The Challenges of Teaching Statistical Inference

Carmen Batanero

University of Granada, Spain

Abstract

Statistics is today taught with almost no exception in all University Majors. However, the use and interpretation of statistical inference are not always adequate and have been criticized for nearly 50 years. Moreover, psychological research shows that even statistically well-trained professionals tend to make erroneous stochastic judgments. In this presentation we summarize some common errors in the use of statistical inference and reflect on possible implications for teaching.

Introduction

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. In spite of this, researchers' excessive confidence and interpretation of statistical inference results have been criticized for years Debates about the misapplication of significance testing have arisen within the social and health sciences, particularly as regards interpretation of p-values and taking significance as the main criteria to accept papers for publication. Other common mistakes in making decisions that involve conditional probability, association, or the Bayes' theorem have been consistently found among both students and researchers. Below we describe some of these issues, that were object of debate in the IASE 2000 Round Table Conference (Batanero, 2001) and finally reflect on possible changes in our teaching practice that can lead to improve the use of statistical inference in applied research.

Research on Stochastic Reasoning and Learning

Psychological research on adults' stochastic reasoning started with the' *heuristics and biases* programme (Kahneman, Slovic & Tversky; 1982), according to which people do not follow the normative mathematical rules that guide formal scientific inference when they make a decision under uncertainty. Instead, they use more simple judgmental heuristics that reduce the complex probabilistic tasks to simpler judgmental operations and are in general useful; however they sometimes cause serious and systematic errors and are resistant to change. For example in the *representativeness*

heuristics, people tend to estimate the likelihood for an event based on how well it represents some aspects of the parent population. An associated fallacy that has been termed *belief in the law of small numbers* is the belief that even small samples should exactly reflect all the characteristics in the population distribution. These errors also extend to problems solved in a teaching setting and have been widely documented in relation to concepts such as randomness, compound probability, association in contingency tables, conditional probabilities, Bayes problems and sampling (Sedlemeier, 1999; Jones, & Thornton, 2005).

A more recent theoretical framework is the *adaptive algorithms* approach (Cosmides & Tooby, 1996; Gigerenzer, 1994). Adaptive algorithms serve to solve adaptive problems (such as finding food, avoiding predation or communicating) and take a long time to be shaped, due to natural selection. Since their natural environments shape adaptive algorithms they are more effective when the tasks are presented in a format close to how data are perceived and remembered in ordinary life. According to this theory people should have little difficulty in solving statistical tasks if the data are presented in a natural format of frequencies (absolute frequencies) instead of using rates or percentages because this lead to simpler algorithms that give immediate accessible solution to many statistical problems (Gigerenzer, 1994)

Sedlmeier (1999) analyses and summarises recent teaching experiments carried out by psychologists and suggest that statistical training is effective if students are taught to translate statistical tasks to an adequate format, including tree diagrams and absolute frequencies. However, learning is assessed through participant's performance to tasks that are very close to those used in the training, so that it is difficult to evaluate the extent to which students would be able to transfer this knowledge to a wider type of problems. While this research provides relevant empirical information and potential theoretical explanations for students' difficulties in advanced stochastics, there is still a large amount of work to be done by mathematics and statistics educators to integrate these results and use them to design and evaluate teaching sequences in natural settings where students are expected to meet wider curricular requirements.

Another branch of educational research comes from the field of statistics, where interest in education arose since the creation of the Education Committee by the International Statistical Institute (ISI (Vere-Jones, 1995). The International Conferences on Teaching Statistics (ICOTS) were started in 1982 by the ISI to bring together statistics teachers at all levels, disciplines and countries and have continued every four

years. These conferences were complemented with a series of Round Table Conferences focussed on specific themes. In 1991 the International Association for Statistics Education (IASE) was created as a separate section of the ISI. The journals *Teaching Statistics* first published in 1979 and *Journal of Statistics Education*, started in 1993, soon became main tools to improve statistics education all over the world. Research in advanced stochastics is also starting to progressively grow under the influence of the IASE. Although this research follows a variety of approaches, some research agendas (e.g. Shaughnessy, 1992; Shaughnessy, Garfield & Greer, 1996; Batanero, Garfield, Ottaviani & Truran, 2000), as well as the creation of *Statistics Education Research Journal* with the specific purpose to promote research show that researchers are now interested in linking the isolated pieces of research and in developing general principles and theoretical knowledge in order to conduct quality research in statistical education.

