

TALLER SOBRE ANALISIS EXPLORATORIO DE DATOS EN LA ENSEÑANZA SECUNDARIA

Carmen Batanero, Universidad de Granada
(<http://www.ugr.es/local/batanero/>)

[Actas de la Conferencia Internacional "Experiências e Expectativas do Ensino de Estatística - Desafios para o Século XXI". Florianópolis, Santa Catarina, Brasil - 20 a 23 de Setembro de 1999]

RESUMEN

El análisis exploratorio de datos, introducido Tukey (1962; 1970), se ha extendido como filosofía de aplicación de la estadística, debido principalmente a la disponibilidad de ordenadores y software estadístico con posibilidades de representación gráfica y tratamiento de conjuntos de datos variados. Las posibilidades didácticas del análisis exploratorio de datos se deben a la sencillez del aparato matemático requerido, la importancia dada hoy día en estadística y matemáticas a los sistemas de representación múltiple y resolución de problemas, las conexiones con otros temas del curriculum, el trabajo en equipo y la posibilidad de desarrollo de proyectos por parte de los alumnos (Batanero, Estepa y Godino, 1992).

El objetivo del taller es presentar a los profesores ejemplos de actividades de análisis exploratorio de datos que puedan utilizar con sus alumnos de secundaria. Un proyecto de análisis de datos, a partir de un fichero tomado de Internet, servirá para analizar los principales contenidos de análisis de datos abordables en este nivel de enseñanza, mostrar ejemplos de cuestiones que requieran el uso de conceptos y técnicas estadísticas, describir algunas dificultades previsibles de los alumnos con los mismos y sugerir criterios para el trabajo en clase con los alumnos. La metodología del taller alternará el trabajo en grupo y discusión de los profesores y profesoras y el resumen de algunos puntos claves por la coordinadora del taller. Se propone dividir el taller en dos sesiones con un breve descanso.

Un ejemplo de proyecto: Análisis demográfico

La actividad se desarrolla en torno a un proyecto a partir de un fichero que contiene datos de 97 países y que ha sido adaptado del preparado por Rouncenfield (1995), quien usó como fuentes Day (1992) y U.N.E.S.C.O. (1990). Este fichero ha sido tomado de Internet, del servidor de Journal of Statistical Education (<http://www.amstat.org/publications/jse/>). Contiene las siguientes variables, que se refieren a 1990:

Tasa de natalidad: Niños nacidos vivos en el año por cada 1000 habitantes;

Tasa de mortalidad: Número de muertes en el año por cada 1000 habitantes;

Mortalidad infantil: Número de muertes en el por cada 1000 niños de menos de 1 año;

Esperanza de vida al nacer para hombres y mujeres;

PNB. Producto Nacional Bruto per cápita en dólares (USA);

Grupo: Clasificación de países en función de la zona geográfica y situación económica, en las siguientes categorías:

1 = Europa Oriental

2 = Iberoamérica

3 = Europa Occidental, Norte América, Japón, Australia, Nueva Zelanda

4 = Oriente Medio

5 = Asia

6 = Africa.

Hemos añadido el número de habitantes en 1990 en miles de personas (*Población*), tomado del anuario publicado por el periódico español "El País". En la tabla 1 listamos los datos del proyecto:

Tabla 1: Fichero de datos del proyecto "Análisis demográfico"

