5	DN
Research methods	3	14.0	32.6	30.2	18.6	0.0	4.7
Statistics application	2	27.9	34.9	18.6	14.0	2.3	2.3
Statistical software	2	37.2	34.9	7.0	7.0	7.0	7.0
Measurement application	1	44.2	18.6	7.0	4.7	0.0	25.6
Statistics theory	1	72.1	11.6	2.3	0.0	4.7	9.3
Measurement theory	1	60.5	4.7	4.7	2.3	2.3	25.6

[a] Responses based on a 5-point scale from 1 = limited to 5 = extensive. DN = don't know.

Student clients are rated as having only moderate to rather limited knowledge, with research methods the only category having a median rank as high as three. As might be expected, the least knowledge is seen in theoretical aspects of both statistics and measurement and in applied measurement. As before, though, over a quarter of the respondents didn't know about knowledge in measurement.

6. PERCEIVED ROLE OF THE STATISTICAL CONSULTANT IN TRAINING RESEARCHERS

In response to an open-ended question about the consultant's role in training researchers in other fields, the most prevalent answer dealt with various aspects of helping them develop quantitative thinking and the importance of the scientific method. This role was seen as being important, and one individual described it succinctly as a way to "*improve the overall quality of scientific research.*"

Although only one respondent said that the "*aim should be to help the researcher become more self-sufficient,*" this concept was implicit in a number of comments. People learn from mistakes, but "*this won't happen unless consultants can diplomatically bring these mistakes to the researcher's attention.*" This goal is not only for the client's benefit, however, as seen from the continuation of this statement: "*So training one's clients is critical, especially if you wish to have an on-going relationship.*"

Five individuals wrote at length about the client and the importance of "*having them take an adequate amount of coursework in statistics.*" This relates, in part, to the lack of time to do extensive teaching in a consulting session, particularly for short-term interactions. Educating clients to be proactive and consult early on research and questionnaire design was seen as an important role for the consultant, probably as a reaction to the number of reactive clients who approach a consultant after data are collected,. Such behaviour would not only enhance the final product, but would help clients "*identify the most effective and efficient approach to the problem at hand and to jointly develop implementable plans*."

Another theme that emerged was that of the consultant as an educator, as someone who had an important role to play because of the utility of learning by doing. The consultant was seen as having "*the unique opportunity of teaching clients with an example that is already of importance to the client*." Some of the respondents offered lists of general types of assistance that a statistical consultant could provide to clients. These included helping them solve research problems, assisting in designing and conducting research studies, providing practical advice on collecting, analysing data, and interpreting data, as well as providing references, suggesting courses, and instructing on specific procedures. The term "*collaborative mentoring*" was used. This seems quite descriptive about the perceived nature of the interaction as a mutual teaching and learning situation.

There were a couple of discordant views. One respondent felt the role of the consultant should be limited to assisting when asked and another that there was no role beyond "*some guidance to the researchers when their designs are poor.*" This latter individual, however, felt that this was a rare occurrence. But these were not the prevalent views in this sample.

One individual summarised the role of the consultant quite well: "*Statisticians certainly have a part to play in the education of researchers about experimental design,*

choice of response variables, and analysis and interpretation of analysis. However, the role of the statistician, and the ability of the statistician to do these jobs well, is limited by the desire and ability of the client. Some clients are very willing to learn a lot and they are motivated to do so. Others are not motivated and want only a brief response. As a consultant, we need to be responsive to those desires. On the other hand, we have to be ready to take every opportunity to help train researchers. By being proactive in this area, by providing training opportunities (short courses, workshops, seminars) and by being willing to invest a little more effort than a client might expect, we can demonstrate our desire to help researchers learn the tools they need to work."

7. TEACHING CONSULTING SKILLS

Nearly all of the respondents thought that it was important to teach statistical consulting skills (93%), but only 74% said that a formal statistics consulting course was offered (67% in the respondent's department and 7% elsewhere). However, an open-ended question that allowed respondents to comment on any aspect of teaching consulting skills produced only 15 responses. While no claim can be made for their representativeness, the comments are interesting and therefore will be discussed.

Several individuals laid out two essential aspects for teaching consulting skills. First, the experience must be problem based with real examples, small group discussions, and model sessions. Second, there needs to be opportunity to practice because experience is critical. Consulting skills are "*hard to teach*" but "*important to learn*," therefore students need "*lots of chances to interact with people from different disciplines with different types of problems."*

Three individuals stressed the importance of providing statistics students with consulting skills, and one provided a graphic metaphor: "*Not teaching these skills is like sending someone out to play professional tennis, possessing only a racquet, tennis balls, and several courses on the physics of bodies in motion."*

One respondent thought that having a consulting course depended on the size of the program, with smaller programs able to teach these skills individually through hands-on experience but that larger departments needed the formalised structure of a course. The latter would provide all students with the opportunity to perform in a consulting situation as well as "*help formalise faculty involvement with student consultants*." Another cited the importance of "*some training in general consulting skills like listening, attention to timeliness and physical constraints, friendliness, respect for all people,* etc." But, also added that "*statisticians might not be able to teach these well in all cases. May need someone besides a statistician to teach these skills effectively*."

In the midst of all the very positive comments, two negative aspects were brought forth. First, the incredible amount of effort a consulting course entails, and that *"teaching such a course takes a very skilled instructor and one with years of experience."* Second, although consulting is important for the students, *"university funds students to be TA's and help with classroom teaching and grading"* but "*is stingy at funding students to serve as assistants to consulting projects."*

8. PROBLEMS STUDENTS ENCOUNTER IN STATISTICAL CONSULTING

Unlike the previous question, nearly everyone answered this one on what is the most common problem that students encounter in statistical consulting. With respect to *student consultants*, a series of responses all dealt with different aspects of problems with interpersonal and management skills. In particular, things like knowing how to listen, establishing clear communications, asking enough questions to truly understand the client's problem, negotiating a reasonable time-frame, knowing how to effectively run a consulting session.

Another problem is *"determining the difference between consulting with someone and doing it for them,"* even though sometimes that is more expedient. Given that student consultants are still learning, they sometimes *"lack confidence in their statistical skills" and "are afraid to make a mistake."* They are not able to tell a client that they don't know the answer; that they need time to think about it. This is where lots of practice, particularly in supervised settings helps. Sometimes, however, they are deficient in the statistical or methodological knowledge needed to help a client. But this can, to some extent, be alleviated by judicious matching of client and consultant, by having formal meetings to discuss difficult cases, or by pairing a novice consultant with an experienced one.

The one difficulty most often mentioned was problem formulation. Respondents mentioned the fact that student consultants have a hard time *"abstracting the real problem from the story that the researcher tells."* They have problems interacting with clients due to not "*understanding enough about the problem from the perspective of the other discipline to give good statistical advice*." Even if they have the proper knowledge, they may not be able to figure out "*how the client's problem relates to the methods they are familiar with*." Another difficult, but related, situation for student consultants has to do with not knowing how to deal with clients who "*have no clear idea regarding their research objectives*" or "*come in with one introductory course in stats and want help carrying out sophisticated analyses*." Because a substantial amount of the assistance provided is reactive in nature, the student consultant is often faced with difficulties from two perspectives. One, trying to make sense of data from a flawed design, and two, having to deal with how to tell a client that most of their research questions are not answerable given the data they collected.

A few respondents provided information about the problems faced by the *student client*. They cited things like misunderstanding the method or analysis presented to them, not being able to communicate effectively, and "*lack of initiative during the meetings*." Although some student clients with little background want guidance in complex analyses, others seem to have the opposite problem. Some are advised by consultants to *"run designs with difficult issues that are well beyond their capability to understand and resolve*." Both situations are frustrating and very likely lead to poor research.

9. DISCUSSION

It appears that there is great diversity in how students gain statistical consulting skills in the United States. A few models emerged from this study: (1) a formal consulting centre or lab where students consult with some form of supervision, (2) a consulting course where students may work in groups or function as individual consultants on a given project, and (3) participating in consulting sessions with a faculty consultant as mentor or role model. In some cases statistics students receive specific

training in consulting skills, while in others they are expected to learn by observing an experienced consultant. Consulting courses may also differ in terms of the content covered and the activities involved (Belli, 2001). At one extreme, the students in the course effectively served as the consulting lab. At the other extreme, they worked in small groups on real or model problems. But, for all the apparent diversity, there seems to be overwhelming consistency in the acceptance that consulting skills are an important part of a statistical education. Some people consider these to be nothing more than a strong statistical background that enables a consultant to resolve a client's problem. Given this view, teaching statistical skills revolves around exposing the student to a range of problems from various disciplines so they can develop their experience. In contrast, what seems to be a majority opinion at least for this sample, is that this is only half the lesson. While statistical and methodological knowledge is certainly important, it is not effective if the consultant cannot adequately communicate it or cannot manage a consulting session properly. In such cases, effective teaching can not occur as part of the interaction. Likewise, a researcher who does not know how to work collaboratively with a statistician, and only utilises one at the end stages for analysis is not able to benefit fully.

The statistical consultants' role in training researchers in different fields may very well be a function of what the client desires. Although some results here were no surprise; e.g., that too many clients tend to seek help only after data collection and arrive for a consultation with poorly defined questions or questions that don't match the data collected. Certainly these are problems and the consultant must learn how to interact with such clients in a reasonable manner. A far more interesting result, however, is how collaboration was mentioned. In particular, how training clients to come early in the research process could lead to joint determination of process and even collaboration in future projects. This transforms consultancy from a service performed to a valuable working relation that benefits both parties. As Svensson (2001) states about one area:

> *"there is a need for more biostatisticians with an interest in collaborative research, not only for the improvement of the applied research, but also for the development of the biostatistical science"* (p. 33).

The same could be said for any applied area. Indeed, Jollife (2001) states that:

> *"significant progress in any field of application needs the participation of both specialists in the field and of statisticians"* (p. 365).

On the one hand, several respondents mentioned how consulting centres *"were not appreciated by the administration"* and were seen by departments as "*a service rather than a contribution to the research infrastructure."* It is therefore difficult to get funding for them and consulting efforts are not rewarded. These sentiments echoed results from a recent survey of heads of biostatistics services in health care research settings in the United States and Canada (Niland, Odom-Maryon, Lee, & Tilley, 1995). Problems with insufficient institutional funding and lack of co-authorship on manuscripts after substantial input were reported. Several presenters at the August 2000 International Association for Statistical Education Round Table Conference commented that similar problems exist in their countries. Some individuals from this sample, however, indicated how collaborative projects and inclusion in grants resulted "*from routine consultations*" and how the centre is *an "excellent source of research problems"* for classroom use and

"*has led to a number of published articles*." One responded summed up by saying that "*these types of collaborations are highly encouraged by the university and are rewarded at the departmental level."*

Professional consultations are a two-way street, and so should be student to student consults. Both parties must bring something relevant to the table and could be taught to do so. The student clients needs some assistance, but must be willing to do their part in terms of communicating effectively to the consultant and learning the necessary statistical concepts. Likewise, the consultants must work to understand the client's domain enough so that a useful solution may be found. The result should be beneficial to both parties. The consultant gains experience and improves in both statistical and consulting skills. The client gains a solution to a research problem and improves in statistical knowledge. More importantly, both can develop an appreciation of collaboration that they carry with them after they graduate and into their professional careers. This is consistent with Godino, Batanero, and Jaimez's (2001) conception of :

> *"statistical consultancy as a device to co-operatively study data analysis problems"* and that *"consultants need the client's contribution, as much as clients need consultant's knowledge"* (p. 347).

Although this survey had a small number of respondents, mostly from statistics or statistics related departments, they provided useful information about the process and problems in university statistical consulting units. One interesting result was the extent to which things like problem formulation and research design issues were part of typical consultation. While one might expect this from more general *research* consultancy units rather than *statistical* consultancy units, this did not seem to be the case. A follow-up survey will be conducted with colleges of education, in part to determine the extent to which any research consulting services there differ from those in more quantitative departments. A search of web pages to find listings for statistics or research consulting in colleges of education under programs in educational psychology, measurement, statistics, or research methods is already in progress. The final list of web pages will be put on a web page, with links to statistical and research consulting services and statistical consulting courses in USA universities. These results will provide a substantial profile of the consulting services available in USA colleges and universities.

REFERENCES

Barnett, V. (1994). The role of consultancy in university education and professional training in statistics. In L. Brunelli, & G. Cicchitelli (Eds.), *Proceedings of the First Scientific Meeting of the International Association for Statistical Education* (pp. 285-297). Perugia, Italy: International Association for Statistical Education.

Belli, G. (2000). The teaching of statistical consulting skills. Paper presented at the *Meetings of the American Educational Research Association,* New Orleans.

Bentley, J., Schneider, T. J., & Bentley, D. L. (1998). Statistical consulting in archeology: Digging for real data. In Pereira-Mendoza, L., Seu Kea, L., Wee Kee, T., & Wong, W. K. (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (347-352). Voorburg: International Association for Statistical Education and International Statistics Institute.

Boen, J. R., & Zahn, D. A. (1982). The human side of statistical consulting. Belmont, CA: Lifetime Learning Publications, a division of Wadsworth, Inc. (Reprinted 1997 by the

Pharmaceutical Education & Research Institute (PERI) and PHRMA's Biostatistics Steering Committee).

Calvin, L. D. (1982). Experience with a student consulting service. In J. S. Rustagi, & D. A. Wolfe (Eds.), *Teaching of Statistics and Statistical Consulting* (pp. 311-326). New York: Academic Press.

Godino, J. D., Batanero, C., & Gutiérrez- Jaimez, R. (2001) The statistical consultancy workshop as a pedagogical tool. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 339-353). Granada: International Association for Statistical Education and International Statistical Institute and International Statistical Institute.

Halavin, J. J., & Mathiason, D. J. (1994). Designing and running undergraduate statistical consulting laboratories. *American Statistical Association 1994 Proceedings of the Section on Statistical Education* (pp. 132-136). Alexandria, VA: ASA.

Herring, S., & Jersky, B. (1995). The statistical consulting needs of a small liberal arts college. *American Statistical Association 1995 Proceedings of the Section on Statistical Education* (pp. 134-137). Alexandria, VA: ASA.

Jolliffe, F. (2001) Learning from experience. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 355-370). Granada: International Association for Statistical Education and International Statistical Institute.

Khamis, H., & Mann, B. (1994). Outreach at a university statistical consulting center. *The American Statistician, 48(3),* 204-207.

Meyers-Tate, C. (1999). A model for applied statistical consulting in a statistics lab environment. Paper presented at the *Annual Meetings of the American Educational Research Association,* Montreal, Canada, 19-23 April.

Niland, J. C., Odom-Maryon, T. L., Lee, J., & Tilley, B.C. (1995). A survey of biostatistical consulting units throughout North America. *The American Statistician,* 49(2), 183-189.

Rangecroft, M., & Wallace, W. (1998) Group consultancy, as easy as falling off a bicycle! In Pereira-Mendoza, L., Seu Kea, L., Wee Kee, T., & Wong, W. K. (Ed.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 359-364). Voorburg: International Association for Statistical Education and International Statistics Institute.

Ruberg, S. J. (1998). Consultancy: The two-way street. In Pereira-Mendoza, L., Seu Kea, L., Wee Kee, T., & Wong, W. K. (Ed.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 365-370). Voorburg: International Association for Statistical Education and International Statistics Institute.

Svensson, E. (2001). Important considerations for optimal communication between statisticians and medical researchers in consulting, teaching and collaborative research – With a focus on the analysis of ordered categorical data. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 23-35). Granada: International Association for Statistical Education and International Statistical Institute.

van Belle, G. (1982) Some aspects of teaching biostatistical consulting. In J. S. Rustagi and D. A. Wolfe (Eds.), *Teaching of Statistics and Statistical Consulting* (pp. 343-385). New York: Academic Press.

Gabriella Belli
Virginia Polytechnic Institute & State University,
7054 Haycock Road,
Falls Church, VA 22043-2311, USA.
E-mail: gbelli@vt.edu

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 339-353.

Printed in Granada, Spain.

JUAN D. GODINO, CARMEN BATANERO, & RAMÓN GUTIÉRREZ JAIMEZ

THE STATISTICAL CONSULTANCY WORKSHOP AS A PEDAGOGICAL TOOL

In this article we present and analyse the results of three related experimental studies: (1) the use of statistics in a sample of mathematics education doctoral theses in Spain; (2) the attitudes towards data analysis and statistical consultancy by doctoral students in education; (3) the future statistics consultants perception of their competence for consultancy work. We also describe a project aimed to implement two didactical devices, which would improve the researchers' attitudes and use of statistics and the future consultants' competence. This project would serve to link together prospective consultants and clients within a Statistical Consultancy Unit at the Faculty of Education.

1. INTRODUCTION

In this work we present information about the statistical training of researchers in mathematics education and about the use of statistics in educational research in Spain. This information, as well as our experience in both teaching research methods courses, and doing co-operative consultancy in educational research projects, led us to realise how difficult it is to carry out their own data analysis for those who have not specialised in statistics, and their need for the co-operation of statisticians.

In Spain, moreover, there is neither a culture that favours statistical consultancy, nor a deserved scientific or economical recognition of this work. It is also necessary to make researchers aware of the resources and possibilities that statistics can offer them, and of the specific technical skills and knowledge required to carry out good statistical analyses. On the other hand, students training in statistics do not have the possibility to do practices in companies, and, therefore, their training is mainly theoretical.

In synthesis, in this paper we analyse the following points:

1. The use of statistics in mathematics education research and the impossibility to provide doctoral students in education with a statistical training that allows them to solve all their data analysis problems. Given the complexity and constant evolution of advanced statistics techniques, it is unrealistic to expect that potential users are able to make an autonomous and appropriate use of statistics.
2. The perception by doctoral students in the field of education of their data analysis needs, knowledge and capabilities.
3. The perception by students training as professional statisticians (in Spain) of their capacity for consultancy work, that is, for solving real data analysis problems in co-operation with researchers in other disciplines.

These points show that the statistics curricula at both graduate and postgraduate levels should be revised to develop a culture favourable to statistical consultancy among

statistics users and suppliers of statistical services. We analyse the educational problems implied by statistical consultancy, in relation to both the practical training of future statistical consultants, and to the creation of positive attitudes towards consultancy on the part of future researchers.