Common Problems in the Use of Statistics Inference

Psychological and educational research has shown widespread misconceptions among scientists who use statistical inference in their daily work, including those who received a previous statistical training (see Harlow, Mulaik & Steiger, 1997; Glaser, 1999). These errors concern particularly the test of hypotheses and lead to a paradoxical situation, where, on one hand, a significant result is required to get a paper published in many journals and, on the other hand, significant results are misinterpreted in these publications (Falk& Greenbaum, 1995; Lecoutre & Lecoutre, 2001).

A particularly misunderstood concept is the *level of significance, α,* which is defined as the probability of rejecting a null hypothesis, given that it is true. The most common misinterpretation of this concept consists of switching the two terms in the conditional probability that is, interpreting the level of significance as the probability that the null hypothesis is true, once the decision to reject it has been taken. For example, Birnbaum (1982) reported that his students found the following definition reasonable: "A level of significance of 5% means that, on average, 5 out of every 100 times we reject the null hypothesis, we will be wrong". Falk (1986) found that most of her students believed that α was the probability of being wrong when rejecting the null hypothesis at a significance level α. Similar results were found by Krauss and Wassner (2002) in University lecturers involved in the teaching of research methods. More specifically they found that 4 out of every 5 *methodology instructors* have misconceptions about the concept of significance, just like their students. The level of

significance is not the only concept misunderstood in significance testing, but there is also confusion between the roles of the null and alternative hypotheses as well as between the statistical alternative hypothesis and the research hypothesis (Chow, 1996).

The wide research on students' and professionals' misunderstanding and misuse of statistical tests has not been followed by related efforts in designing and evaluating teaching experiments oriented to help students and researchers overcome these difficulties. Recommendations to substitute or complement statistical tests with confidence intervals (e.g. Wilkinson, 1999) do not take into account the fact that their appealing feature is based on a fundamental misunderstanding (Lecoutre, 1998), which consists of thinking of the parameters as random variables and assuming that confidence intervals contain the parameters with a specified probability.

Bayes' thinking is more intuitive than the frequentist probability for students and reflects their everyday thinking about uncertainty better. It is for this reason that some researchers (e.g. Lecoutre & Lecoutre, 2001) are suggesting changing the practice and teaching of statistics towards Bayesian methods. However, although there is an increasing number of publications about how best introduce Bayesian concepts to students coming from non scientific specialities (e. g. Berry, 1997, Albert & Rossman, 2001) reported results from experiments or research focussed on teaching Bayesian statistics are very limited. Moreover, some experiences reported suggest that students can make mistakes in interpreting their inferential results or in specifying their prior distributions (Albert, 2000).

Given the relevance of a correct understanding and application of statistical inference to improve empirical research (including research in mathematics education) this is a possible area where research is an urgent need. Both the teaching of classical and inferential statistics are clear priority areas for further research in advanced stochastics.

Implications for Teaching

Ten years ago, statistics teaching at University level used to emphasize the teaching of formulas for calculating statistics (e.g. correlation coefficients or confidence intervals) without much concern towards the data context or interpretative activities. In other cases, the courses were over-mathematised for students without any prior or experience of advanced algebra or calculus. As a consequence many students finished the courses been able to manipulate definitions and algorithms with apparent

competence, but lacking understanding of the connections among the important concepts of the discipline (Schau & Mattern, 1997) and ability to decide what statistical procedure to apply when facing a real problem (Quilici & Mayer, 1996).

Fortunately, increasingly easy access to powerful computing facilities has saved time previously devoted to laborious calculations and encouraged less formal, more intuitive approaches to statistics (Garfield, Chance & Snell, 2001; Biehler, 2003). Consequently, changes are recommended in statistics teaching at University level, both in course content and in the teaching approach (Moore, 1997). The effectiveness of these changes will depend, however, of the extent to which we take into account results from educational research that we are summarising below.

Different orientation for different students

On the other hand, the influences from different philosophical views about randomness, probability and statistical inference are still reflected in stochastics teaching and research: (Batanero, Henry & Parzysz, 2005). These conceptions influence changes in the role given to probability within the curriculum, from being the central core to trying to teach statistics without resort to probability (focus on exploratory data analysis only) or favouring classical, Bayesian, computer-intensive (resampling methods) or mathematical-abstract approaches to inference. However, there is no consensus about which of these approaches is better suited for a particular type of student or which are best ways to introduce a given approach. We need to find which approach is needed for each type of student and which are the best ways to introduce a given approach.