País	Grupo	Tasa natalidad	Tasa mortalidad	Mortalidad infantil	Esperanza vida hombre	Esperanza vida mujer	PNB	Población (miles)
Afganistán	5	40.4	18.7	181.6	41.0	42.0	168	16000
Albania	1	24.7	5.7	30.8	69.6	75.5	600	3204
Alemania (Oeste)	3	11.4	11.2	7.4	71.8	78.4	22320	16691
Alemania Este	1	12.0	12.4	7.6	69.8	75.9	.	61337
Algeria	6	35.5	8.3	74.0	61.6	63.3	2060	24453
Angola	6	47.2	20.2	137.0	42.9	46.1	610	9694
Arabia Saudí	4	42.1	7.6	71.0	61.7	65.2	7050	13562
Argentina	2	20.7	8.4	25.7	65.5	72.7	2370	31883
Austria	3	14.9	7.4	8.0	73.3	79.6	17000	7598
Bahrein	4	28.4	3.8	16.0	66.8	69.4	6340	459
Bangladesh	5	42.2	15.5	119.0	56.9	56.0	210	111590
Bélgica	3	12.0	10.6	7.9	70.0	76.8	15540	9886
Bielorusia	1	15.2	9.	13.1	66.4	75.9	1880	.
Bolivia	2	46.6	18.0	111.0	51.0	55.4	630	7110
Botswana	6	48.5	11.6	67.0	52.3	59.7	2040	1217
Brasil	2	28.6	7.9	63.0	62.3	67.6	2680	147294
Bulgaria	1	12.5	11.9	14.4	68.3	74.7	2250	9001
Camboya	5	41.4	16.6	130.0	47.0	49.9	.	8250
Canadá	3	14.5	7.3	7.2	73.0	79.8	20470	26302
Colombia	2	27.4	6.1	40.0	63.4	69.2	1260	32335
Congo	6	46.1	14.6	73.0	50.1	55.3	1010	2208
Corea (Norte)	.5	23.5	18.1	25.0	66.2	72.7	400	21143
Checoslovaquia	1	13.4	11.7	11.3	71.8	77.7	2980	15641
Chile	2	23.4	5.8	17.1	68.1	75.1	1940	12980
China	5	21.2	6.7	32.0	68.0	70.9	380	1105067
Dinamarca	3	12.4	11.9	7.5	71.8	77.7	22080	5132
Ecuador	2	32.9	7.4	63.0	63.4	67.6	980	10329
Egipto	6	38.8	9.5	49.4	57.8	60.3	600	51390
Emiratos Arabes	4	22.8	3.8	26.0	68.6	72.9	19860	1544
España	3	10.7	8.2	8.1	72.5	78.6	11020	39161
Etiopía	6	48.6	20.7	137.0	42.4	45.6	120	48861
Filipinas	5	33.2	7.7	45.0	62.5	66.1	730	61224
Finlandia	3	13.2	10.1	5.8	70.7	78.7	26040	4974
Francia	3	13.6	9.4	7.4	72.3	80.5	19490	56119
Gabón	6	39.4	16.8	103.0	49.9	53.2	390	1105
Gambia	6	47.4	21.4	143.0	41.4	44.6	260	848
Ghana	6	44.4	13.1	90.0	52.2	55.8	390	14425
Grecia	3	10.1	9.2	11.0	65.4	74.0	5990	10039
Guayana	2	28.3	7.3	56.0	60.4	66.1	330	95
Holanda	3	13.2	8.6	7.10	73.3	79.9	17320	14828
Hong_Kong	5	11.7	4.9	6.10	74.3	80.1	14210	5735
Hungría	1	11.6	13.4	14.8	65.4	73.8	2780	10587
India	5	30.5	10.2	91.0	52.5	52.1	350	832535

Indonesia	5	28.6	9.4	75.0	58.5	62.0	570	178211
Irán	4	42.5	11.5	108.1	55.8	55.0	2490	50204
Iraq	4	42.6	7.8	69.0	63.0	64.8	3020	18271
Irlanda	3	15.1	9.1	7.5	71.0	76.7	9550	3537
Israel	4	22.3	6.3	9.7	73.9	77.4	10920	4525
Italia	3	9.7	9.1	8.8	72.0	78.6	16830	57537
Japón	3	9.9	6.7	4.0	75.9	81.8	25430	123045
Jordania	4	38.9	6.4	44.0	64.2	67.8	1240	4041
Kenya	6	47.0	11.3	72.0	56.5	60.5	370	23277
Kuwait	4	26.8	2.	15.6	71.2	75.4	16150	2020
Líbano	4	31.7	8.7	48.0	63.1	67.0	.	2900
Libia	6	44.0	9.4	82.0	59.1	62.5	5310	4395
Malasia	5	31.6	5.6	24.0	67.5	71.6	2320	17340
Malawi	6	48.3	25.0	130.0	38.1	41.2	200	8230
Marruecos	6	35.5	9.8	82.0	59.1	62.5	960	24567
México	2	29.0	23.2	43.0	62.1	66.0	2490	85440
Mongolia	5	36.1	8.8	68.0	60.0	62.5	110	2128
Mozambique	6	45.0	18.5	141.0	44.9	48.1	80	15357
Namibia	6	44.0	12.1	135.0	55.0	57.5	1030	1300
Nepal	5	39.6	14.8	128.0	50.9	48.1	170	18431
Nigeria	6	48.5	15.6	105.0	48.8	52.2	360	113665
Noruega	3	14.3	10.7	7.8	67.2	75.7	23120	4215
Omán	4	45.6	7.8	40.0	62.2	65.8	5220	1486
Pakistán	5	30.3	8.1	107.7	59.0	59.2	380	109950
Paraguay	2	34.8	6.6	42.0	64.4	68.5	1110	4161
Perú	2	32.9	8.3	109.9	56.8	66.5	1160	21142
Polonia	1	14.3	10.2	16.0	67.2	75.7	1690	38061
Portugal	3	11.9	9.5	13.1	66.5	72.4	7600	10333
Rumania	1	13.6	10.7	26.9	66.5	72.4	1640	23148
Sierra Leona	6	48.2	23.4	154.0	39.4	42.6	240	4040
Singapur	5	17.8	5.2	7.5	68.7	74.0	11160	2664
Somalia	6	50.1	20.2	132.0	43.4	46.6	120	6089
Sri_Lanka	5	21.3	6.2	19.4	67.8	71.7	470	16779
Sudáfrica	6	32.1	9.9	72.0	57.5	63.5	2530	34925
Sudán	6	44.6	15.8	108.0	48.6	51.0	480	24423
Suecia	3	14.5	11.1	5.6	74.2	80.0	23660	8485
Suiza	3	12.5	9.5	7.1	73.9	80.0	34064	6541
Swazilandia	6	46.8	12.5	118.0	42.9	49.5	810	761
Tailandia	5	22.3	7.7	28.0	63.8	68.9	1420	55200
Tanzania	6	50.5	14.0	106.0	51.3	54.7	110	25627
Túnez	6	31.1	7.3	52.0	64.9	66.4	1440	7988
Turquía	4	29.2	8.4	76.0	62.5	65.8	1630	54899
U.K.	3	13.6	11.5	8.4	72.2	77.9	16100	57270
U.S.A.	3	16.7	8.1	9.1	71.5	78.3	21790	248243
Ucrania	1	13.4	11.6	13.0	66.4	74.8	1320	.
Uganda	6	52.2	15.6	103.0	49.9	52.7	220	16722
Uruguay	2	18.0	9.6	21.9	68.4	74.9	2560	3067
URSS	1	17.7	10.0	23.0	64.6	74.0	2242	287664
Venezuela	2	27.5	4.4	23.3	66.7	72.8	2560	19244
Vietnam	5	31.8	9.5	64.0	63.7	67.9	.	65758
Yugoslavia	1	14.0	9.0	20.2	68.6	74.5	.	23707
Zaire	6	45.6	14.2	83.0	50.3	53.7	220	34442
Zambia	6	51.1	13.7	80.0	50.4	52.5	420	7837
Zimbabwe	6	41.7	10.3	66.0	56.5	60.1	640	9567