We finally describe plans for implementing two interrelated didactical tools that might contribute to solving these problems: (1) the Statistical Consultancy Workshop (directed to trainee statistics students), which is conceived as an optional subject where students would contact real applied research problems; (2) the Statistical Consultancy Units where these students would practice, guided by a Tutor. These units are conceived as meeting points for future clients (applied researchers) and consultants (statistics students), where applied researchers' positive attitudes towards statistical consultancy and trainee consultants' practical experience would be developed. We analyse the basis and characteristics of both didactical tools, and the human and material resources needed to implement this project. We also present the results of a survey carried out to the different agents implied in this project, as regards their attitudes and preconceptions about statistical consultancy.

2. THE USE OF STATISTICS IN MATHEMATICS EDUCATION RESEARCH: AN EMPIRICAL STUDY

2.1. THE STUDY

In this section we analyse the use of statistics in a sample of doctoral dissertations in mathematics education carried out in Spain, to show that educational researchers need the statistician's support in the analysis of their data and in the interpretation of their results. We chose this area because of our knowledge of the specific research problems and our own experience in teaching research methods and in doing statistical consultancy in this area in the University of Granada and other Universities in Spain. On the other hand, we think that the problems related to the use of statistics found among these researchers, who have an advanced training in mathematics, will be higher among researchers in other areas of Education, where the training in mathematics and statistics is weaker.

In Spain, doctoral programs in mathematics education were started in 1988 at the Universities of Granada and Valencia. In addition to these specific programs, doctoral dissertations in this area of knowledge have been carried out in mixed doctoral programs (sciences and mathematics education), pedagogy, general didactics or mathematics. The production of theses has concentrated in the period 1991-2000, since mathematics education was not recognised as an official area of knowledge until 1983, and the doctors who started the first doctoral programs had carried out their doctoral theses in other branches of mathematics, like statistics or analysis. The majority of students entering the doctoral programs have graduated in mathematics, and have taken one or more courses of mathematical statistics during their training. Many of them have also taken a course of statistics applied to education.

In the doctoral program at the University of Granada, we offered four methodological courses in the period 1988 to 2000, one of which (course 1) was compulsory and the remaining were optional. The content of these courses, strongly centred in the specific problems of the area of knowledge, and which lasted 30 hours each, is summarised below:

1. *Research methods in education.* The research process and its main stages. Research paradigms and agendas in mathematics education. Different approaches to research. Bibliographical sources in mathematics education. Methods and techniques of data collection.
2. *Educational research design.* Research variables and hypotheses. Types of variables and their control. Basic types of experimental and quasi-experimental design. Their application to educational research. The process of causal inference. Validity and reliability.
3. *Analysis of educational data.* Coding and preparing data for analysis. Exploratory data analysis. Introduction to inference. Correlation and regression. Analysis of variance. Introduction to the use of SPSS.
4. *Applications of multivariate data analysis to educational research.* Geometrical representation of multivariate data. Classification methods. Factorial methods. Correspondence analysis. Implicative analysis.

In practice, most students took some of the courses 2, 3 and 4, besides the compulsory course 1, although few students took the four courses, due to the fact that there were a limit of 32 credits in the Program, including 9 credits of research work. The situation in the new Doctoral Program is getting worse, since the maximum number of credits that a student can take is 20, courses 3 and 4 have been suppressed, and both courses 1 and 2 are optional. In other doctoral programs in education, the number of methodology and data analysis courses is still more reduced.

2.2. METHODOLOGY USED TO ANALYSE THE DOCTORAL THESES

The analysis has been carried out in a total of 25 doctoral thesis in the area, from the Universities of Barcelona, Cadiz, Granada, Huelva, Basque Country, Santiago de Compostela and Zaragoza. For each of them, we analysed the methodological approach, type and size of samples used, research design, data analysis methods, their adequacy for the problem being researched and the interpretation of their results.

Most researchers took intentional samples of students or teachers, and in 13 theses one or several pilot samples were also analysed. One work of research used a multi-stage-stratified sampling of schools where a proportional number of schools were selected among all the schools in each stratum of a Spanish region. The number of children in each school was also considered in the second stage of sampling, so that the final sample was random. In another thesis a sample of 60 schools was taken, and a very complete statistical study of the sociological and economical variables that define the sample was done. In our opinion, the sample was a good representation of the population studied.

Two theses only performed a study of documents and four of them used a qualitative approach, with scarce use of statistical methods (just a descriptive study). The sizes of the samples were very variable, ranging from the study of 3 cases up to a sample of 1904 students. In general, the number of variables analysed was quite extensive, and thus, even with a sample not too big, the data set was quite large.

The software used was, generally SPSS, although some more specific software were also used, such as Bilog, Lisrel, CHIC, Iteman and Statworks. Auxiliary software were built for collecting data (for example, to record the students' interaction with the computer when solving a task), and we even found a case were a scalogram analysis was carried out with the help of a calculator. Many theses combined more than one

statistical software, since the variety of methods used was not available in just one package.

The frequency of use of different statistical methods is presented in Table 1, where we can appreciate the difficulty of preparing the doctoral students in such a wide range of methods within the scarce time available for data analysis courses. Sometimes, the statistical analysis was carried out by statisticians who participate in the research teams (this is our case for some of the theses analysed).

Table 1: Frequency of Statistical Methods Used in 25 Doctoral Theses

Statistics methods used	(1)	(2)
Descriptive study	22	2
Item analysis: difficulty and discrimination indexes	17	
Validity, reliability	11	6
Item response theory, generalizability theory	7	2
Association: contingency tables, association coefficients; correlation	5	4
Hypothesis tests, confidence intervals, estimation of effects	3	3
Non parametric inference: Kruskal - Wallis, range correlation	5	4
Lineal models: anova, ancova, regression; multiple regression; repeated measures	5	3
Lineal logarithmic models	1	2
Lisrel models	1	
Analysis cluster	5	1
Descriptive analysis of variables deduced from content analysis of texts	7	
Factor analysis, principal components	5	5
Correspondence analysis	7	1
Scaling (Guttman scalogram)		1
Multivariate inference: T2 Hotelling, manova, mancova	4	2
Experimental design	6	3
Total	111	39

(1) Correct use; (2) Errors in analysis or interpretation

In other cases the analysis was performed by external statisticians, specialists in educational research methods with no specific statistical studies, or by the doctoral students, who learned to operate the statistical software and carry out the analysis of their own data.

In relation to these last two cases we observed an excess of use of statistical methods, consisting of testing a given hypothesis or analysing a variable by several logically equivalent methods that produce the same result. For example, in some theses the reliability was computed, by the two-halves method, Alpha and Theta coefficients, and by generalizability coefficients deduced from analysis of variance with repeated measures. Of course the four values obtained only varied in the last decimal figures. Another example is using correspondence analysis to study the association between rows and columns in a contingency table with only 3 columns, where the associations are observed by simple visual inspection. This suggests the researchers' lack of confidence in the method chosen or recurring to statistics in order to assure the "scientific status" in a work of research that, otherwise, would scarcely be relevant.

Another fact observed is that, statistical methods were sometimes adapted to the specific field of education and quite a thorough knowledge of this field was required on the part of the data analyst. For example, some works used correspondence analysis to relate students' strategies or errors with task variables in the items of a questionnaire.

This questionnaire was elaborated with the help of experimental design in such a way that projecting certain supplementary variables in the correspondence analysis could serve to carry out informal hypotheses tests about qualitative data (Batanero, Estepa, & Godino, 1995). Another example was using implicative analysis to show the evolution of students' conceptions as a consequence of a teaching experiment (Batanero, Godino, & Navarro-Pelayo, 1995). Experimental designs such as Latin- square, grecolatin-square, or factorial fractional, were used in several works to control task variables in the questionnaires or to divide an over-sized questionnaire into equivalent parts that were distributed to diverse groups of students. The analysis of variance with repeated measures was used to compute different components of generalizability.

2.3. SOME DIFFICULTIES OBSERVED

The difficulty that the correct use and interpretation of statistics implies for researchers is shown in the current controversy around significance tests (Harlow, Mulaik, & Steiger, 1997; Batanero, 2000), as well as in the students' difficulties in understanding even elementary statistical methods and concepts (Batanero et al.., 1994; Vallecillos, 1999). In spite of the small number of works analysed, we found a variety of incorrect uses and interpretations of statistical methods and results, some of which had been previously described in research literature, particularly in mathematical education (White, 1980; Menon, 1993). Below we describe some of these errors and their implications for the quality of research work.

Association

Computing several association coefficients between two variables, without discussing their relevance and taking the coefficient with the highest value to argue the existence of association.

Analysis cluster

Using unclear reasons to determine the number of clusters, when neither the dendogram nor the distances serve to clearly differentiate a given number of them. Arguing, based on the said clusters, that a given number of subject typologies were determined, even when the association coefficients between the dependent variables and the groups identified are near to zero.

Using a contingency table crossing each task with a list of strategies to carry out a cluster analysis of students' strategies in a series of tasks, instead of using the original data set. That is, significantly reducing the data set and using the percentage of each strategy in each problem, instead of each student's strategy in each problem

Contingency tables and logarithmic lineal models

Not taking into account the minimum advisable expected frequency in the table cells. For example, we found a Chi-square test of a large contingency table where only 10% of the cells had frequency greater than 5, and 20% of the cells were empty.

Using logarithmic lineal models in contingency tables where the hypothesis of independence among diverse dimensions in the table did not apply. The data were obtained from several samples, each of which provided data for a certain combination of categories for each dimension. Deducing from the analysis that there was an interaction among the different dimensions.

Factor analysis

Applying factor analysis to data sets with too few cases in relation to the number of variables analysed (less than 2 cases per variable in a thesis), without noticing that correlation coefficients have very wide confidence intervals in small samples.

Oblique rotation was used to get a number of factors bigger or smaller than the number of factors obtained in the initial extraction by principal components, as a way to show the validity of a questionnaire. For example obtaining only one factor by principal components and using the oblique rotation to justify that there were two differentiated factors, even when the two factors were correlated and the variance explained by the second factor was very small.

Confusion among the total variance explained by a factor and the variance explained in the reduced factorial space. In this way a researcher interpreted that a given group of factors explaining 70% of the variance before rotation, could explain 100% of the variance after rotation. It is symptomatic that these errors appear in doctoral students with a high mathematical preparation, who previously studied analytical geometry. The relevance of the context in the understanding of concepts is shown in these examples. None of these researchers would doubt that a rotation of a solid in the space preserves the solid form (number of factors) and relative dimension of each axis (contribution to the explained variance).

Variance analysis

Confusing random effects (for example the effect of different schools) with fixed effects. In particular not keeping in mind the lack of robustness of random effects models when the distribution are no normal or the variances are too heterogeneous.

Finding a significant effect of the school on the dependent variable and considering all the schools equivalent when carrying out inferences in the remaining analysis.

Carrying out a series of one-way variance analysis to study the effect of a series of factors, instead of a factorial analysis of variance or a multivariate analysis of variance.

Scaling techniques

To show that a given competence increases in levels, one author applied scalogram analysis to a questionnaire with 16 tasks that was designed to assure the validity and to control the relevant variables. Since the complete questionnaire did not fit a scale pattern, he recursively applied scalogram analysis and suppressed tasks, until he only kept 4 of the initial tasks. Even then two different factors with a very close proportion of explained variance appeared in factor analysis, yet the author thought he had proved the unidimensionality of the construct studied.

Experimental design

Few studies used experimental design to explicitly control concomitant variables and to assure a better possibility of generalising the results. Since the samples are, generally, intentional, it is important to study the instrument validity and to analyse the type of tasks to which the results could be generalised in students similar to those who participate in the study. Out of 20 researchers who prepared their own questionnaires, only 9 of them controlled the task variables with experimental design and only 17 of them analysed the instrument validity and or reliability.

The methodology of quasi - experimental design suggested by Cook and Campbell (1979) should also be considered to make inferences to other students or to evaluate the effectiveness of instruction. This is based on controlling the threats to validity and on analysing the changes' patterns. Ten out of the 25 doctoral theses implemented teaching

experiments; seven of them did not use a control group; only in one thesis the study suggested by Cook and Campbell was carried out and in two of them the initial knowledge of the participant sample was compared with a bigger sample. Finally we also found a comparison between two groups, in one of which there was a year of instruction before the pre-test was applied.

Inference

One author interpreted significance level in a hypothesis test as "probability of error." Another researcher thought he proved the null hypothesis, because he had found a non-significant result. We have also found confusion between probability and percentage.

In the later theses the study of effects size in statistical tests was included to substitute or supplement the classical T or F tests. We think that this change is due to the current controversy regarding hypothesis tests and to the recommendations to improve its use (Harlow, Mulaik, & Steiger, 1997; Levin, 1998). However, after computing the effects size, some of these authors only analysed whether the effect was significant or not, and only reported the effect's p-value, not doing any analysis of its practical significance. In this way, a method criticised (only reporting the T or F value and its significance) is changed by another equivalent method, that in fact, presents the same philosophical problems.

The problem of multiple comparisons was also found in these theses, where the significance level that would be necessary to carry out a great number of tests on the same sample was usually not taken into account. Finally we found a wrong formula for the sampling error of the mean.

Sampling

*C*onsidering random an intentional sample of children from the same school, from which a random subsample was selected, after eliminating the students that showed learning problems. Considering that a sample size of only 12 cases is enough to apply methods, where the basic assumptions do not hold, but that are robust enough for big samples; confusion between conglomerate and stratified sampling; using stratified sampling, with a very variable size of sample in heterogeneous stratum (regarding the variable measured, for example, different educational levels) and not keeping in mind the subsamples or strata sizes in computing global estimates.

All these results show the difficulty that researchers who have not specialised in statistics find in carrying out their own data analysis and the consequences that this might imply for the quality of their research work. The difficulties found in the theses analysed increase in other related areas, such as language education, psychopedagogy, psychology, science education and physical activity education, where we have sometimes carried out consultancy work. Therefore, it is necessary to recognise that mastering advanced statistical concepts and methods to guarantee their appropriate and pertinent use in solving real data analysis problems is the competence of professional statisticians.

3. ATTITUDES AND NEEDS IN RELATION TO STATISTICS BY STUDENTS IN THE DOCTORAL PROGRAMS IN EDUCATION

Besides analysing the use of the statistics by educational researchers, we aimed to

evaluate the extent to which future researchers were aware of their limitations and to which they value the relevance of the consultant's work in relation to the final quality of their research work. To assess these attitudes we gave a questionnaire to 50 doctoral students in mathematics education (10), sciences education (8), pedagogy / psychology (17), and other educational field (15). 20 of them had good previous training in statistics (2-3 statistics courses in their undergraduate studies), 11 had some training and 19 had no previous training in statistics. 28 students were still defining their research project and the remaining were in different stages of collecting or analysing data.

Analysis and discussion of results

A first set of questions were intended to evaluate the researchers' attitudes regarding data analysis and how they valued the consultant's work (Table 2) .

Table 2: Frequency of Answers to Questions about Researchers' Attitudes Towards Data Analysis and Consultancy

Researchers' attitudes in relation to data analysis	Yes	No	No answer
Have you carried out some statistical analyses in your previous research?	28	21	1
Do you plan to carry out some statistical analyses in your current research?	46	3	1
Do you consider that your statistical knowledge is enough to carry out your data analysis and to interpret your results yourself?	2	46	2
Would you be willing to request collaboration from a statistical consultant?	48	2	
Would you invite a statistician to co-supervise your thesis ?	36	10	4
Would you invite the person who did the data analysis to co-author some derived publications?	26	20	3
Do you consider that your training allows you to properly set your problems to a statistician?	25	19	6

Table 3: Researchers' Subjective Perception[a] as Regards the Need of Statistical Help

Research stages and methods of data analysis where help is needed	Median	Interquartile range
Research design, identifying variables, selecting samples	1	1
Identifying the statistical techniques appropriate to the problem	2	1
Coding and recording data	2	0
Producing descriptive univariate tables and graphs	2	1
Producing descriptive bivariate tables and graphs	2	1
Computing statistical summaries (central position, spread, shape)	2	1
Studying association in simple or multiple contingency tables	2	1
Correlation analysis and simple or multiple regression	2	1
Variance and covariance analysis	2	1
Fitting of distributions of probability	2	1
Estimation and hypothesis testing	2	1
Time series and longitudinal studies	2	1
Multivariate methods (cluster analysis, factor analysis, etc.)	3	1
Analysis of questionnaires and studies of reliability	2	1
Using statistical software	2	1
Interpreting the results of the statistical programs	2	1
Writing the report	2	1

[a] Responses based on a 4 point scale from 0=not needed to 3= very important

Most subjects were planning to carry out some statistical analyses in their current research and many of them had also done some data analysis before. We note the high percentage of students (40%) that do not consider necessary to include the statistician as an author of the publications derived from the consultant's contribution, which suggests a scarce recognition of the knowledge contributed by data analysis experts. We believe that, even when consultants are paid for their work, they do not lose the responsibility and property of the scientific knowledge produced by their collaboration. We also gave the doctoral students a series of questions to assess (in a scale 0-3) their subjective perception about their need of help for doing specific data analyses. In Table 3 we present the average values for each question in the whole sample.

The median score was 2 or greater in all the questions, except in “research design, identifying variables, selecting samples”. This suggest that doctoral students are aware of their needs for collaboration from a statistical consultant, especially in the most advanced methods.

4. A DIDACTICAL PROPOSAL: THE STATISTICAL CONSULTANCY WORKSHOP AND UNITS

Belli (1998) analyses the consultants' work features, stressing their educational role: Consultants have to make users understand their data limitations and possibilities and the requirements of research designs to reach the intended conclusions. A process of collaboration among educational researchers and statistical consultants incorporated in the research teams would significantly reduce part of the problems and needs described.

We conceive statistical consultancy as a device to co-operatively study data analysis problems, and therefore as a didactical system. Consultants need the client's contribution, as much as clients need consultant's knowledge, and both of them require some adequate structure supporting this work (material resources, knowledge, attitudes, etc.). It is clear that the circumstance that leads to statistical consultancy is the lack of enough statistical knowledge on the part of the client. However, a minimum knowledge is required to value and be aware of the necessity of the expert's collaboration.

In the same way, the consultant should know enough of the application area to guarantee the mutual communication and understanding. Barnett (1988) indicates that the consultant should be a solver of problems posed by another person, and therefore he/she should be a translator and a communicator: he/she should understand enough of other disciplines to appreciate the problems, must express them in statistical terms and what is more important, to communicate the results in an understandable way.