In spite of our efforts intuitions do not progress with teaching and misconceptions remain after formal training in statistics. We should concentrate on clarifying why current teaching of statistics does not improve stochastic intuition and think of some alternative ways, for example including some elements of psychology or philosophy in this teaching. *"In any case statistics should be taught in conjunction with material on intuitive strategies and inferential errors* (Nisbett & Ross, 1980, p.281).

Web based teaching

Advances in technology and increasing student enrolment numbers have led many universities to offer on-line courses although few studies have compared online and traditional methods of teaching of advanced statistics and results are inconclusive.

While Hilton & Christensen (2002) evaluated the impact of incorporating multimedia presentations into the traditional lecture format and found that it did not improve student's learning or attitudes, Utts et al., (2003) found that students in a traditional and mixed setting (using Internet and traditional teaching) had similar performance. At the IASE Satellite Conference on Statistics Education and the Internet, most presentations focussed on analysing Internet resources for teaching statistics or presenting examples of these resources.

Theoretical frameworks that take into account the new meanings of statistical concepts in virtual environments and research methods that are useful to analyse the enormous amount of information generated in the interaction between students and teacher in such as environment are an urgent need. Given the increasing access of students to virtual learning and the extensive and varied resources available there is a need to analyse how the meaning of statistics concepts is affected by these resources and how can we incorporate them in the teaching both in group and individual learning.

The nature of advanced stochastics

Whilst in other areas of mathematics the boundaries between elementary and advanced thinking are reasonably well defined, such a distinction is much more fuzzy in the case of stochastics, where current secondary school curricula in many countries include ideas about association and inference. Also, situations which require mature stochastic thinking for correct interpretation, such as voting, investment, research planning or quality control increasingly form part of the information to which many citizens and professionals are exposed. So, while advanced mathematical thinking tends to be used only in a formal way and after some systematic training, advanced stochastic thinking is now being formally or informally used by many people with little formal training in either mathematics or stochastics.

Moreover some apparently simple concepts, such as randomness, independence and variation are complex because each of them describes a separate continuum, are interconnected and usually are taken for granted in teaching (Gal, 2005). We therefore need to reflect on the exact amount of formalism needed to teach these concepts and what the best way to use computers to help us in this teaching is In this sense, research in statistics and probability can be paradigmatic for finding ways and approaches to introduce advanced mathematical ideas to wider audiences, and to rethink the very meaning of what is advanced mathematical thinking.

References

Albert, J. (2000). Using a sample survey project to assess the teaching of statistical inference, *Journal of Statistical Education,* 8. On line: [http://www.amstat.org/publications/jse/secure/v8n1/albert.cfm]

Albert, J. H., & Rossman, A. (2001). *Workshop statistics. Discovery with data. A bayesian approach.* Bowling Green. OH: Key College Publishing.

Batanero, C. (2001) (Ed.). *Training researchers in the use of statistics. Proceedings of the 1996 IASE Round Table Conference.* Granada, Spain: International Statistical Institute.

Batanero**,** C., Garfield, J. B., Ottaviani, M. G., & Truran, J. (2000). Research into statistical education: Some priority questions. *Statistics Education Research Newsletter* 1(2), with discussion in *SERN* 2(1) and response in *SERN* 2(2). On line: [http://www.ugr.es/~batanero/Newsletters/Newsjan01.PDF]

Batanero, C., Henry, M. & Parzycz, B. (2005). The nature of chance and probability. In G. Jones (Ed.). *Exploring probability in schools: Challenges for teaching and learning.* (pp. 16-42). Dordrecht, The Netherlands: Kluwer

Berry, D. A. (1995). *Basic statistics: A Bayesian perspective.* Wadsworth: Belmont.

Biehler, R. (2003b) Interrelated learning and working environments for supporting the use of computer tools in introductory courses. CD-ROM *Proceedings IASE Satellite conference on Teaching Statistics and the Internet.* Berlin: IASE.

Birnbaum, I. (1982). Interpreting statistical significance. *Teaching Statistics, 4*, 24–27.