We think that the practical training of professional statisticians should include specific courses where the diverse contents and necessary skills for the application of statistics are approached in a systematic way. Even when the student was finally directed to teaching, the practice of consultancy carried out would have a positive impact in his teaching, as he could use real cases of application of statistics to make more attractive and less abstract his/her lectures (Wisenbaker & Scott, 1998). Consultancy courses should have an essentially practical orientation and be based on the philosophy of workshops and seminars, with the support of a network of centres where practical work would be carried out.

A bibliographical survey in statistical education reveals that the various aspects of statistical consultancy has been analysed by different authors. A first and important issue analysed is the consultants' training. For example, Rangecroft and Wallace (1998) describe the process of consultants' preparation at Sheffield Hallam University, which is

based on an intense year of practice in companies, after the student has carried out some team works in the second year of University studies. Ruberg (1998) describes the training of statistical experts within a multinational pharmaceutical company, where graduate statisticians form part of multidisciplinary research teams.

Not all these contents are currently considered in the curricula for training statisticians in Spain. The main difficulty in organising practice for future statisticians in some regions like Andalusia is in fact the shortage of suitable companies in the University geographical area. Since small companies do not invest many resources in research, students of statistics (and other university specialities, such as economy) are prevented from having periods of practice as an main component of their preparation.

In cities like Granada, the University is in fact the biggest company and the place where there are more opportunities to use statistics. The University is also the place where the professionals who will need the consultant's collaboration are trained. Therefore, we consider that the conditions that facilitate the meeting and mutual exchange of these potential data analysis clients and consultants should be created within the University. We notice that in other university specialities like, teacher training, engineering, interpretation or computer science, students can practice in companies or other institutions.

Based on these contextual factors, we describe a project for implementing two interrelated didactical tools aimed to improve the application of statistics by researchers: (1) A course offered to statistics students, which is conceived as a Workshop of Statistical Consultancy (WSE); (2) A series of Units of Statistical Consultancy (USC), which will be started up at the different University Faculties where experimental research is developed; these units would be assisted by the students entering the WSE, who will be supervised by an expert statistician in the role of Tutor.

This WSE workshop could be offered as a free configuration subject, since in the Spanish University curricula students can optionally take a percentage of credits from other Faculties or degrees. Therefore, the WSE could be offered in specific faculties like the Faculty of Education or Medicine to provide future consultants with a specialisation in the use of statistics in education, medicine, psychology, etc. Students would receive a practical training, and would receive some payment from Projects, Research Groups, or different types of scholarships (collaboration grants, teaching assistantships, etc.). Another possibility is that the work carried out in the statistical workshop could be valued as a part of the future statistician's graduation projects or Master's thesis.

This workshop would be supplemented by the creation of Statistical Consultancy Units (SCU) within some University centres for graduate degree students who would in fact be assisted by the statistics students entering the WSC. We think that future researchers and future statistical consultants should meet at the University centres and departments, where collaboration between them should be started.

One problem hindering the use of consultancy and the incorporation of statistical experts in research teams at Universities is the reduced size of such teams. At the University, where the attitudes and work habits are modelled, research teams are frequently formed by only a doctoral student and his/her thesis supervisor; or at best the research group is constituted by a few researchers. Consultancy has to be faced at a higher level, either the Department or Faculty, or even at the whole University. Consultancy work at the SCU would be supervised and assessed by professors or lecturers, who would be selected among professionals with experience in statistical consultancy. The SCU would in addition be a good resource to favour the development of positive attitudes toward statistical consultancy, and an opportunity for statistics students to practice their theoretical knowledge.

Below we describe the objectives, contents, procedures and resources needed for starting up these two didactical devices.

4.1. OBJECTIVES

With this project we intend to reach the following purposes:

1. Creating a network of centres where trainee statisticians could practice statistical consultancy.
2. Favouring the development of a positive attitude towards statistical consultancy among future applied researchers.
3. Reinforcing the future researchers' culture of working with statistical experts, through consultancy and/or by including expert statisticians into research teams.

4.2. CONTENTS

The course content would be essentially practical and would be organised in two components: a) student's practice of co-operative data analysis of real problems, inside research teams; b) Seminars guided by the workshop supervisor, where students would present their work in progress to their classmates and the difficulties found would collectively be discussed by the students and the lecturer.

It would also be interesting to organise specific sessions directed at increasing the students' knowledge of the following non-statistical knowledge, which is basic for doing consultancy and for practising statistics: Defining the data analysis problems; interviewing the client; identifying questions of interest, presenting and discussing results; communicative strategies; psychological and educational aspects of consultancy work.

4.3. PROCEDURE AND RESOURCES NEEDED

Besides the SCW course, we intend to start up a *statistical consultancy unit* (SCU) at the University Faculties where experimental research requiring data analysis is being carried out by diverse departments and groups. These units would support, in particular, the data analysis of research carried out by doctorate students.

Human resources

At least two students training as statisticians would take over the responsibility of each SCU, during their period of practice or when holding a scholarship. These students will be supported by the lecturer responsible for the SCW and the lecturers of the remaining subjects, who will answer any doubts during tutoring time. The users requiring the services of the SCU will also be supported by their respective tutors and theses' supervisors.

The lecturer responsible for the SCW would co-ordinate the different SCU within the University, and this work (teaching, advising and evaluating the consultancy work carried out) would be recognised as teaching credits. This lecturer should have specialised in statistics and have a wide experience in consultancy within the Faculty where the SCW is functioning.

Material resources

In each Faculty a physical, identified space would be prepared to host the SCU. The University computing resources would be used, either at the computer labs or in the research support units at each Faculty. The bibliographical resources at the different Departments and University libraries would also be available.

Academic frame

The consultancy work carried out by a student training as statistician will be valued with credits in his/her academic record, as the workshop can take part in the curriculum as an optional subject. If enough resources were available the course might also be offered to students majoring in other university careers, whenever they have an appropriate previous statistical formation (e.g., Economy and Psychology).

5. EVALUATION OF THE PROJECT BY TRAINEE STATISTICIANS

It would not be possible to carry out the project presented without an interest on the part of students training in statistics, who should take the responsibility for an important part of the consultancy work inside the SCW, guided by the tutor. To evaluate the interest in the project on the part of these students, we carried out a study of their attitudes towards the consultancy work. In particular we were interested in their interest in doing practice and we wanted to know if they considered themselves well enough trained to perform consultancy work in their future professional life.

With this aim, we gave a questionnaire to 43 students in their 4º year of statistics studies at the University of Granada. All of them were interested in enrolling in the "Statistical Consultancy Workshop", in case the project was carried out, assigning between 4.5 and 10 credits ($x = 6.48$, $s = 1.16$) to the course.

Table 4: Future Statisticians' Subjective Perception of their Capacity for Consultancy Work[a]

Student's preparation in diverse aspects of consultancy work	Median	Interquartile range
Knowledge of the fields where statistics is applied	1	1
Main stages in experimental research and the role of statistics in the same	1	0
Research types and approaches (exploratory/confirmatory; transversal/longitudinal; experimental / quasi-experimental	1	1
Philosophical and ethical problems in the application of statistics	1	1
Working in teams to solve data analysis problems	1	1
Dealing with clients and understanding their problems	1	1
Research design, identifying variables, selecting samples	2	1
Identifying the appropriate statistical techniques	2	1
Using the software needed to perform the statistical methods	2	0
Interpreting the software results	2	0
Writing the report	1	1
Presentation of reports to clients or in an audience	0	1
Oral or written communication of reports	1	1

[a] Responses based on a 4 point scale from 0=not needed to 3= very important

Only 12 of the students had had the opportunity to carry out practical work of data analysis for companies and only 3 of them considered that their studies qualified them well enough for their future functions of statistical consultants. In Table 4 we present

the median and interquartile range of the scores given by the students to the different items in the questionnaire, valued in a scale 0 (nothing), 1 (little), 2 (enough) to 3 (much).

The students training in statistics considered themselves to have “little” preparation to confront real problems, except by using the software, interpreting the results, identifying statistical techniques and working in teams. These students receive a high technical preparation, that should be reinforced with additional practical knowledge about ethical and philosophical problems, dealing with clients and understanding their problems, producing and communicating reports, points that received a punctuation equal to 1 or even 0 and that are needed for their future work

6. DISCUSSION AND IMPLICATIONS FOR TRAINING RESEARCHERS AND FUTURE STATISTICIANS

A realistic vision of researchers' training should recognise the complex character of statistical knowledge, even when increasing the teaching time and improving the didactical resources. It is difficult for researchers who are not specialised in statistics to acquire a complete mastering of statistical concepts and methods, beyond the most basic content, or that which becomes familiar due to its frequent use. The time assigned to study statistics in undergraduate and graduate courses in human, social, scientific and technological careers is too limited to prepare these investigators to be self-sufficient in solving their data analysis problems. As Belli wrote:

> *"graduate students in applied areas had some introductory background in statistics, but lack many of the skills to actually analyse the data they collect"* (1998, p. 344).

We also observe the need to encourage a more positive attitude among researchers in human, social, scientific and technical disciplines towards statistical consultancy. Many of these potential users are unaware of the necessity to consult statisticians or to incorporate statistical experts in the research teams, and they do not value the difficulty of this work. As a consequence, the statistical training curricula should take into account this problem, and try to make the users aware of their own limitations. We consider it necessary to increase the team work culture among researchers, by integrating the applied statistician's role in the design and data analysis phases of experimental investigations and by fostering statistical consultancy.

At the same time it is necessary to revise the statistical experts' training program (in mathematical or applied disciplines) to qualify them for developing co-operative work and consultancy. In the case of many Spanish universities, we observe that future statisticians do not carry out adequate practical work to apply their theoretical knowledge.

> *"Lack of consulting experience, which is a chance to apply the tools of statistics, leaves them without an appreciation for the artistic side of statistical reasoning. They learn the formulas and theories of analyses, but have no opportunity to become involved in planning a study, gathering real experimental data, nor having to present the fruits of their analyses to professional audiences".* (Bentley, Schneider, & Bentley, 1998, p. 347-48).

The project of Statistical Consultancy Workshops and Units can serve as a meeting point for potential clients and suppliers of statistical consultancy and can follow the

example of this type of consulting services in other countries (as in the examples discussed in Belli, 2001; Jolliffe, 2001, Ospina, 2001). This will allow students of experimental disciplines (humanities, social, sciences, technology) to appreciate the benefits of using statistics in a reasonable way. For statistics students, this project can serve to learn the bases of consultancy and to appreciate that

> *"The problems faced by statisticians acting as consultants are varied, not only by the origin of data and research questions posed, but also by the type of personal abilities required to solve them with success. Communication with researchers with insufficient statistical and mathematical preparation is an arduous task, and sometimes, frustrating, and it will require in good measure a dose of patience and tolerance. Nevertheless, working within a team to solve practical problems can be very exciting, and contribute a great intellectual recompense.* (Hand & Everitt, 1987, p. 9).

ACKNOWLEDGEMENT

This research has been supported by the grant BSO2000-1507 (Ministry of Science and Technology, Madrid).

REFERENCES

Batanero, C. (2000). Controversies around the role of statistical tests in experimental research. *Journal of Mathematical Teaching and Learning,* 2(1& 2), 75-97.

Batanero, C., Godino, J. D. Green, D. Holmes, P., & Vallecillos, A. (1994). Errors and difficulties in understanding statistical concepts. *International Journal of Mathematics Education in Science and Technology,* 25(4), 527-547.

Batanero, C., Estepa, A., & Godino, J. D. (1995). Correspondence analysis as a tool to analyse the relational structure of students' intuitive strategies. In R. Gras, & M. Artigue (Eds), *Colloque "Méthodes d'analyses statistiques multidimmensionnelles en Didactique des Mathematiques."* (pp. 155-166). Caen: A.R.D.M.

Batanero, C., Godino, J., & Navarro-Pelayo, V. (1995). The use of implicative and correspondence analysis for assessing pupil's combinatorial reasoning. In R. Gras, & M. Artigue (Eds.), *Colloque "Méthodes d'analyses statistiques multidimmensionnelles en Didactique des Mathematiques."* (pp. 245-256). Caen: A.R.D.M.

Barnett, V. (1988). Statistical consultancy. A basis for teaching and research. In R. Davidson y J. F. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics* (pp. 303-307). University of Victoria.

Belli, G. M. (1998). The teaching aspects of consultancy. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (Vol. 1, pp. 341-346). Singapore: International Association for Statistical Education and International Statistics Institute.

Belli, G. M. (2001). The teaching/learning process in University statistical consulting labs in the United States. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 325-338). Granada: International Association for Statistical Education and International Statistical Institute.

Cook, T. D., & Cambell, D. T. (1979). *Quasi-experimentation. Design and analysis issues for field setting.* Chicago: Rand Mc.Nelly.

Hand, D. J., & Everitt, B. S. (1987). Statistical consultancy. In D. J. Hand, & B. S. Everitt (Eds.), *The Statistical consultant in action* (pp. 1-10). Cambridge: Cambridge University Press.

Harlow, L.L., Mulaik, S. A. & Steiger, J. H. (Eds.), *What if there were no significance tests?*

Mahwah, N.J.: Erlbaum.

Jolliffe, F. (2001) Learning by experience. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 355-370). Granada: International Association for Statistical Education and International Statistical Institute.

Levin, J. R. (1998). What if there were no more bickering about statistical significance tests? *Research in the Schools*, 2, 45-53.

Menon, R. (1993). Statistical significance testing should be discontinued in mathematics education research. *Mathematics Education Research Journal*, 5(1), 4–18.

Ospina, D. (2001). Statistical research and consulting in Universities from developing countries: The Colombian case. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 295-310). Granada: International Association for Statistical Education and International Statistical Institute.

Rangecroft, M., & Wallace, W. (1998). Group consultancy, as easy as falling off a bicycle?. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics*. (Vol. 1, pp. 359-364). Singapore: International Association for Statistical Education and International Statistics Institute.

Ruberg, S. J. (1998). Consulting: The two way street. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics*. (Vol. 1, pp. 365-370). Singapore: International Association for Statistical Education and International Statistics Institute.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher*, 25(2), 26–30.

Vallecillos, A. (1999). Some empirical evidencies of students difficulties in understanding statistical tests. *Bulletin of the International Statistical Institute*, Tome 52, Book 2, p. 201-204. Helsinki: International Statistical Institute.

White, A. L. (1980). Avoiding errors in educational research. In R. J. Shumway (Ed.), *Research in mathematics education* (pp. 47-65). Reston, Va: National Council of Teachers of Mathematics.

Wisenbaker, J. M., & Scott, J. S. (1998). Examining the consultancy -teaching relationship for university based statisticians. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.) *Proceedings of the Fifth International Conference on Teaching Statistics*. (Vol. 1, pp. 371-376). Singapore: International Association for Statistical Education and International Statistics Institute.

Juan D. Godino and Carmen Batanero,
Department of Mathematics Education, Faculty of Education
University of Granada,
Campus de Cartuja, 18071 Granada, Spain
E-mail: godino@ugr.es; batanero@ugr.es

Ramón Gutiérrez Jáimez
Department of Statistics, Faculty of Sciences
University of Granada,
Campus de Fuentenueva, 18071 Granada, Spain
E-mail: rgjaimez@ugr.es

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 371-377.

Printed in Granada, Spain.

BEN CHANG SHIA

HOW TO THINK ABOUT STATISTICAL CONSULTATION? LEARNING FROM DATA

The purpose of this paper is to reflect on statistics teaching and practice. To begin with, I will reflect on the present status of statistics education and its importance. Then I will discuss about seven main statistical topics: 'Descriptive statistics', 'Basic concepts', 'Statistical process control', 'Random walks', 'Introduction to statistical tools', ' How to do data analysis' and ' How to consult'. With the expansion and development of IT and the Internet, it is the e-century now, and therefore statistics education should change. Finally I suggest the idea of "statistical electronic school (statistical e-school)".

1. INTRODUCTION

Nowadays, many kinds of scientific fields need statistical analysis, and statistics plays an important role in many fields of research. Many researchers and other people use statistical methods; however, they do not know what they are using. Here is a simple and essential example about how statistics can be taught.

If you want to inform someone else about the height of your classmates, what should you tell him? I believe you would not give him the height of everyone in your class, because it is worthless and useless to tell others about each piece of data. You only have to provide him with the average height of your class. The average height is a statistic.

Furthermore, you also can present him with the variance or standard deviation of the height of your class to describe the average deviation from the centre (mean) of the data. They are also statistics. From mean and variance (standard deviation), we can roughly understand the average of the height and how it is distributed. After we have found the average and distribution, we know the structure of our data well enough to make any decision. No matter whether we want to decide a marketing policy or to forecast a future trend, we will need the underlying structure in a data set.

Maybe your students will doubt that statistics is so powerful and they will be right. It is not exactly that you cannot make decisions or carry out a prediction without statistics, however statistics will help you to make more accurate and better decisions.

Today we can gather the information and data that we want easily. Before facing up to lots of data, it is better to think about what you want, what you should do and what you are doing. I believe you must realise what you can do with the data. You try to analyse the data and want the data to help you make decisions in management. Unfortunately, you are involved in statistics now.

No matter what kind of analysis and what kind of research topic you are interested in, you will always be involved in statistics. Hence when trying to analyse the data, you are using statistical thinking. And statistics is a very useful tool for data analysis and to help you make decisions.

2. TOPICS ABOUT STATISTICS

Briefly speaking, there are seven main statistical topics: 'Descriptive statistics', 'Basic concepts', 'Statistical process control', 'Random walks', 'Introduction to Statistical tools', ' How to do data analysis' and ' How to consult'.

We know that statistics is an important tool. It can figure out what the problem is, where the problem is, and after repeated experiments, we can see the probable trend. In fact, the point that statistics emphasises is not a completely correct answer but a reasonable field of answers. Before consulting, what we have to do is integrate the other professional fields. Hence, to make a good statistical inference, we should develop a second professional ability.

Therefore, we have to understand how the data are distributed and what structure it has. We can use the elementary statistical techniques such as mean, variance and statistical charts etc to get some ideas about the distribution. From 'Descriptive statistics' we will get the key trend and structure the data. We also need to understand some 'Basic concepts' about data.

And then reasonable, appropriate and further statistical analysis tools should be provided, such as making inference and decisions. 'Statistical process control' and 'Random walks' will help us to figure out where the problem is, what trend we cannot ignore and how to control the uncertainty.

What I have aforementioned are only concepts that we should make operative, so some powerful tools for the calculations should be available. They are statistical software, such as SAS. At the end of completing the consulting process, the results should be presented in many ways, such as graphs, tables, reports and files.

3. RELATIONSHIP BETWEEN DATA ANALYSIS AND MANAGEMENT

To follow, we talk about 'data analysis' and 'management'. First, we must take some definition of 'data' and 'analysis'. 'Data' is any collection of numerical values. It must have some structure in which we are interested and which we do not know yet. 'Analysis' is a process to extract the useful information from data. The typical process of data analysis is 'summary and display of data', 'formulation of tentative models (explanations)', 'fitting models' and 'diagnostic checking of model adequacy'.

Before analysing data, the most important task of all is collecting data. The source of data is 'process'. Now, we must give a definition of 'process'. 'Process' is a number of steps or operations aimed at a single goal. Data is a sample from the process; such as data from a time series.

Then we can make the summary and display the raw data. By means of the summary and display of data, we can have the essential understanding of our information. Then we can establish some models and find useful ones among them for the managers and, therefore, we can achieve the goals of our statistical study. For time series data, we can use time series models for improvement and forecasting of processes.

Today we all know that information is very important. It can help us make the correct decisions and do the right things. Where does the information come from and who generates it? You can get the information from the news or research papers written by someone else. But is it suitable for you? I do not think so. We can use the information collected by others as reference; however we should also create new information by ourselves. Therefore, we must use the process of data analysis to create the information that is more suitable for us and for our decisions.

There is now a new information technology (IT), namely "data mining", which means a series of actions to select, discover and model from the mass of data, and find out the unknown patterns to transfer to the people responsible for decision making in the company and improve the profits.

The series of actions involved in data mining includes five major steps: sample, explore, modify, model and assess. Maybe you will think that data mining is just information technology. However, we can find above the five major steps that are almost always involved in statistics. Data mining can also include regression analysis, decision trees, neural networks and other statistical analyses. The process of data analysis is just the extension of the process I described previously and the kernel of data mining is data analysis. The only difference between them is the target, since the target of data mining is the mass of data from a lot of databases.

On the other hand, what role should a researcher or a business manager play in the statistical analysis? He cannot understand the results of data analysis if he has not been involved in it. A researcher or business manager must think about 'what kind of data to collect and analyse', 'how the data should be analysed' and 'what special inquiries and actions are needed'. In other words, he must provide his professional knowledge for the statisticians. We can make more useful and correct decisions by combining professional knowledge and the statistical analysis process.

To sum up, what is statistical control of the trend? There are different ways of thinking with different variables through time. No matter what way we follow, the most important of all is that we have to understand the trend and control it, to make correct conclusions.

4. FUTURE OF STATISTICS EDUCATION

With the expansion and development of IT and the Internet, it is the e-century now. There are over 250 million people using the Internet. Everyone can use the Internet from the school, office and home easily. We will not be able to live without it in the future. It probably means everyone may become a distance learner. So statistics education should be different.

According to the University Law, education has three tasks, which are teaching, research and service. The roles of Universities are creating knowledge, preserving knowledge, integrating knowledge, transmitting knowledge and applying knowledge. Of course these are also the roles of education. Especially in the e-century, education is no longer restricted to the traditional in-campus type. A new kind of channel should be developed to allow everyone to get knowledge from any corner of the world.

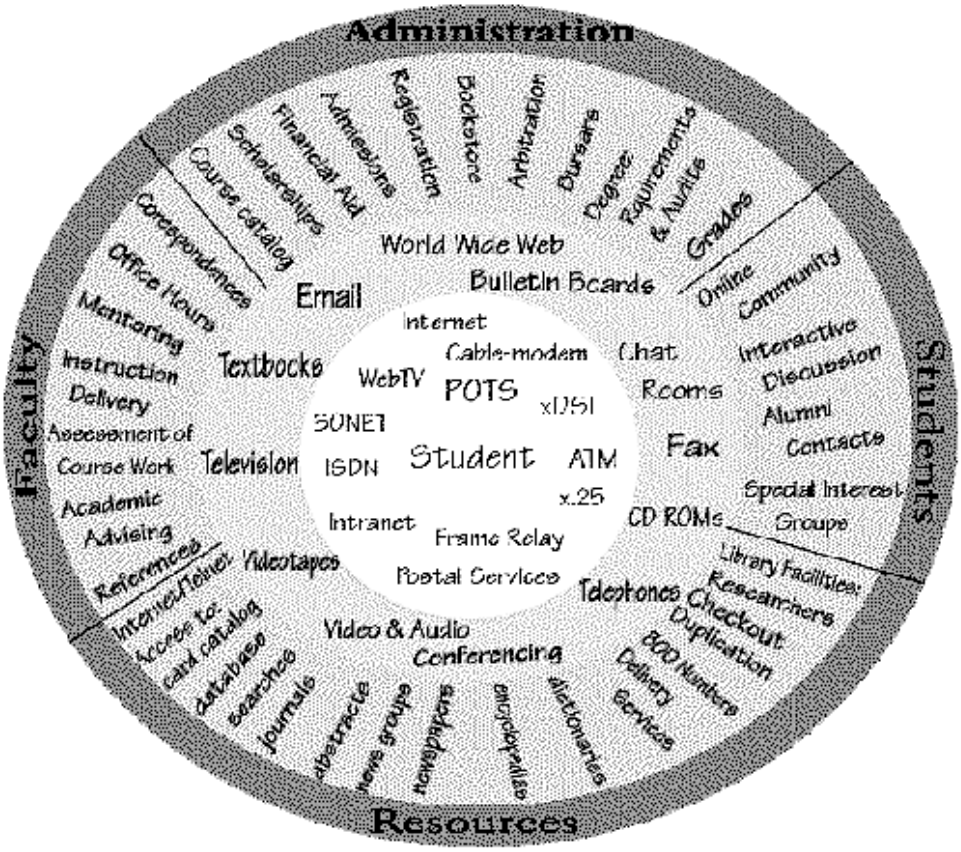
We should also expand the range of education in statistics, that should be combined with the Internet, and diversify the ways and channels of education. In doing this, we will let students from anywhere and from any level understand the importance of statistics and its application. We can then help them acquire the correct understanding of statistics science and so we can achieve the goal of expanding statistics science.

A virtual University will include distance teaching, distance education, online course, asynchronous learning network (ALN), virtual campus and classroom. This new virtual University has many competitive advantages such as responsiveness, accessibility, convenience and quality at a reduced cost (Chen, 1998). Aoki and Pogroszewski (1998) describe the infrastructure needed for providing students with a learning experience and related support services to complete a degree program partially or totally online and for providing faculty members with resources for teaching and

doing research effectively online. They present the Virtual University Reference Model (VURM), which is reproduced in Figure 1. This model is intended to be a guideline or a framework for colleges and universities that plan to deliver instruction and support services to distance learners, and to be a checklist to evaluate existing distance learning programs.

In the model, a virtual university is broken down to four major components: administrative services, student services, resource services, and faculty services. Each component has a different purpose and provides students with different facilities.

Figure 1. Virtual University Reference Model



The second outer ring in the model shows the types of services a student receives from each of the four noted component areas. The inner three rings represent (from the innermost):

1. The student and his or her relationship to each of these four areas;
2. Transmission systems with which the services can be accessed by students;
3. Applications and tools to be used in offering the service elements in the outer ring.

It is important to note that in this model students are the centre of the model and all the service components and elements are depicted in relation to the students (Kumiko Aoki and Donna Pogroszewski, 1998).

So far the WWW is a very efficient tool for presenting statistical information in new ways that contribute to attracting the interest of ordinary people: The classic social facts

(economy, demography, etc) are often presented by exploiting the "technological appeal" as much as possible (Galmacci, 2001). From Galmacci's paper we can realise how many kinds of powerful statistical resources are spread in the WWW, at different web sites in each corner of the world. They should be integrated to a complete and efficient library of statistics education resources in the Internet.

This is the reason why I bring up the idea of "statistical electronic school (statistical e-school)". This new type of educational institution will not be based on the traditional campus and classrooms. In other words, it may not have a physical presence consisting of buildings and departments, but it might exist in the cyberspace (Aoki & Pogroszewski, 1998).

The statistical e-school is not only an e-book database, an introduction site, and the channel where Video Education is provided. The statistical e-school is an integrated site in the Internet world, where we are willing to include: statistical classroom for 3 levels, e-book database, Video Education database, survey and economics database, statistical consulting centre, internet survey and other services. These services include virtual personal assistant, planning learning programs, searching within the e-school site, inquiring about the main library, introducing new and good books, bulletin, game zone, and service interface.

4.1. STATISTICAL CLASSROOM

First level

The aim is to teach students to understand the application of elementary statistics and the meaning of descriptive statistics, such as percentile, percentage, mean, variance and other statistical charts etc. Using interesting examples in ordinary life we will let students learn how to make preliminary analyses.

Second level

The aim is to teach students and researchers to learn more advanced techniques of statistical analysis, such as regression, analysis of variance, time series analysis, sampling survey, contingency table, categorical data analysis, etc. In addition, students can try to use statistical software, such as SPSS, Statistica, SAS, S-plus, etc.

At this level, which is relevant to the professional statisticians and researchers, we will provide the Video Education and hire a teacher to answer the questions.

Third level

This is the upper level of statistical education, including mathematical statistics, Bayes inference, generalised linear model etc. The emphasis is on deriving the theory and teaching correct statistical concepts.

At this level we also let students know how other sciences use statistics, such as sociology, biology, information and industry technology, etc. Like the second level, we will provide Video Education and hire a teacher to answer the questions. And we will set up a group belonging to the consulting centre to solve problems from any other science.

4.2. E-BOOK DATABASES, AND STATISTICAL CONSULTING CENTER

Statistical e-school provides not only e-book and Video Education but also the functions of inquiring and consulting. Hence we will establish the databases for e-book and Video Education. Besides, using the techniques of data warehouse and data mining

we will establish the statistical survey databases, including government survey and economic data.

On the other hand, the e-school can also solve statistical problems through the consulting centre. The members of this centre should include some lecturers and graduate students. Of course, the leader should be a lecturer. In addition, we charge for using the databases and consulting. The fee will be the cost of maintaining this e-school.

4.3. INTERNET SURVEY AND OTHER SERVICES

If the website only contains the above functions, it may not interest all the intended audience. Therefore, the e-school should provide Internet surveys to increase the interest in studying statistics. We will make the topics of these Internet surveys change and will try to find some interesting ones often.

Moreover, the on-line students can use and study in the statistical e-school more conveniently and easily. The e-school provides them with lots of kinds of friendly functions and interfaces, like intranet map, virtual personal assistant, planning learning programs, searching within the e-school site, enquiries about the main library, introducing new and good books, bulletin, game zone and service interface.

5. CONCLUSIONS

The purpose of education is to spread knowledge and to expand statistics education. With the expansion and development of IT and the Internet, we can achieve the target easily by each kind of channel. It is possible to set-up a statistical e-school to provide statistical resources and consultant services for students, teachers, statisticians, researchers, business and general people. When doing this, however, we cannot forget that the essence of statistics is analysing data. Data is coming from everywhere and has any kind of pattern. It contains a lot of information and unknown patterns, that should be discovered by statistical analysis and thinking. The statistical tools of analysis and discovery should also be improved to adapt to each kind of data, and this is why we speak about “learning from data”.

APPENDIX: SOME IMPORTANT WEBSITES

Asynchronous Learning Networks, including Magazine Journal of ALN: www.alnn.org

On Line Journal of Distance Learning Administration: www.westga.edu

Web-Course tools

I-CARE: http://Pride-sun.poly.edu/icare
Socrates Forum: http://www.environmentslintal.com/socrates
WebServer: http://www.madduck.com
WebMentor: http://www.avilar.com
Learnspace: http://www.lotus.com/learningspace
*Top class:*http://www.wbtsystems.com
Webct: http://www.webct.com

REFERENCES

Aoki, K., & Pogroszewski, D. (1998). *Virtual University reference model: A guide to delivering education and support services to the distance learner.* Available online at http://www.westga.edu/~distance/aoki13.html

Chen, W. S. (1998) *Establishment and management of Virtual University*. Taiwan: NKFU.

Consultant Centre of NCKU (1993). *Case reporter*. Taiwan: Department of Statistics, NCKU.

Consultant Centre (1994). *1994 Price survey*. Taiwan: Department of Statistics, FJU.

Consultant Centre (1995). *Price survey*. Taiwan: Department of Statistics, FJU.

Consultant Centre (1996), *Quality survey of Nissan new car.* Taiwan: Department of Statistics, FJU.

Galmacci, G. (2001). The impact of the internet on researchers' training. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 159-169). Granada: International Association for Statistical Education and International Statistical Institute.

Hand, D. J., & Everitt, B. S. (1987). *The statistical consultant in action*. Cambridge University Press.

Lin, H. L., & Chen, C. C. (1996). *Statistics method and application I*. Taipei: Yeh-Yeh Publishing.

Lin, H. L., & Chen, C. C. (1996). *Statistics method and application II*. Taipei: Yeh-Yeh Publishing.

Statistics Science. November, 1995. Taiwan: NSC.

Survey Research. April 1996. Taiwan: Office of Survey Research Academia SINICA.

The Statistical Consultant, (1995) Volume 12, No. 2.

The Statistical Consultant, (1996) Volume 13, No. 2.

Ben-Chang Shia
Department of Statistics, Fu Jen University,
510 Chung Chen Rd. Hsin Chuang, 24205,
Taipei Hsien, Taiwan, China.
E-mail: stat1001@mails.fju.edu.tw

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 355-370.

Printed in Granada, Spain.

FLAVIA JOLLIFFE

LEARNING FROM EXPERIENCE

Training courses for researchers are discussed in some detail. The preparation of researchers and of statisticians for consulting sessions, and the opportunities such sessions provide for training, are considered.

1. INTRODUCTION

Training of researchers in the use of statistics is a large topic and involves consideration of the subject of statistics itself, of the interaction between statisticians and researchers, and even the training of statisticians. In this paper a *researcher* refers to someone researching in any area except statistics, whereas a *statistician* refers to someone with the qualifications and expertise required in a statistical post. See Smith (1993) pages 146-9 for a slightly different view.

Many researchers will have followed a course in statistics as part of their training in their main discipline, but will possibly need some further training in statistics when they are undertaking research. Others might get their initial statistical training while researchers, through formal courses, or, less satisfactorily, by referring to books or by using statistical packages. A further opportunity for training occurs when researchers consult with statisticians. As well as learning something about statistical design and analysis they would learn when to consult and what questions to ask the statisticians. In helping researchers, statisticians learn teaching and consultancy skills and, if presented with a challenging statistical problem when acting as consultants, might learn statistical techniques previously unknown to them or develop new techniques. Thus both researchers and statisticians *learn from experience*, learning from mistakes as well as from successes.

This paper focuses on two aspects of training – on courses in statistics for researchers, and on the consultancy process. It draws on the author's experience in teaching courses and giving statistical advice to students from other disciplines and to researchers. Ideas have also come from a selective review of relevant papers published during the last twenty years. Some of the comments made about the training of researchers in statistical methods and in the skills needed for a successful outcome from a consultancy session apply also to the training of statisticians.

2.1. TRAINING THE RESEARCHER – GENERAL MATTERS

The spread of statistics teaching both at school and at university level in recent years means that more people than previously have at least a basic knowledge and

understanding of statistics and its potential. Thus it is likely that researchers in fields other than statistics will already have some expertise in statistical techniques. It is assumed in this section that this is the case, but many of the points made apply as much to the initial training as to further training – and indeed to the training of specialist statisticians. Training in statistics is, of course, a continuing process that comes partly through experience. Developing statistics from the perspective of the scientific user is "complex and challenging" (McPherson, 1989). He puts forward as important criteria underlying a consideration of statistics for scientists - the need to develop self-sufficiency, understanding of assumptions, and use of meaningful language. Sometimes researchers do not really know the scope of statistics or its relevance to their work, and might even have some wrong ideas. In fact, a study of researchers' errors in understanding and using statistics should help in the design of training courses, as awareness of errors is the first step in helping people avoid such errors. Greenfield (1993) makes some thought-provoking comments on reaching out to non-statisticians, and Hahn (1999), although writing in terms of industrial statisticians, makes many general points of relevance for training.

A course for researchers should give guidance as to a general approach to be taken in a statistical analysis and of how to balance out the effort spent on data collection and on data analysis, on probabilistic versus descriptive methods, and between numerical and graphical techniques (Cox, 1981). Successful courses will encourage a critical attitude towards data and the results of statistical analyses. Researchers need to be warned not to use too elaborate a technique on a poor quality data set, as well as to exploit extensive data sets fully, and to beware of what techniques they use on non-random data or on data collected by a complex sampling method. The importance of obtaining good quality data and of checking at all stages of an investigation – in the field, at coding stage, input to a computer, copying of results – should be stressed. Researchers also need to understand why the design stage in a study is important, the difference between a parameter and an estimate, and how variation can be described by the random part of models (Nelder, 1986, p.118).

The statistician has to some extent been replaced by the computer, so it is particularly important that researchers are made aware of the dangers of misusing statistical packages and of the errors inherent in some routines. References to known errors in Excel's statistical facilities are given on the ASSUME web page (http://www.mailbase.ac.uk/lists/assume/files/). To some extent the analysis which can be done is determined by what is available in the package, and the menu on offer reflects the producer's view rather than the customer's. It is all too easy to over-analyse the data, instead of being selective and to produce large quantities of output, almost as difficult to digest as the raw data (see Preece, 1987 for some comments on this). On the other hand the ease with which interactive analysis can be done, enabling each step and alternative steps to be studied before the next is taken, should not be under-rated.

Use of computer packages has enabled more efficient and sophisticated analyses to be undertaken, but the downside is that users will not necessarily have the theoretical background to understand what has been done, and it is more difficult to communicate the results than in those arising from more elementary techniques. Even if the technique itself is difficult to understand, it is important to help the researcher to understand and question underlying assumptions, and to appreciate the implication of the results of an analysis in practical terms. This should mean that in turn the researcher would report the results correctly and in an understandable way to others. Pictorial representation of

results, when this can be done, is made easier by computer packages, and helps in communication (Bradstreet, 1999).