Chow, L. S. (1996*). Statistical significance: Rationale, validity and utility.* London: Sage.

Cosmides, L. & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58, 1-73.

Falk, R. (1986). Misconceptions of statistical significance. *Journal of Structural Learning, 9*, 83–96.

Falk, R., & Greenbaum, C. W. (1995). Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory and Psychology, 5* (1), 75-98.

[http://www.amstat.org/publications/jse/v4n1].

Gal, I. (2005). Towards "probability literacy" for all citizens: building blocks and instructional dilemmas. In G. Jones (Ed.), *Exploring probability in school. Challenges for teaching and learning* (pp. 39-64). Dordrecht. The Netherlands: Kluwer.

Garfield, J., Chance, B., & Snell, J. L. (2001).Technology in college statistics courses. In D. Holton et al. (Eds.); *The teaching and learning of mathematics at university level: An ICMI study* (pp. 357-370). Dordrecht: Kluwer.

Garfield, J. B., del Mas, B., & Chance, B. (1997). Assessing the effects of a computer microworld on statistical reasoning. Paper presented at the *Join Meetings of the American Statistical Association,* Ana heim, CA, August, 1997. On line [www.gen.umn.edu/faculty_staff/delmas/asa_1997_paper/asa_paper.htm].

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice-versa). In G. Wright & P. Ayton (Eds.), Subjective probability (pp. 129-161). Chichester: Wiley.

Glaser, D. N. (1999). The controversy of significance testing: Misconceptions and alternatives. *American Journal of Critical Care,* 5. 291-296.

Godino, J. D., & Batanero, C. (1998). Clarifying the meaning of mathematical objects as a priority area of research in mathematics education. In A. Sierpinska & J. Kilpatrick (Eds.), Mathematics Education as a research domain: A search for identity (pp. 177–195). Dordrecht: Kluwer.

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

Hilton, S. C. & Christensen, B. H. (2002). Evaluation of the impact of multimedia lectures on student learning and attitudes. In B. Phillips (Ed.), *Proceedings of the sixth International Conference on the Teaching of Statistics.* Cape Town: International Statistical Institute. CD ROM.h

Jones, G. A., & Thornton, C. A. (2005). An overview of research into the teaching and learning of probability. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 65-92).. Dordrecht, The Netherlands: Kluwer Academic Publishers.

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

Krauss, S. & Wassner, C. (2002). How significance tests should be presented to avoid the typical misinterpretations. In B. Phillips (Ed.), *Proceedings of ICOTS-6.* CD Rom. Ciudad del Cabo: International Association for Statistical Education.

Lecoutre, B. (1998). Teaching bayesian methods for experimental data analysis. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth nternational Conference on Teaching of Statistics* (v.1, pp. 239-244). Singapore: International Statistical Institute.

Lecoutre, B., & Lecoutre, M. P. (2001). Uses, abuses and misuses of significance tests in the scientific commumity: Won't the Bayesian choice be unavoidable? *International Statistical Review,* 69(3)-399-417.

Moore, D. S. (1997). New pedagogy and new content: the case of statistics. *International Statistical Review*, 635, 123-165.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgments.* Englewood Cliffs, NJ: Prentice Hall.

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how sutdents learn to categorise statistics word problems. *Journal of Educational Psychology,* 88(1), 144-161.

Rossman, A. (1997). Workshops statistics: Using technology to promote learning by self- discovery. In J. Garfield, & G. Burrill (Eds.). *Research on the Role of Technology in Teaching and Learning Statistics*. International Statistical Institute.

Sedlmeier, P. (1999). *Improving statistical reasoning. Theoretical models and practical implications.* Mahwah, NJ: Erlbaum.

Shaughnessy, J. M., Garfield, J., & Greer, B. (1996). Data handling. In A. Bishop et al. (Eds.), *International handbook of mathematics education* (v.1, pp. 205-237). Dordrecht, Netherlands: Kluwer.

Utts, J., Sommer, B., Acredolo, C., Maher, M. W. & Matthews, H. R. (2003). A study comparing traditional and hybrid Internet- based instruction in introductory statistics classes. *Journal of Statistics Education,* 11(3). On line [http://www.amstat.org/publications/jse/v11n3].

Vere-Jones, D. (1997). The coming of age of statistical education. *International Statistical Review,* 63(1), 3-23.