Statistics educators are in broad agreement that it is important to teach concepts rather than recipes. There have been many important contributions relating to this – research into different methods of teaching, suggestions of strategies to follow, and development of software and other learning material. There are, however, difficulties to be overcome, not least of which are the time constraint imposed on many courses at school and university, and the need to assess students on demonstrable skills. There is also something of a reluctance on the part of students to think about underlying concepts. Researchers are often busy people and possibly even more likely than students at a lower level to resent what they see as a waste of time. An attempt should be made, however, and here too the computer has made things easier for the teacher. For example, the modules produced by the STEPS consortium, designed to be used in conjunction with lectures to non-specialists are useful and interesting. They are based on real problems taken from biology, geography, psychology, and business. They enable the user to explore and analyse data, and can be worked through by students on their own.

Examples showing the importance and relevance of statistical methods to the researcher's area of interest, preferably based on problems the researcher faces (Hirotsu, 2001) are essential. These should, as far as possible, be real rather than contrived, and with the large amount of data available via the Internet, there can be no excuse for not getting students to work on real data. Here too the computer has changed what is feasible – computation and dealing with relatively large amounts of data are not the great problem now that they were at one time. However, some of the time which previously might have been spent on statistical concepts, now has to be spent on how to use software (and sometimes hardware in the case of those who are not computer literate). Reference to real examples should also help researchers understand the use of the more complex statistical methods. Some examples might be taken from articles in the researcher's area of interest. Helping researchers understand the results of statistical analyses in such articles is an important function of training.

Models are important in statistics and have a central role in the theoretical development of the subject, but matching of models to data is important in applications of statistics. Explaining to researchers that models are simplifications of systems and usually wrong, but that they are useful and necessary for successful quantitative thinking would help break down communication barriers between researchers and statisticians. See Nelder (1986) for more comments on this. A model might be considered to be a good approximation to reality if predictions from the model are consistent with observations. Finding the perfect model can be less important than finding a parsimonious model that "works". Sensitivity analyses, made easier by computers, help here.

An important part of any training is to draw attention to areas that have not been covered fully or at all, so that researchers will realise *when* they need to consult with statisticians. More difficult to do perhaps, but researchers do need to be cautioned that as a technique might be inappropriate for their data or problem, it is advisable for them to check with a statistician that a proposed analysis is meaningful. Similarly they should be advised to check on the suitability of a proposed design. Consultants would rather participate from the beginning than be asked to salvage an experiment. In some disciplines issues which are partly the concern of statisticians are considered to be part of general research methodology rather than statistical (Cox, 1981) so that researchers

would not necessarily think to consult with statisticians on these. Training also needs to prepare researchers on *how* to consult, that is what questions to ask the statistician, and what information the statistician is likely to need about the research (and see Belli, 2001). Ideally training should also aim to give researchers *confidence* that a statistician can help contribute to the research, and stimulate them to want to learn more about statistics.

2.2. TOPICS COVERED IN TRAINING COURSES FOR RESEARCHERS

The specific topics taught in a statistics course for researchers will depend on their disciplines and needs, their background, and the length of the course. There is a large number of papers on various aspects of teaching statistics to those whose main area of interest is not statistics, for example some of the papers included in this book, and several have been given at the various International Conferences on Teaching Statistics (ICOTS). Some general observations on broad topics are made in this section.

Data is an integral part of research in an applied area, and there are a number of issues in connection with data that could be covered in some depth in a training course. It might seem trivial to those who are accustomed to using a spreadsheet or statistics package, but it is important that researchers know how to enter the data in a way that is conducive to statistical analysis. This sometimes depends on which package is being used and on what analysis is to be done. However, common mistakes include entering variables as rows rather than columns, entering text in a cell where a numerical entry is expected, and entering tabulated data in the form of a nicely presented table. Data collected by different researchers or coming from different sources can be difficult to combine because different units of measurement have been used or because data have been entered in different ways in the spreadsheet or a statistical package. Electronic forms of the data might be incompatible. Thus it is useful if researchers learn how to manipulate data both within and between packages. Increasingly automated methods of data collection are being used, often with direct input to a computer, and here too the form of the electronic version is important. For some groups of researchers it might be important to discuss the construction, management and maintenance of large data bases, and the associated documentation. Some researchers might be involved with data collected for administrative purposes, and there could be issues of confidentiality and linkage to other data bases to consider. See also points made in the paper by McDonald (2001).

The structure of data is important – their context, units of measurement, how and why they were collected, and properties of the units on which measurements were made (Nelder, 1986). Researchers need to appreciate that the accuracy of recording determines the degree of accuracy in results, and understand the effect of transformations (which can be useful for statistical analysis) on accuracy and on the interpretation of results. Examples can be found in Preece (1981, 1982). Computers are useful in checking on outliers, on missing observations, and consistency between observations.

Most training courses make use of software, usually the instructor's favourite packages or those readily available. Updated and new versions of software keep appearing, and the researchers might not have access to the software used in the course, so courses need to place more stress on the principles of using software than on specific commands (and see Blumberg, 2001). It is important to make researchers aware that

there are differences in the menu offered by different general statistical packages and that there are specialist packages. Showing researchers how to analyse the same data set in different packages, and how to move output produced by one package into another package for further analysis, would be good preparation for when they are analysing data on their own. Spreadsheets are very popular in the teaching of introductory courses and it is the author's experience that researchers enter their data on spreadsheets and use spreadsheets for preliminary analysis. It is therefore particularly important to teach researchers how to use spreadsheets well. A useful reference, written for natural resources researchers, is *Excel for statistics – tips and warnings*, available from the ASSUME web page (http://www.mailbase.ac.uk/lists/assume/files/).

Hypothesis tests are both widely misunderstood and over-emphasised, and the ease with which P-values can be obtained from statistical packages has not helped. Rather than discuss the fundamental principles and limitations of inference some instructors might be tempted just to tell students that if the P-value is less than 0.05 the result is *significant*. This is a poor teaching practice, which leaves students with little understanding of procedures, or even the difference between a null and an alternative hypothesis. There have been notable attempts to wean researchers away from tests and to present results in terms of confidence intervals, for example Gardner and Altman (2000), but while confidence intervals are obtained from packages as part of a test procedure we might be fighting a losing battle. On the other hand some statisticians recommend that all of test results, confidence intervals, and power analysis or estimation of effects should be presented. Others favour using a Bayesian approach in inference, for example see Lecoutre (1999) who gives arguments in support of this with many references, and Stangl (2001) who discusses why medical researchers should be trained in Bayesian methods.

Exploratory data analysis (EDA) and the initial analysis of data (IDA) (Chatfield, 1985) are very important. Too many courses rush over these important stages and rush on to advanced techniques and methods of analysis. Yet sometimes in a research problem scrutiny of the data and simple summaries and presentations of the data are sufficient. See Chatfield's paper for some examples. Simulation is important in non-standard problems, which do not lend themselves to use of existing techniques (see Mullins and Stuart 1992 for some examples), and researchers could perhaps be taught something about its potential.

2.3. EXAMPLE OF A TRAINING COURSE

In the Autumn of 1998 the author taught the statistics component of a course on research methods given to students taking either an MSc in Sustainable Agriculture or an MSc in Natural Resources at the Natural Resources Institute (NRI), which is part of the University of Greenwich. The statistics content which the author was requested to teach in six 3-hour blocks on 23rd, 27th, and 29th October was "Basic statistics (summary statistics, graphs, etc.), Design of experiments, surveys and sampling schemes, Methods of statistical analysis, including analysis of variance and regression". The statistical backgrounds of the students were unknown at the time the course was prepared, but it was suspected, and was indeed the case, that they had had little previous instruction in statistics.

Clearly, the statistics content of this course in Research Methods covers a very wide

syllabus and it was neither possible nor appropriate to cover it in depth. The author decided to base the statistics teaching on the use of SPSS (computer skills were covered in another part of the course) and to use a data set (Johnson, & Wichern, 1998) in practical sessions.

Hand calculations and the reading of statistical tables were not covered, but references were provided. The author's objectives, as circulated to the students, were:

- To make students aware of the scope of statistics;
- To familiarise students with statistical terms and methods, including the assumptions behind the methods and the situations when the methods would be used;
- To show students how to obtain and interpret output from SPSS;
- To encourage students to consult with statisticians when they have reached the limits of their knowledge of statistics or wish to check they are applying their knowledge correctly, and to give them the confidence to approach statisticians;
- To emphasise the importance of planning in studies involving statistical analysis and that statisticians can give useful advice at this stage;
- To indicate how easy it is to obtain meaningless output from statistical packages.

Four of the five students on the course were from overseas, had little computing experience and were hesitant in English. On the first and third day classes took place in a computing teaching room, but on the second day we were in a classroom without computers in the morning and in a computer laboratory in the afternoon. It would have been preferable to have had all sessions in a computing teaching room, and to have had more time between each of the six 3-hour blocks to give the students a chance to assimilate the material between sessions. Students were given copies of overheads and a few longer handouts to read on their own. The content of the sessions is indicated in Appendix 1. For the most part little more than definitions and ideas were given, but there were linking sections between topics giving, for example, some explanation of concepts. No attempt was made to give derivations, and only a very little notation was used.

It has to be said that the course was not a great success in the sense that students appeared to be overwhelmed by it (hardly surprising given its intensive nature and the amount to be covered) and they did not perform very well on the examination question. This focussed on the understanding of when particular methods of design and analysis were appropriate, and on interpretation of SPSS output, and accounted for 50% of the marks in a 1-hour examination. A colleague had approved the question, but with hindsight there were too many different things for students with almost no experience of statistics to think about and to do in the time. I think that my general approach to the course was satisfactory, but the syllabus is rather long and either should be pruned or more time should be allowed for teaching it, and over a longer period. SPSS was chosen because it was the only statistics package available on the site. The overseas students would probably not have access to any package on return to their own countries, but hopefully it will only be a matter of time before they do, and teaching a calculator-based course would have been restrictive.

On the basis of my experience, if I were to repeat the course, or to teach a similar course elsewhere, I would: a) reduce the amount of algebraic notation used in the notes

even more, b) include examples of SPSS output in the notes (the request to give the course came at a late stage so that it was prepared in something of a rush), c) give students more specific tasks to do as computing exercises instead of open-ended tasks.

2.4. CONVERSION COURSES IN STATISTICS

Researchers who realise the importance of statistics in their area of work or speciality might decide to follow a conversion course in statistics. The University of Greenwich has developed a MSc course of this nature. The University won a contract to provide the college-based component of the UK Government Statistical Service Trainee Statistician Scheme for the three-year period starting in October 1993. Persons in the government statistical service designed the training scheme, to give graduates who had little or no formal statistical training the practical skills and knowledge to become a government statistician. The trainee statisticians would be involved in research during their careers as government statisticians, and so were potential researchers. While preparing the proposal for providing the college-based component the University had the course validated as an MSc degree in Applied Statistics to be taken in part-time (over 2 years), full-time (over 1 year), or block release mode.

When the scheme was first run trainees followed two intensive eight week blocks at university separated by a period of similar length in the government departments to which they were attached, where they worked on projects under the direction of their line managers. The scheme has since been dropped, partly because the internal funding arrangements within the statistical service made it expensive for departments to have trainee statisticians, with the result that very few bids to have trainees were being made.

During the third year of the training scheme, that is in 1995-96, the university offered the MSc degree in part-time mode to be taught on two evenings a week for two years. In the following two years it was also offered in full-time mode to be taught in the evenings as a combination of the year 1 and year 2 of the part-time degree. No entry was taken in 1998-9, but in the 1999-2000 session the degree is being run during the day and there are both full-time and part-time students, the latter attending on one day per week only. The degree does not attract many applicants. There are several possible reasons for this – statistics is not a popular subject, employers tend not to support employees who wish to study either financially or in release from work commitments, those not in employment might not be able to afford the fees. In the case of a conversion degree people also have to be convinced that it is going to be of direct benefit to them in their work or enhance their career prospects. Entrants to the degree have to have basic mathematical skills (see Jolliffe, 1997, p.446) and this might deter some potential applicants.

The syllabus and the way of dividing the degree into teaching units have evolved since it was first proposed. A brief description of the contents of each unit as taught in 1999 to 2000 is given in Appendix 2 and a fuller description is given in Jolliffe (1997). For the award of the MSc students have to spend three months on project work written up as a dissertation after successfully completing the taught part of the course. A postgraduate diploma can be obtained for successful completion of the taught course and this was awarded to those on the trainee statistician scheme. As students are not required to know any statistics when they start, but are studying for a master's degree, a lot of material has to be covered in a short time. This means that whenever a new topic

is introduced there will be a quick progression from an introduction at a low level to a fairly advanced level. As the degree is in *applied* statistics the emphasis is on the practical skills of applying statistical methods to data rather than on mathematical proofs of statistical results.

As a conversion degree it appears to be successful in that several of those who passed the assessment on the taught part of the course (some are still working on their projects) have obtained a first statistical post or have changed to another statistical post since completion. A few of those who started have found the work difficult, although they have had an appropriate background, and a few have dropped out because of heavy employment commitments. Some of those who followed the course are currently in research posts, but working as statisticians. A researcher from other discipline who becomes a statistician in that discipline is in an ideal position to give statistical advice, but it might be argued that training researchers in statistics is counter-productive if the researchers do not remain researchers!

2.5. OTHER METHODS OF TRAINING RESEARCHERS

Another kind of statistical training course is a short course on a specific topic designed with a particular target group of researchers in mind. Such courses typically take place over a one to five day period. Sometimes these are in-house courses, but possibly more frequently are a commercial venture. Topics related to the design and analysis of surveys, and topics in medical statistics are fairly popular, but "brush-up" courses, for example for psychologists, and courses on the use of a specific statistical package, are also offered. The problem with such courses is that they are very tiring both for those attending and for the facilitators, so concentration and attendance are likely to flag before the end of the course.

The main disadvantages of courses on general offer are that people have to spend time travelling to them, and those attending are likely to differ from one another as regards their needs and backgrounds. Researchers do not usually have to travel to in-house courses, and both their needs and backgrounds should be familiar to the course facilitators, which are advantages. The disadvantage is that those attending such courses might be tempted or required to miss some sessions in order to do their ordinary work – which suggests holding residential courses away from the work-place. Even on in-house courses those attending could be a very heterogeneous group as regards their needs and backgrounds. Successful on-site training programmes, which have been developed over several years, are described by Horgan et al (1999) and by Saville (2001) together with the thinking behind the programmes and an assessment of their effectiveness. Although the programmes described are for biological researchers (Horgan, 1999) and agricultural researchers (Saville, 2001), much of what is said is of relevance for courses for researchers in other disciplines.

When the author and some colleagues gave a two-day course to social science researchers at the NRI in 1998 an attempt was made to find out what statistical topics they were familiar with. The lists which came back varied from researcher to researcher, and many of them added such comments as "very rusty", so that this information was almost useless. Similarly the researchers were not unanimous in suggestions as to what they would like us to teach, perhaps because they were not fully aware of how statistics might help them in their work. It would be interesting and useful to perform fairly

detailed surveys of researchers to ask them about their past statistical training, what they perceive their current needs to be, and what kinds of training they think are most successful. There is much anecdotal evidence, but few hard facts. The methodology and results of the MeaNs project (e.g. Holmes, 1996), which looked at employment needs in statistics in general, would be useful background to such studies.

Distance learning courses either by correspondence or over the Internet are another possibility, but these tend to work better if there is also some face to face contact, and they are not necessarily going to match the researcher's needs. Researchers can also be trained "on the job" as and when the need arises, perhaps by another researcher in the case of a routine research exercise, perhaps by consulting with a statistician.

3. CONSULTANCY

3.1. TRAINING STATISTICIANS

It might be thought that a discussion of the training of statisticians as consultants is inappropriate in a meeting to discuss the training of researchers in statistics, but if statisticians are good consultants those who ask their advice will learn some statistics. Much has been written on consultancy skills and on training statisticians to act as consultants. Barnett (1994) gives an overview, and papers on consultancy have been given at all the ICOTS. In some ICOTS there have been whole sessions on consulting, the latest being one at ICOTS5 (Pereira-Mendoza et al., 1998). The paper by Hunter (1981), which is based on actual examples of statistical consulting, contains many useful comments on what to do (or not do) when acting as a statistical consultant. Preece (1987) discusses the role of a statistician in a research project through from planning to writing up, with an emphasis on good practice.

Like any practical skill, statistical consultancy is learnt by practice and experience. Teaching of consultancy, if to be effective, must attempt to give students practice in doing consultancy, for example, by having them sit in on or participate in consulting sessions (see Rangecroft, & Wallace, 1998), or by setting up consultancy situations based on real problems. See also Belli (2001), Godino et al (2001) and Ospina and Ortiz (2001) for discussion of training in consultancy. Ideally facilitators of training courses in consultancy should themselves have experience of acting as consultants, or at least have worked as practical statisticians, and, as Nelder (1986) comments, every statistician's training should involve collection and analysis of their own data – and that involves learning about the area of application. This makes statisticians better able to identify with those who are at the receiving end in a statistical consultancy session. Students who do a period of industrial training might gain experience of research problems, and hence of consultancy.

Training in communication is of prime importance. Statisticians who are strong in theory but weak in application are at a disadvantage working alongside experts in particular subject fields and tend to be pushed into a support role, especially if they are poor communicators (Moser, 1980). Lack of training in written and oral communication skills might partly explain why there has not been the increase in the demand for statistics and the services of statisticians we might have expected to follow developments in information technology (Nicholls, 1999). Requiring undergraduate and

postgraduate statistics students to write a dissertation, and to give an oral presentation on it, could help statisticians to acquire communication skills. Statisticians also need training in listening and questioning skills to help them become effective in finding out the researcher's needs.

Statisticians need to sell themselves and the contribution that statistics can make to other disciplines. One way to do this is to demonstrate the utility of statistics and to build on success stories, that is, to communicate the usefulness of statistics. Greenfield (1993) argues that statisticians must change their culture in order to bring about a greater acceptance of statistics among non-statisticians. Statisticians need to be trained to communicate statistical information in everyday language free of statistical jargon and concepts (Nicholls, 1999). They might need to communicate results to clients, at meetings of research groups, to committees, at courses, at conferences. They could well be involved in making an input to a research proposal.

There is a strong argument that even courses for specialist statisticians should include some teaching in an application area. The trend towards joint degrees in statistics and another subject, for example economics or psychology, goes some way towards meeting this (though cynics would say that such joint degrees are introduced with the aim of attracting more students into statistics departments, not in itself a bad thing). Reading relevant journals and attending conferences in an applied area will help statisticians become known and accepted, and enable them to discuss recent developments with specialists in the area on equal terms. In turn their greater involvement with the application will make them better teachers and consultants.

3.2. INTERACTION BETWEEN THE STATISTICIAN AND THE RESEARCHER IN CONSULTANCY

Perhaps the most important condition for a consulting session to be successful is that the statistician and the researcher can communicate with one another (Belli, 2001; Saville, 2001). Notation, abbreviations, vocabulary and jargon can be a barrier in communication, something which both the statistician and the researcher will need to bear in mind, and to remind one another that they do not always speak the same language. The Royal Statistical Society and the Institute of Electrical Engineers have in 1999 to 2000 held a series of joint meetings because some statisticians and engineers recognised that improving communication would benefit both groups. Differences in vocabulary have been a barrier to communication between these two disciplines.

First the researcher has to communicate the problem to the statistician – and here the researcher must be careful not to assume that (s)he knows a suitable method and ask only that the statistician works out the details. The statistician has to have at the very least a general understanding of the research problem. It helps if the statistician has some expertise in the researcher's field, but is not essential provided the statistician is interested in the subject matter of the research and is willing to learn more. In fact by explaining the problem in layman's terms the researcher might see the problem in a new light. It could also be the case that a "general" statistician is more likely than a statistician who has specialised in the research topic to realise that a statistical method not usually used in that area of research is appropriate. Related studies and relevant theory from the subject area need to be taken into account and the statistician needs to find out about these. If simplifying assumptions are needed for the statistical analysis the

statistician has to ensure that these are sensible from the point of view of the subject matter.

In any investigation the planning, what data are collected, and the analysis depend on the objectives of the study, and the statistician needs to be aware of what these are, and to be involved as a member of the research team from the start. If necessary, objectives may have to be modified in the light of financial, time and staffing constraints, and even by the statistical techniques and methodology of the day. As Chatfield (1985) comments, the true objectives of the study might turn out to be different from those suggested in the initial consultation. Here the statistician can help clarify the questions that the researcher wishes to consider (Hand, 1994) and both researcher and statistician will learn from this as they move towards a common understanding. What are sometimes called "errors of the third kind" – giving the right answer to the wrong question – must be avoided.

The statistician needs to be able to communicate in statistical terms at a level appropriate to the statistical knowledge of the researcher. It can, however, be difficult to explain some of the more complex statistical techniques in simple terms so that not only the researcher, but also those to whom the researcher plans to disseminate results, can understand what is involved. Nevertheless this is an important part of the statistician's work as a consultant. Bradstreet (1999) makes a strong case for communicating effectively through graphics, and for placing more emphasis on this in statistical education.

Collaboration between researcher and statistician at all stages is important and the consulting aspect should perhaps be played down. Consulting has "overtones of pretentious servitude" (Cox, 1981. P.294). Significant progress in any field of application needs the participation of both specialists in the field and of statisticians. This has happened in the pharmaceutical industry, but in few other areas. Most statisticians would prefer to be collaborators and joint authors rather than thanked in a foot-note if their contribution is at all substantial, but the survey results reported by Godino et al (2001) suggest that it will be some time before statisticians are readily accepted as joint authors. In all cases statisticians must be given the opportunity to veto any incorrect statistical work or misunderstandings on the part of the researchers, and mistakes can occur as Stangl (2001) comments. Saying that a statistician was involved does not give work statistical credibility.

It is not uncommon to hear of difficulties in getting papers involving statistical methods and results, and papers correcting erroneous statistics, published in application areas. One argument is that those who will use the results are not interested in how they were obtained. Yet description of the statistical analysis, including the assumptions made and limitations of models, and perhaps comparison of using different statistical techniques, are as important as a description of the method of data collection (and sometimes even that is omitted). Collaboration at the report writing stage is crucial for the public image of statistics and statisticians, and statisticians should review papers with a substantial statistical content. Depending on the intended readership, the statistical section could be written in a form suitable for the lay person, or as a technical report for fellow statisticians.

On occasion the statistician might have to overcome negative attitudes towards statistics and the role of the statistician such as the attitude that only those with the appropriate professional qualifications can be a medical doctor or social worker or whatever, but anyone who can add up or click on a mouse can analyse data (see Cox,

1981, and Bangdiwala, 2001). Both researchers and statistical consultants might find the paper by Brook (1994) useful preparation for consultancy sessions.

3.3. HOW THE STATISTICAN AND RESEARCHER LEARN FROM ONE ANOTHER

By consulting with statisticians, researchers learn when to consult and what questions to ask, as well as learning about statistical concepts, design and analysis (Belli, 2001; Saville, 2001; Svensson, 2001). It is sometimes easier for the consultant to suggest a design for a study and to analyse the data than to tell the researcher how to do this, but by helping the researcher to do these things him or herself the consultant is giving the researcher statistical training. There might be an opportunity to encourage the researcher to repeat the study and under standardised conditions – to look for significant sameness as opposed to a significant difference in a single experiment (Nelder, 1986). In general the statistician might be able to help the researcher think quantitatively.

Researchers often come for help with statistical analysis after they have collected their data, and sometimes they have already entered the data in a spreadsheet or have attempted some preliminary analysis. If not done well this can create many problems for the statistician, and sometimes the design is not ideal for the researcher's purpose. This does, however, provide an opportunity to extend researchers' statistical knowledge, and to save them from making the same mistakes on another occasion. Statisticians learn teaching and consultancy skills when they help researchers, and consultancy sessions are a guide as to the development of suitable training (see Caulcott, 1987).

As well as learning about the research itself, statisticians might get into another area of statistics, and they might even learn some statistical techniques previously unknown to them from the researcher. Smith (1993) takes the view that the great majority of work which could be classified as statistics is carried out by specialists who have no wish to be called statisticians, for example agronomists, biologists, and chemists, and says that subject matter specialists, such as psychometricians and econometricians are statisticians working as scientists. It may also be the case that some researchers become more expert in a specialised area of statistics than a "general" statistician.

If presented with a challenging statistical problem when acting as a consultant a statistician might develop new techniques (for example, Svensson, 2001). Practical problems stimulate statistical research (Barnett, 1994) and the growth of statistics as a discipline depends a great deal on application areas. In particular, agricultural research has had a huge influence on the development of statistical methods (see Gower, 1988). Yet many statistical papers tend to deal with theories looking for data rather than with real problems needing theoretical treatment (Moser, 1980). Some examples of interesting problems arising from consultancy are given in Barnett (1994) and Chatfield (1985) amongst others.

4. CONCLUSION

This paper has considered various aspects of training and consulting, from the viewpoint of both the researcher and the statistician. It is suggested that both would learn from the experience of interacting with one another and that this experience should

improve the training of researchers in statistics.

APPENDIX 1. CONTENT OF SESSIONS IN A SHORT TRAINING COURSE

1. Types of data. Frequency tables, bar charts. Frequency distributions, histogram, stem and leaf, median, quartiles, box plot, mean, standard deviation. SPSS session on this material.
2. Basic idea of probability as a relative frequency. Discrete random variables. Discrete uniform, binomial and Poisson distributions. Continuous random variables. Uniform, normal and exponential distributions. Idea of a sampling distribution, sampling distribution of the sample mean, use of the normal distribution for finding a confidence interval for a population mean. SPSS session on confidence intervals for a mean, and on areas under normal curves.
3. Tests of hypotheses – concepts, terminology. Tests of normality. Inference re means (1 sample t, independent samples t, paired samples t) and non-parametric alternatives to t. Bonferroni inequality. Definition of bias. Scatter diagrams. Pearson's coefficient of correlation. Simple linear regression, comments re multiple linear regression, treatment of categorical variables. Spearman's coefficient of rank correlation. χ^2 tests of association, homogeneity.
4. SPSS session exploring a data set and trying different techniques.
5. Design of experiments – terms, principles. Matched pairs designs, one-way ANOVA, two-way ANOVA, Latin square design. SPSS session on one-way ANOVA.
6. Sampling from finite populations – overview of methods. Notation and basic results for simple random sampling and stratified sampling. Estimation of sample size for estimating a mean and proportion under s.r.s. Proportional and optimal allocation in stratified sampling. Non-sampling errors, imputation, deff.

APPENDIX 2. DETAILS OF AN MSC CONVERSION DEGREE IN STATISTICS

Outline syllabuses

Statistical Methodology and Techniques: Inference (point and interval estimation, tests of hypotheses including non-parametric methods); Methods of estimation and properties of estimators; Multivariate analysis; Computer intensive methods; Quality control

Statistical Modelling: Probability and probability distributions; Simple and multiple regression; Design of experiments, analysis of variance; General linear model; Logistic, loglinear, and other models

Applied Statistics: Survey methodology (qualitative aspects, sampling methods); Time series analysis

Statistical Investigations: Data analysis; Communication skills; Case studies

Medical Statistics: Epidemiology; Clinical trials

Assessment

Statistical Investigations is assessed 100% by coursework, which includes open-bookwork tests in a computer laboratory. Assessment in the other courses is 80% by examination and 20% by coursework.

Statistical Methodology and Techniques, *Statistical Modelling*, and *Applied Statistics* are worth 20 credits each and *Statistical Investigations* and *Medical Statistics* are each worth 10 credits. The project is worth 40 credits. 120 credit points are needed for award of an MSc.

A 10 credit course is time-tabled as 3 hours a week for 13 weeks.

REFERENCES

Bangdiwala, S. I. (2001). Training of statisticians worldwide to collaborate as co-investigators within country clinical epidemiology units: The experience of the International Clinical Epidemiology Network (INCLEN). In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 265-275). Granada: International Association for Statistical Education and International Statistical Institute.

Barnett, V. (1994). The role of consultancy in university education and professional training in statistics. In L. Brunelli, & G. Cicchitelli (Eds.), *Proceedings of the First Scientific Meeting of IASE* (pp. 285-297). University of Perugia.

Belli, G. (2001). The teaching/learning process in university statistical consulting labs in the United States. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 325-338). Granada: International Association for Statistical Education and International Statistical Institute.

Blumberg, C. J. (2001). Training regular education and special education teachers in the use of research methodology and statistics. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 231-244). Granada: International Association for Statistical Education and International Statistical Institute.

Bradstreet, T. E (1999). Graphical excellence. The importance of sound principles and practices for effective communication. Invited paper at the fifty-second session of the International Statistical Institute, Helsinki. http://www.stat.fi/isi99/proceedings.html

Brook, R. J. (1994). Questions to ask a statistical consultant. *Proceedings of the Fourth International Conference on Teaching Statistics*, (vol 1, 257-264). Marrakesh: National Organising Committee.

Caulcott, R. (1987). Statistics in industry: a failure of communication. *The Statistician*, 36(5), 555-560.

Chatfield, C. (1985). The initial examination of data. *Journal of the Royal Statistical Society*, Series A, 148(3), 214-253.

Cox, D. R. (1981) Theory and general principle in statistics. *Journal of the Royal Statistical Society*, Series A, 144(3), 289-297.

Gardner, M. J., & Altman, D. G. (2000). Confidence intervals rather than P values. In D. G. Altman, D. Machin, T. N. Bryan, & M. J Gardner (Eds.), *Statistics with confidence*, 2nd ed (pp. 15-27). London: British Medical Journal Books.

Godino, J. D., Batanero, C., & Gutiérrez Jaimez, R. (2001). The statistical consultancy workshop as a pedagogical tool. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 339-353). Granada: International Association for Statistical Education and International Statistical Institute.

Gower, J. C. (1988). Statistics and agriculture. *Journal of the Royal Statistical Society*, Series A, 151(1), 179-200.

Greenfield, T. (1993). Communicating statistics. *Journal of the Royal Statistical Society*, Series A, 156(2), 287-297.

Hahn, G. J. (1999). Opportunities and challenges for industrial statisticians in the 21st century. Invited paper at the *Fifty-Second Session of the International Statistical Institute*, Helsinki. http://www.stat.fi/isi99/proceedings.html.

Hand, D. J. (1994). Deconstructing statistical questions. *Journal of the Royal Statistical Society*, Series A, 157, 3, 317-356.

Hirotsu, C. (2001) Statistical training of researchers in total quality management: the Japanese experience. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 53-63). Granada: International Association for Statistical Education and International Statistical Institute.

Holmes, P. (1996). Matching education, assessment and employment needs in statistics: The

MeaNs project. *Teaching Statistics* ASLU supplement, 1-3.

Horgan, G. W., Elston, D. A., Franklin, M. F., Glasbey, C. A., Hunter, E. A., Talbot, M., Kempton, R. A., McNichol, J. W., & Wright, F. (1999). Teaching statistics to biological research scientists. *The Statistician*, 48 (3), 393-400.

Hunter, W. G. (1981). Six statistical tales. *The Statistician*, 30(2), 107-117.

Jolliffe, F. (1997). How much theory should be covered in an MSc conversion course? *Proceedings of the 51st Session of the ISI, Bulletin of the International Statistical Institute*, (Tome LVII, Book 1, pp. 445-448). Istambul: International Statistical Institute.

Johnson, R. A., & Wichern, D. W. (1998). *Applied multivariate statistical analysis*. 4th edition. Prentice Hall, New Jersey.

Johnson, R. A., & Wichern, D. W. (1998). *Applied multivariate statistical analysis*. 4th edition. New Jersey: Prentice Hall.

Lecoutre, B. (1999). Beyond the significance test controversy: Prime time for Bayes? *Proceedings of the 52nd Session of the ISI, Bulletin of the International Statistical Institute*, (Tome LVIII, Book 2, pp. 205-208). Helsinki: International Statistical Institute.

McDonald, S. (2001). Practical and educational problems in sharing official microdata with researchers. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 119-128). Granada: International Association for Statistical Education and International Statistical Institute.

McPherson, G. (1989). The scientists' view of statistics: A neglected area. *Journal of the Royal Statistical Society*, Series A, 152(2), 221-240.

Moser, C. (1980). Statistics and public policy. *Journal of the Royal Statistical Society* Series A, 143(1), 1-31.

Mullins, E., & Stuart, M. (1992). Simulation as an aid in practical statistical problem-solving. *The Statistician*, 41(1), 17-26.

Nelder, J. (1986). Statistics, science and technology. *Journal of the Royal Statistical Society*, Series A, 149(2), 109-21.

Nicholls, D. (1999). Future directions for the teaching and learning of statistics at the tertiary level. Invited paper at the *Fifty-Second Session of the International Statistical Institute*, Helsinki. http://www.stat.fi/isi99/proceedings.html

Ospina, D., & Ortiz, J. E. (2001). Statistical research and consulting in universities from developing countries: The Colombian case. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 295-310). Granada: International Association for Statistical Education and International Statistical Institute.

Pereira-Mendoza, L., Seu Kea, T. Wee Kee, & W. K. Wong (Eds.) (1998). *Proceedings of the Fifth International Conference On Teaching Of Statistics* (pp. 341-376). Singapore: International Association for Statistical Education and International Statistical Institute

Preece, D. A. (1981) Distributions of final digits in data. *The Statistician*, 30(1), 31-60.

Preece, D. A. (1982). t is for trouble (and textbooks): a critique of some examples of the paired-samples t-test. *The Statistician*, 31(2), 169-195.

Preece, D. A. (1987). Good statistical practice. *The Statistician*, 36 (4), 397-408.

Rangecroft, M., & Wallace, W. (1998). Group consultancy, as easy as falling off a bicycle? In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics* (pp. 359-364). Singapore: International Statistical Association and International Statistics Institute.

Saville, D. (2001). A hands-on, interactive method of teaching statistics to agricultural researchers. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 197-213). Granada: International Association for Statistical Education and International Statistical Institute.

Smith, T. M. F. (1993). Populations and selection: limitations of statistics. *Journal of the Royal Statistical Society*, Series A 156(2), 145-166.

Stangl, D. (2001). Design of an internet course for training medical researchers in Bayesian statistical methods. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 147-157). Granada: International Association for Statistical Education and International Statistical Institute.

Svensson, E. (2001). Important considerations for optimal communication between statisticians and medical researchers in consulting, teaching and collaborative research – with a focus on the analysis of ordered categorical data. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 23-35). Granada: International Association for Statistical Education and International Statistical Institute.

Web addresses

ASSUME (Association of Statistics Specialists using Microsoft Excel): *http://www.mailbase.ac.uk/lists/assume/files/*

STEPS (Statistical Education through Problem Solving) project. University of Glasgow:. http://www.stats.gla.ac.uk/cti/

Flavia Jolliffe
Computing and Mathematical Sciences,
University of Greenwich, Maritime Greenwich Campus
Old Royal Naval College, *Greenwich, London, SE10 9LS U.K*
E-mail: F.R.Jolliffe@gre.ac.uk

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 379-382.

Printed in Granada, Spain.

YUKI MIURA

DISCUSSION

The four papers presented in this session deal with the problems of training of statisticians in consulting skills, training of researchers in the use of statistics, and enhancing interaction between statisticians and researchers. There is no doubt that the scope of areas of applications of statistics is expanding, and owing to a rapid progress in information and computer technologies and to the widespread use of friendly computer software, the use of statistical methods for data analysis has become common in research work in fields different from statistics and in practical work in industry and government.

There is a general agreement that the interaction and communication between statisticians and researchers is very important. Statisticians who will be consulted by researchers in other fields or by practitioners require consulting skills. Client researchers or practitioners who will deal with data need a basic knowledge and understanding of statistics and statistical thinking.

Take the example of medical doctors. A medical practitioner who is specialised in internal medicine, surgery, ophthalmology, or dermatology must have studied all areas of medical science before becoming a specialist or a practitioner in specific fields of medicine. However good he or she is at medical discipline and techniques, he or she may not be regarded as a good physician by patients if he or she lacks a good ability to communicate with patients. Likewise, to be a good statistician, he or she must have good communication skills to listen to clients, understand their problems and be able to give appropriate explanations and advice.

Programmes for training statisticians in consulting skills, and training researchers in other fields in the use of statistics and in how to consult with statisticians are then essential. The authors of the four papers presented in this session agree on the importance of these matters, although the current state of approaches to training varies from country to country.

When considering statistical consultations, there will be different categories of clients. These include doctorate students, researchers in fields different from statistics, and practitioners working in industry or the government who deal with data. There may be variations in the statistical backgrounds of the clients.

Fields of interest and the problems clients face also vary. In areas such as medical and pharmaceutical sciences, agricultural science, engineering, sociology and psychology, the clients may be concerned with experimental design, sampling design, data collection, data scrutiny, data processing, and data analysis, and in each of these phases problems may arise. On the other hand, in the fields of economics or demography where data might have already been aggregated, clients will usually handle published official statistics or micro-data from censuses and sample surveys. In this case, one may be interested in the method of data collection, the concepts and definitions used, the reliability of data, and the method of data analysis. When clients use a certain statistical method, they should understand the underlying assumptions and limitations of the method and interpret the results correctly.

Belli's paper presents the interesting results of an electronic survey of 106 US departments. I agree that in order for statisticians to be able to communicate effectively with researchers and practitioners, they should have good consulting skills. I noted that some US universities have statistical consulting laboratories or centres, and such units are also a forum for interaction between statistician and researcher from diverse fields. I would like to know more about the formal course on consulting, and who give the courses and what their qualifications are. What is the impact of the development of computer and information sciences on statistics education in the US universities? Is statistics education in the universities in decline or is it expanding?

The paper by Godino, Batanero and Gutiérrez Jáimez presented a proposal to organise Workshop of Statistical Consultancy and to establish Statistical Consultancy Units in the university faculties. This proposed scheme will in future cover different areas of discipline where needs arise. Target areas may be quite diverse, and I wonder what the qualifications for tutors of these workshops will be.

Jolliffe's paper focuses on two aspects of training: on courses in statistics for researchers, and on the consulting process.

The situation in Japan

I should like to explain the present situation in Japan with regard to statistical consulting. There is, I believe, no formal course for training statisticians in consulting and communication skills. Such skills are usually acquired through practice and experience.

In official statistics, there is a reference unit in the Statistics Bureau open to the public. This unit acts as a clearing-house of official statistics, and attracts many clients including laypersons as well as researchers who are interested in the use of official statistics, and gives them advice on sources and nature of data, how to use the data and other relevant technical matters.

The Institute of Statistical Mathematics has an information centre for consultation. The Institute receives many requests from outside for statistical help and consultation. The centre will refer the clients to appropriate statisticians by matching the problem areas and the area of interest of respective statisticians. This consulting process is working well.

In the universities there is no formal statistical consulting unit. The researchers or practitioners may sometimes approach professors in statistics in the faculty for help and consultation. Thus the statistical consultation usually takes place on a personal basis, depending on the area of interest of professors.

Training of researchers in other fields is also important. The Statistical Training Institute attached to the Statistics Bureau of Japan offers various courses to those employees of the central and local governments who are engaged in statistical work or data analysis. The core course of six months offers intensive curricula from theory to applications. The Institute of Statistical Mathematics organises three or four tutorial programmes a year on different subjects. The programmes provide a forum where researchers and practitioners can become acquainted with new and useful statistical methods.

In the industry sector, the Japanese Standards Association and the Japanese Federation of Science and Technology both organise many statistical seminars and courses on statistical techniques useful to industry. These contribute to the advancement of statistical capabilities of the staff working in quality management and production

processes in industry. On-the-job training and in-company training are also common.

During the annual meeting of the Japanese Statistical Society, two tutorial seminars are organised on newly developed areas of statistics, and these sessions are attended by many researchers and practical statisticians as well as statisticians in different fields. Topics of this year's tutorial seminar were "non-linear multivariate analysis" and "finance engineering and statistical analysis".

Theoretical statisticians must have some areas of application of statistics that they are interested in. By studying problems arising from the real world, they can discover and develop new theories and methods, and thus contribute to the development of statistics. So consultation and collaboration with researchers and practitioners who have problems sometimes lead to new area of statistical research. Although statisticians can give advice on statistical problems of a general nature, they cannot always cope with the problem in specific areas that are unfamiliar to them.

Nowadays one can find many textbooks on statistics authored by non-statisticians, mostly by experts in computing. Those books are written in a friendly way, focusing on how to use statistical tools rather than statistical theories or statistical thinking. They are easy to read for non-statisticians. They seldom touch upon underlying assumptions behind the statistical methods or limitations of the methods. The worst thing is that they do not discuss problems with the quality of the data that are going to be analysed. Here is a danger of misusing statistical tools. Training of researchers in statistics emphasising conceptual aspects and statistical thinking is thus needed.

Training in communication and consultation skills of statisticians will be very useful and effective for doctoral students or potential researchers. Training in statistics of researchers in other fields is also needed. Interaction between statisticians and the other groups will be beneficial to all.

However, the problems that researchers or practitioners face in the real world, say, in official statistics or in industry, are so complex that collaboration of experts in related fields is absolutely essential.

Take as an example the treatment of missing data in a census or a survey conducted by the government statistical office. In order to find out appropriate models and develop rational, workable methods, the Statistics Bureau here established a small research group consisting of theoretical statisticians, demographers, a sociologist, and experts familiar with data collection and data processing of real large-scale statistical surveys or censuses. Similar small groups are formed for other problems.

This type of consultation process is useful for both academic members and practitioners. For academic statisticians and researchers, this forum provides a good opportunity to become acquainted with complex problems in the real world. For practitioners working in the Statistics Bureau, it provides a valuable chance to learn new developments of statistical theory and practice.

Consultation processes in a similar form exists in the interaction between academia and industry. For instance, the pharmaceutical field is one of the most successful and active areas of statistical consultation or collaborative statistical research. Such collaborations are common in this field. Pharmaceutical firms themselves organise in-house training courses for their researchers in new statistical methodologies. Quite a few researchers working in the pharmaceutical firms always attend scientific sessions on medical and pharmaceutical topics in the annual meeting of our statistical society.

Collaborative research efforts in improving quality of products are carried out in the motor industry and some other industries, as well. Academic statisticians in quality management will play a major role in the project.

At the Institute of Statistical Mathematics, most statisticians are engaged in theoretical research as well as in research in application areas. There are many collaborative research projects involving statisticians and researchers in other fields from other universities and research institutes. These collaborative research projects thus provide forums for interaction between statisticians and researchers from different institutions.

As mentioned by Belli, Jolliffe and Godino, Batanero and Gutiérrez Jáimez, statistical consultation of statisticians will definitely be beneficial to client researchers, but it is often not appreciated by statisticians as it has no novelty and is merely an application of the existing theory or method. Even if the statistician's advice or contribution is essential to the substance of the research, the statistician's role in the research is often not appreciated properly.

The future of statistics education

Shia discussed the future of statistical education and suggested the idea of an electronic school. He suggested that the function of such an e-school should cover not only the e-book and video education, but also consultation and the services of a data warehouse. I am sure that databases to be constructed for these purposes will be very useful if they can serve to solve problems in the real world as well as for research work. Shia also suggested Internet surveys should be included in the function of an e-school. I could not get his idea clearly. Are such Internet surveys for statistics education purposes or for solving real problems?

Anyway, the world, particularly the environment surrounding statistics, is changing rapidly. To be benefited by new technologies, we need to strengthen international co-operation and the IASE can play an important role in this.

REFERENCES

Belli, G. (2001). The teaching/ learning process in University statistical consulting labs in the United States. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 325-338). Granada: International Association for Statistical Education and International Statistical Institute.

Godino, J. D., Batanero, C., & Gutiérrez Jáimez, R. (2001). The statistical consultancy workshop as a pedagogical tool. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 339-353). Granada: International Association for Statistical Education and International Statistical Institute.

Jolliffe, F. (2001). Learning from experience. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 355-370). Granada: International Association for Statistical Education and International Statistical Institute.

Shia, B.C. (2001). How to think about statistical consultation. Learning from data. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 371-377). Granada: International Association for Statistical Education and International Statistical Institute.

Yuki Miura
Surugadai University,
Hanno, Saitama-ken, Japan
E-mail: miura@parkcity.ne.jp

Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 385-396.

Printed in Granada, Spain.

CARMEN BATANERO

MAIN RESEARCH PROBLEMS IN THE TRAINING OF RESEARCHERS

In this last chapter I will try to summarise the main arguments made in the preceding chapters and the points raised during the Tokyo IASE Round Table Conference debates. During a very productive week at the Institute of Statistical Mathematics, we discussed the issues and problems concerning the Training of Researchers in the Use of Statistics. As suggested by Jolliffe, this is an extensive topic that involves consideration of statistics itself, the interaction between statisticians and researchers, and even the training of statisticians.

My aim in writing this summary is to provide the reader with the view that emerged from our joint reflections about this complex issue and its many facets. The conclusions below are organised around the research questions initially set out in the Discussion Document, which is reproduced in the first chapter of the book, so that the extent to which those questions were taken into account in the different contributions can be more easily perceived.

All the authors highlighted the relevance of improving researchers' statistical training and of making statisticians better acquainted with other research fields for ensuring an optimal communication between statisticians and researchers. However, it was also made clear from the arguments throughout the book that there are no easy or definitive answers to the questions raised.

In addition to the complexity of the topic, little research has been done until now on the teaching of statistics to postgraduates and on the difficulties of learning advanced statistical topics. Even when some of the papers present examples of useful research and innovative teaching experiences, the Conference participants felt that there is an urgent need to increase research in this area.

This chapter then presents not a list of final agreed solutions to the problems of training of researchers, but our view about the main areas in which further research on this field is needed. Some recommendations about ways in which the teaching of statistics might contribute to increasing statistical knowledge among researchers and to better use of statistics in experimental research are also made.

1. SPECIFIC STATISTICAL COMPETENCIES IN THE TRAINING OF RESEARCHERS

Few researchers can do their work effectively today without reference to empirical information and statistics provides a set of tools to manage, organise, describe and interpret this information. As suggested by Miura, due to a rapid progress in information and computer technologies and to the widespread use of friendly computer software, the use of statistical methods for data analysis has become common in research work in fields different from statistics and in practical work in industry and government.

In a wide study of research literature in the biological and health sciences, Harraway et al. classified the statistical methods used in 16 high impact journals during a period of two years. This study, some previous similar studies in education (Elmore, & Woehlke, 1988, 1998) and veterinary journals (Hammer & Buffington, 1994), the analysis of doctoral theses in education by Godino et al., as well as the deep analysis by Blumberg reflect the tremendous variety of both elementary and complex statistical methods that researchers in different areas are using today to carry out their investigations.

This number of useful statistical methods and the quick pace of change and development in statistics mean that researchers' statistical knowledge is insufficient for them to be independent. Consequently, it was emphasised in different papers that it is unrealistic to expect researchers to be their own statisticians and solve all their data analysis problems by themselves. This was also perceived by the researchers themselves, according to the results of the surveys carried out by some participants in the Conference. As argued by Phillips, it is not necessary for particular researchers or departments to do everything by themselves. It is more important that they make use of the diverse and rich resources which exist in their institutions, including statistics consultancy. Chadjipadelis suggests that it is not clear if we should teach statistics to researchers or if it is better to make them understand that they need a statistician as a help, a colleague or even as a leader.

We discussed the main abilities to be emphasised in the training of researchers and there was an agreement that it is important for them to understand the basic ideas in survey and experimental design and data collection. Researchers should be able to read, interpret, communicate and defend arguments based on statistical information. They should develop understanding of assumptions behind statistical methods and use meaningful statistics language. Schuyten also pointed out the need to take into account the interrelationships between research methodology and statistics in the training of researchers.

In Jolliffe's words, training in statistics is a continuing process that comes partly through experience and thus requires a long time. Successful courses will encourage a critical attitude towards data and the results of statistical analyses will allow researchers to experience the variability and accept uncertainty, and increase their interest in statistics. The importance of obtaining good quality data and of exploiting data sets fully should be stressed.

An important part of any training is to draw attention to areas that have not been covered, so that researchers will realise when they need to consult with statisticians. Recommendations were made that short continuing education courses focussed in specific topics are essential for the trained researchers as a means to prevent them making some well known and serious errors; and also that it is convenient to offer these courses in connection with meetings of scientific societies.

2. PARTICULAR NEEDS AND PROBLEMS IN THE STATISTICAL TRAINING OF RESEARCHERS IN SPECIFIC FIELDS

A field where the use of statistics has a long tradition and is now well established is agriculture. Saville describes a number of workshops to fill the training needs of these researchers, based on his long experience of consulting and teaching in an agricultural research institute.

Harraway and his colleagues complement their study of research journals with the description of some case studies and a survey given to researchers and postgraduate students in five different departments. It is apparent from the paper that a great deal of statistics is used by researchers in the biological and medical sciences, although the specific techniques used vary to some extent between and within disciplines. Given the time constraints, they suggested that generic courses should be focused on basic statistics and advanced techniques should be offered only when required by different areas.

While the previous papers focused on statistical techniques and concepts, Bishop and Talbot argue that researchers do not fully grasp the essence of statistical thinking and its ways of reasoning and, there is therefore a need to provide training in statistical strategic skills.

They describe a web-based project to develop statistical thinking directed to researchers in the biological sciences, which is based on the cyclic model by Wild and Pfannkuch (1999), who propose five components: Problem-Plan-Data- Analysis and Conclusions. Even though this particular facility was designed for the biological sciences, the didactical problem of developing statistical thinking in researchers is applicable in general to researchers in different sciences.

In the context of training special education and regular education teachers, Blumberg analyses the usefulness and requirements if training in a number of statistical topics. In his discussion, Brian Phillips argues that not only the teachers, but other professionals such as doctors, social workers, etc. are continually collecting and analysing data on people's behaviours and Blumberg's analysis is valid for them too. Glencross and Mji mention as relevant for the training of researchers in social sciences many of the topics analysed by Blumberg and also describe their workshops in such advanced topics as principal component analysis, correspondence analysis and other multivariate methods.

They recognise that there are problems in presenting the workshops, including the researchers' lack of computer literacy and little formal knowledge of basic statistics and mathematics. However, there was an agreement among participants that a minimum, intuitive knowledge of statistical tools should be provided for researchers if we want them to be able to critically read and interpret research literature in their fields.

3. MAIN LEARNING PROBLEMS, MISCONCEPTIONS AND ERRORS CONCERNING PARTICULAR ADVANCED STATISTICAL CONCEPTS

In the second part of the book we are including a set of papers that specifically dealt with the analysis of the main reasons why training in particular statistical topics is needed. Topics cover association (Estepa & Sánchez-Cobo), categorical data analysis (Svensson), stochastic processes (Wang), quality control (Hirotsu), statistical models (McLean) and Bayesian statistics (Iversen). Other papers in the book analyse the teaching and learning of experimental design and linear models (Saville), statistical thinking (Bishop & Talbot) and multivariate analysis (Crivisqui et al.) or refer to the increasing use of complex methods. A main conclusion from these papers is that statistics is misunderstood and misused by researchers, not only as regards advanced methods such as stochastic processes or multivariate analysis, but also in relation to very basic concepts.

In research carried out with undergraduates, Estepa and Sánchez-Cobo systematically describe errors and misconceptions in a number of concepts related to the idea of association, including functional and random dependence, covariance, correlation and regression. They also summarise other previous research that points to misinterpretation of contingency tables, confusion between correlation and causation, and the effect of previous theories on the interpretation of association (see, for example Beyth- Marom, 1982, for a survey of psychological research on the interpretation of association).

As remarked by Mukherjee in his reaction, it is pretty difficult to provide adequate guidance to people involved in training researchers in this topic. He proposes a number of practical suggestions, such as systematic introduction of themes, stress of real-life situations, starting with data sets and knowledge of the background research. Aliaga also suggests more use of technology and that teaching should move from passive to active and should emphasise statistical thinking. Since previous research (Batanero, Estepa, & Godino, 1997; Morris, 1997; Truran, 1997; Batanero, Godino, & Estepa, 1998) shows that some of these misconceptions are resistant after instruction based on use of computers, and students' active work with real data, we deduce that much more research to evaluate and improve the effectiveness of such approaches is still needed.

Hand (1996) argued that little consideration is given to the measurement level of the data, even when this condition affects the type of applicable statistical analysis. In her paper, Svensson indicates that this is very common when assessing qualitative variables, such as feelings, attitudes, preferences, etc. Since categorical responses are often transformed into numerical scores, there is a temptation to treat such data as numerical values. A problem is that the teaching of statistics is focussed on methods appropriate for quantitative data and, therefore, even well educated researchers might be unaware of the fact that there are statistical methods suitable for ordinal or qualitative data.

Hirotsu and Wang present detailed analyses of the content of courses dealing with quality control and stochastic processes. These analyses again indicate that teaching such topics is not easy at all. In the particular case of stochastic processes a main difficulty is that very advanced mathematical techniques are needed. Additionally, some results in probability and stochastic processes are counter intuitive and researchers can be frustrated in their attempts to understand the same.

However, Wang argues that computer capabilities of simulation and visualisation of abstract ideas and stochastic phenomena might be used to help overcome these difficulties. Conditional probability and the translation of problems from verbal statement to probability formulae are also problematic. As suggested by the author there is scarce didactic material dealing with stochastic processes, in spite of the abundance of formal books on the same, and here is another clear need to continue research on the training of researchers.

Difficulties in understanding statistics do not involve only particular concepts or methods. More importantly researchers do not sufficiently appreciate the role of statistics in the research process and do not realise the series of steps going from theoretical constructs to raw data.

As is remarked by McLean and Iversen, our construction of the world comes from the combination of previous theoretical frameworks, data, statistical analyses and the interpretation of the same. Models are important in the theoretical development of statistics, and modelling takes place at all levels of statistical analysis, although models appear very little in elementary statistical courses. This is perhaps the reason why researchers do not always discriminate between models and data, and do not see models as simplifications of systems.

In McLean's opinion, this can explain the controversy about the nature and role of statistical testing in scientific research (Morrison & Henkel, 1970; Harlow, Mulaik & Steiger, 1997). This discussion was also summarised in Batanero (2000), where I also argue that people confuse not just theoretical hypotheses and data, but they also mix and confuse the different levels of hypotheses in a research study that includes scientific (theoretical), experimental, research and statistical hypotheses.

It is important to make researchers conscious of these different levels of abstraction and also make them understand that finding the perfect model can be less important than finding a simple model that serves to make predictions that are consistent with observations. Researchers should understand that statistical analysis cannot prove that one model is better than another and that statistics cannot be applied in a mechanical way. A lot of personal judgement is needed on the part of the researchers, including the definition of variables and categories, the definition of the model concerned, selection and size of the sample and, of course, the significance level.

In the context of Bayesian statistics, Iversen presents a simple example to show students how different analysis of the same data can lead to very different results. Today, when computers make it easy to perform statistical analysis, people very often carry out complicated calculations, without understanding why they are needed or if they are needed at all, and without thinking about possible alternative methods of analysis.

We use inferential statistics in a non critical way, without reflecting that there are different views about inference and different meanings for the word probability. Moreover, students assign subjective meaning to concepts in classical inference, such as confidence coefficient or significance level. In Iversen’s view Bayesian inference is closer to students' intuitions and brings the model and the data closer. There is however an inertia to teach Bayesian statistics or to apply Bayesian statistics, even in situations where a priori information is available and this results in a lack of feedback for decision making from statistical analyses.

4. DESIGN/EVALUATION OF COURSES FOR TRAINING RESEARCHERS

Ottaviani suggests that similar problems are mentioned in several papers, no matter if the situation described happens in a developed or a developing country. The different type of solutions described by the authors also show the importance of the comprehension and attention which must be given to the local situations, the necessity of enhancing statistics, and the intelligent use of local human resources, tools and equipment.

The problems in the training of researchers are increased by the different ways in which the initial statistical training takes place. While the majority of researchers get their training in traditional courses, in some cases statistics is taught by people with no specific training in statistics, who might contribute to the spreading of all the misconceptions and misuses described. In other cases researchers get their training in statistics by referring to books or by using statistical packages, without any formal training in the topic. In Iversen's words, statisticians have completely lost control of their field.

Some authors present their experiences in training researchers in statistics. For example, Saville describes in detail the contents of workshops that have been successfully run with the aim of introducing basic statistical ideas to agricultural researchers. He gives a list of essential ingredients for these courses: starting from the

beginning, going slow, providing hands-on work with data sets in the area, encouraging participation and interaction, experiencing variability, learning to cope with uncertainty, building confidence and interest in statistics.

A number of authors also agree that the methodology for such courses should be based on encouraging participation and interaction, and working with data sets in their research's areas. In some cases courses focus on the complete process of research as a coherent integrated activity going from the formulation of a research problem to writing the research report (Glencross, & Mji). Examples of innovative solutions include Internet courses (Bishop, Stangl, Lee, and Shia); courses on critical appraisal of statistical analysis in research bibliography (Bangdiwala), in writing research reports, project proposals or papers, or in supervising research (Glencross, & Mji).

The majority of papers describe teaching in a University setting. However, in countries like Japan there are no departments of statistics at Universities. To solve the problem of covering the statistics training needed in order to perform total quality management, Japanese companies have organised their own training systems that involve all the staff and departments. Hirotsu described some courses as well as the role played by the Japanese Standards Association and Japanese Union of Scientists and Engineers to complement these training needs in Japan.

The goals of the INCLEN program (Bangdiwala & Muñoz), were to develop units of excellence in clinical epidemiology research at the participating medical schools in the developing countries. That program not only provided training of physicians in statistics and training of statisticians in clinical epidemiology methods, but also gave participants time for conducting research activities, and provided them with other support required when returning to their countries. Currently, INCLEN has trained over 500 health professionals world-wide and has created Clinical Epidemiology Research and Training Centres (CERTC) in such diverse countries as Brazil, Chile, Colombia, Thailand, the Philippines, and India.

PRESTA (Crivisqui et al) is another successful international training programme in applied statistics for teachers and researchers in South American universities, sponsored by the European Union. The seminars organised in its first quinquennium were attended by 2,500 researchers and lecturers from about 300 universities. A co-operative strategy in five stages was devised to create an autonomous local system for statistical training. This strategy included the organisation of regional seminars to train future trainers in the region, the progressive incorporation of those trainers in the teaching tasks, providing bibliographical and software resources, developing centres for distance education and promoting joint local research projects with the support from European laboratories.

The Universidad Nacional de Colombia (UNC) is the only institution of higher education in Colombia which offers undergraduate and graduate programs (Specialisation, Master and Ph.D.) in statistics. Ospina describes and analyses these programs, including the results of an international evaluation of the Master's programme made by a committee of the American Statistical Association two years ago.

Finally Wei analyses the training needs in statistics produced by the change that China is going through in transferring from planned economy to market economy. China had until recently a huge planning system where local statistical data was collected and analysed, by different levels of statistical offices. Right now, the system employs over 2 million people, and the training of official statisticians is a huge task. In his paper Wei describes the complex statistical training system, types of training programmes, including distance education, and the role played by the Chinese Statistical Society to help in developing this training system.

This last set of papers posed big challenges for statistics education, and, as Ottaviani suggests, they also show how statistics and its teaching are connected with the socio-economic and political situation of a country.

We were glad to know the role played by local associations of statistics education, research resources centres, and international projects to solve these problems in an imaginative way. We hope to be able to see many more similar examples of national and international support to develop research excellence through supporting the training of researchers in statistics and the adequate use of statistics in research in the near future. We agree with Miura that the IASE can play an important role in strengthening international co-operation and promoting research and development in the training of statistics at the different educational levels.

5. THE EFFECTS OF TECHNOLOGY ON THE STATISTICAL TRAINING OF RESEARCHERS

Technology is creating new didactical problems. McDonald describes the experience of Statistics New Zealand in sharing micro data with researchers. As discussed in her paper, statistical agencies and other institutions rely heavily on public trust and good will and this affects their policies of data access and confidentiality. Since these agencies cannot always undertake in-depth analysis of the data they collect, there is an increasing tendency to share some of the data on the request of researchers and this benefits both parties.

It is important that researchers who have access to these facilities carefully consider the data and variables they require to minimise disclosure risk. In some cases using large data sets would imply access to specific software that is unfamiliar to researchers. During their training researchers only deal with data they have collected themselves or with data sets that have being specially prepared for teaching. In this case, data are “error free” and the original file has usually been simplified.

The researcher is then not conscious of the full complexity of data and does not realise all the processes that are applied after the data has been collected, such as coding, recording, editing and transforming the variables. There is also a perception that the data have a high degree of precision since they were collected by an official agency, without being conscious of the limitations due to the sample design and instrument used to collect the data. McDonald recommends that courses directed to these researchers should include all the points above as well as discussion of the advantages and disadvantages of using large complex data sets.

As discussed by Shimada with an example, an important danger is misusing statistical software or using software without a careful evaluation of the same. Since computers and software are widely available, the question about whether a particular analysis is worth doing does not even arise. As discussed by Lee, access to computers increases the requirement for a statistical understanding among scientists and professionals in order to choose the appropriate statistical method and to interpret the results obtained by the computer. The possibilities given by interactive software, enabling each step to be studied before the next is taken, should not be under-rated. Researchers without a deep statistical knowledge just use the standard options since they do not possess a full knowledge of the software possibilities and how they relate to the different types and conditions of statistical analysis.

When the analysis is carried out by a statistician there is also difficulty in communicating the results that arise from complex statistical techniques to users because they do not always have the theoretical background to understand what has been done.

Technology is also offering didactical possibilities. While researchers perceive their data analysis problems to be too specific to be discussed in a general course or seminar, in the experience of Bishop and Talbot, the Internet made it possible for each student or researcher to concentrate on the particular concept or the particular research stage he/she needs to study more in depth. In Korea the government is encouraging partnership among universities and the private sector to share existing resources to provide instruction to students and adults supported by new technologies. The idea of "electronic school" and the changes and possibilities of this new type of education is also discussed by Shia.

Stangl analyses the didactical features of the Internet, which provides interactivity, can adapt to the pace of changes, is accessible to a wide number of students and is flexible. She suggests that it is not aimed to replace the lecturer or the textbook, but to complement both and can make education more individualised. Via discussion groups students and lecturers can interact and come together across great distances and thus facilitate continuing education.

A good design of a course, however, requires a series of decisions as regards the software, slides, scripts, testing, implementation and updating. It also requires a complex production process that will ensure that the project will meet the user’s needs and is sustainable in the long run. All of this adds new research points as well as new training needs for the course developers, lecturers and students.

Finally, as stated by Galmacci, Internet is changing both the way we work in statistics and the way we teach statistics. The large amount of data available from the Internet facilitates teaching based on working with real data and with real projects. It also provides a variety of resources to help researchers in developing their research and in particular to do the elementary or routine data analysis. Faster communication, electronic journals, discussion lists, electronic books and expert systems incorporated in statistical software are contributing to the diffusion and democratisation of statistical knowledge, and at the same time are giving students more responsibility in their own learning.

As stated by Araya, the Internet has brought about a tremendous effort in the education sector. However the traditional ways of keeping close contact between teachers and students and doing some graphics and computation by hand or calculators is also important for beginners. The quick revolution of the past few years has shown the advantage of this new technology; however it is a task for researchers to explore and to exploit these opportunities to improve the teaching and the use of statistics.

6. ERRORS AND ATTITUDES IN THE USE OF STATISTICS BY RESEARCHERS

Some chapters refer to the researchers' errors in using and interpreting statistics methods. Svensson, Lee, and Harraway et al. provide references in various reviews of medical journals, where the poor quality of methodology and statistics in research journals in medicine is highlighted. In analyses of a small number of doctoral dissertations in mathematics education Godino et al. found a variety of incorrect uses and interpretations of statistical methods and results, some of which had been

previously described in research literature, particularly in mathematics education (White, 1980). Their results suggest the difficulty that researchers who have not specialised in statistics find in carrying out their own data analysis and the consequences that this might imply for the quality of their research work.

The incorrect use of statistics is also reflected in the current controversies around statistical tests mentioned before. Several organisations have recently established special committees to study the use of statistics in experimental research. For example, the American Psychological Association (APA), in a 1994 publication manual, noted that significance testing does not reflect the importance or magnitude of an effect, and encouraged researchers to provide effect-size information (American Psychological Association, 1994, p. 18). Subsequently, the Task Force on Statistical Inference established by the APA published an article to initiate discussion in the field prior to revising the APA publication manual (Wilkinson, 1999).

As regards attitudes, Svensson's surveys of researchers with good basic knowledge of statistics showed that contact with a statistician had a low priority for these researchers, due mainly to a lack of experienced statisticians and the lack of a common language between statisticians and researchers. According to Bangdiwala and Muñoz, statisticians are placed on a pedestal and there is not a true collaboration with researchers, which results in poorly planned, conducted, analysed, interpreted or presented research.

Belli's survey suggests that many researchers are reactive and only consult the statistician after their data have been collected, since they are not conscious of the relevance of experimental control and random sampling. According to Jolliffe, sometimes they have already entered the data in a spreadsheet or have attempted some preliminary analysis. Such researchers often present their results to a statistical consultant hoping that a suitable analysis will rescue a poorly designed study (Bishop, & Talbot).

It is necessary to increase the appreciation of statistical work on the part of researchers. Statisticians are viewed as a necessary evil that must provide significant results and there is not a true collaboration with researchers. Their consultancy work is not always valued as it should be and it is difficult to get institutional funding or get co-authorship on published work where the statistician provided substantial input.

Another important issue is the potential disagreement between members within a research group when introducing new statistical approaches (Svensson). Since they are anxious not to lose comparability with other studies there is a preference for traditional statistical methods. Although the reasons for such inertia might also be found in a lack of knowledge, the survey by Svensson also confirmed potential conflicts between the use of non-standard statistical methods in applied research in terms of acceptance by referees and journals.

7. CONSULTATION AND A TEACHING /LEARNING PROCESS

A good opportunity for training occurs when researchers consult with statisticians. In Jolliffe's opinion, the researcher has to communicate the problem to the statistician, including some ideas of relevant theory from the subject area and the aims of the study. The statistician and the researcher should agree about simplifying assumptions, in case they are needed for the statistical analysis and the statistician can help clarify the questions that the researcher wishes to consider. The statistician, in turn, should be able to communicate the statistics techniques and results in a way which is understandable to

the user and also to the audience of the research report.

Saville describes many different types of consultations and the steps where the statistician may give advice. Through this process researchers will learn about research design and statistical analysis, and also about the type of question they can ask a statistician. Statisticians, in turn, will learn teaching and consultancy skills, and will increase their knowledge of statistics through applying new techniques to challenging problems.

Consultancy practices carried out by students is a didactical device that is planned or used in different countries as a means for both providing practical experience to future statisticians, and creating a culture favourable to the value of statistical consultancy by future researchers. In Belli's survey consultants are conscious of their role in helping to develop researchers quantitative thinking. They have the unique opportunity of teaching clients with their own data and examples. They also perceived that a main educational goal for consultants is to change "reactive" researchers who approach the consultant only when the data have been collected to "proactive or collaborative" researchers, that is, clients who count on the statistician from the very beginning of their research.

Consultancy skills are an important part of the statistician' s training and this is mostly done by exposing students to a range of problems from various disciplines. It was noticed in the surveys by Godino et al. that statistics knowledge is only a part of the training needed and that future statisticians feel they lack abilities for producing written reports, communicating with clients, and managing a consulting session, and they also need the support of expert statisticians while doing consultancy practice. Developing ability in problem formulation was another point suggested in Belli's survey. The relevance of making statisticians better acquainted with other research fields for assuring an optimal communication among statisticians and researchers was also made clear in the discussions by different participants.

Since institutional services of consulting are not currently available in some Universities or countries, it is important to create statistics labs and statistics consultancy services in Universities and research centres, not only to help researchers, but also to contribute to the training of both researchers and statisticians. A very interesting example is the research resource centre described by Glencross & Mji where regular on-going research training and consultancy is provided to researchers and postgraduate students. This includes, project planning, writing of research proposals, collecting and analysing data, and providing information about funding resources. However, as pointed out by Ottaviani, it still seems to be very difficult to create a correct “equilibrium” between teaching and practising statistics.

8. LEARNING FROM RESEARCH LITERATURE

An important opportunity to learn statistics is provided by reading statistics materials included in research papers. Researchers, however do not take advantage of this chance and often ignore the methods section when reading research papers in their own scientific fields. This also suggests that they have difficulties in reading this material and thus they cannot critically evaluate research nor effectively collaborate in a research team (Bangdiwala, & Muñoz).

One cause of this situation is that editors of research journals suggest that description of statistical methods be minimised or just replaced by reference to a statistical book, even if the method is new or scarcely used. However, the description of the statistical analysis, including the assumptions made and limitations of models is as

relevant as a description of the method of data collection and of the way the sample was taken. It is not uncommon to have difficulties in getting papers describing new statistical methods or correcting erroneous statistics, published in research journals. We think that an urgent task for statistics education is to write didactical explanations of new statistical methods available in an understandable way for applied scientists.

Finally, since mistakes and misinterpretations in the use of statistics are frequent in published research, these errors and misuses might be spread when new researchers try to find how statistics was used in a particular research or try to replicate a given research. An important recommendation is that top quality research journals include a statistical review of submitted papers in addition to the traditional scientific review, before accepting a paper for publication.

ADDITIONAL REFERENCES

American Psychological Association (1994). *Publication Manual of the American Psychological Association* (Fourth edition.). Washington, DC: American Psychological Association.

Batanero, C (2000). Controversies around the role of statistical tests in experimental research, *Journal of Mathematics Thinking and Learning*, 2(1-2), 75-98.

Batanero, C., Estepa, A., & Godino, J. D. (1997). Evolution of students' understanding of statistical association in a computer based teaching environment. In J. B. Garfield, & G. Burril (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 191-205). Voorburg, The Netherlands: International Association for Statistical Education and International Statistical Institute.

Batanero, C., Godino, J., & Estepa, A. (1998). Building the meaning of statistical association through data analysis activities. In A. Olivier, & K. Newstead (Eds.), *Proceedings of the 22nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 221-236). Stellenbosh, South Africa: University of Stellenbosh.

Beyth-Marom, R. (1982). Perception of correlation reexamined. *Memory and Cognition*, 10(6), 511-519.

Elmore, P. B., & Woehlke, P. L. (1988). Research methods employed in *American Educational Research Journal, Educational Researcher,* and *Review of Educational Research* from 1978 to 1987. Paper presented at the *Annual Meeting of the American Educational Research Association*, New Orleans.

Elmore, P. B., & Woehlke, P. L. (1998). Twenty years of research methods employed in *American Educational Research Journal, Educational Researcher* and *Review of Educational Research*. Paper presented at the *Annual Meeting of the American Educational Research Association*, San Diego.

Hammer, A. S., & Buffington, C. A. (1994). Survey of statistical methods used in the veterinary medical literature. *Journal of the American Veterinary Medical Association*, 205, 344-345.

Hand, D. J. (1996). Statistics and the theory of measurement. (with discussion). *Journal of the Royal Statistical Society, Series* A, 159, 445-492.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.

Morris, E. J. (1997). An investigation of students' conceptions and procedural stills in the statistical topic correlation. *Centre for Information Technology in Education*, Report n. 230. Milton Keynes (U.K): The Open University.

Morrison, D. E., & Henkel, R. E. (Eds.) (1970), *The significance tests controversy. A reader* Chicago: Aldine.

Truran, J. M. (1997). Understanding of association and regression by first year economics students from two different countries as revealed in responses to the same examination questions. In J. Garfield, & J. M. Truran, (Eds.), *Research papers on stochastics educations*

from 1997 (pp. 205-212). Minneapolis, MN: University of Minnesota.
White, A. L. (1980). Avoiding errors in educational research. In R. J. Shumway (Ed.), *Research in mathematics education* (pp. 47-65). Reston, VA: National Council of Teachers of Mathematics.
Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry, *International Statistical Review*, 67(3), 223-265.
Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604.

Carmen Batanero,
Department of Mathematics Education,
Faculty of Education, University of Granada,
Campus de Cartuja, 18071 Granada, Spain
E-mail: batanero@ugr.es

“There is no doubt that Training Researchers in the Use of Statistics is very important to improve the quality of empirical research and to foster technical and economical development. However, since the logic of statistical inference is difficult to grasp, its use and interpretation are not always adequate and have been criticised for nearly 50 years. In this book the reader will find various analysis of the problems related to the training of researchers, and a number of views of ways in which some of these problems might be solved."

ISBN: 90-73592-19-4