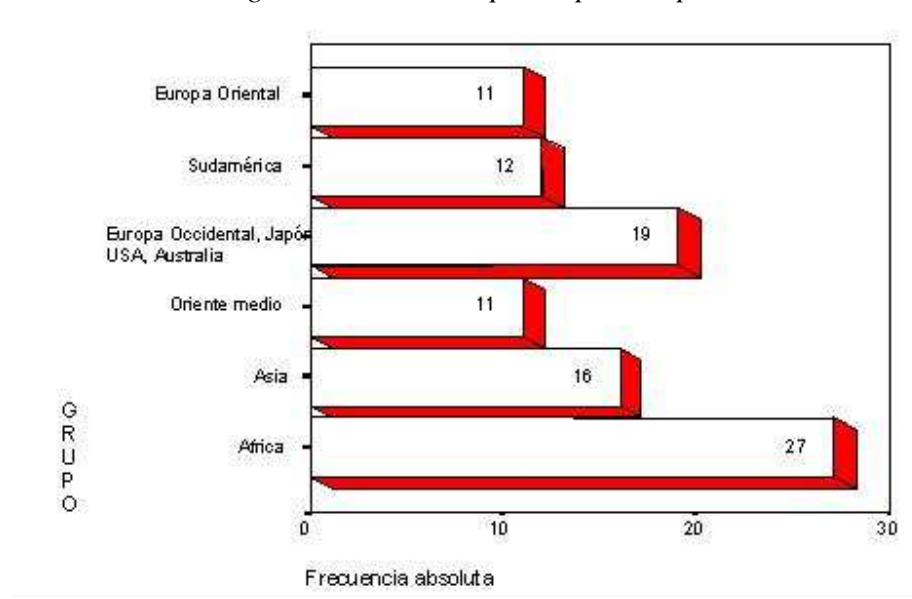
Rouncenfield (1995) presenta este proyecto sobre el cual los alumnos pueden trabajar en grupos, comparando las variables en los diferentes grupos de países y formulando por si mismos preguntas de su interés, incluso completando las variables del fichero cuando la solución de las preguntas así lo requiera. Mostramos a continuación algunos ejemplos de las posibles actividades a desarrollar en un curso de análisis exploratorio de datos, esperando que propicien la reflexión sobre los contenidos del curriculum y las dificultades previsibles de los alumnos en su proceso de aprendizaje. Otras actividades se sugieren en el artículo de Rouncenfield.

Actividad 1. Variables: La actividad inicial consiste en discutir el significado de las variables de este fichero y analizar como se han calculado las diferentes tasas: natalidad, mortalidad, esperanza de vida, PNB. Los alumnos podrían investigar qué otros indicadores alternativos se emplean para obtener un indicador demográfico o económico de la riqueza de un país. El profesor podría pedir a los alumnos que busquen artículos en la prensa en que se hable de alguno de estos indicadores y que expliquen con sus propias palabras la utilidad que pueden tener y que averigüen quien y como los calcula.

En este fichero se ha usado un código para agrupar los países en función de la zona geográfica y desarrollo económico. Los alumnos podrían sugerir otras variables de clasificación de los países o añadir otras variables de países al fichero.

Actividad 2. Diagramas de barras: La Figura 1 muestra un diagrama de barras donde representamos el número de países en los diferentes grupos de países. A partir del mismo es fácil construir una tabla de frecuencias y discutir el significado de las frecuencias absolutas, relativas y porcentajes. Los alumnos pueden analizar las ventajas que el diagrama de barras tiene frente a la tabla para visualizar el grupo que tiene mayor /menor número de países. Asimismo pueden elaborar otros gráficos adecuados para representar esta variable.

Figura 1. Número de países por Grupo



Actividad 3. Promedios: La elaboración de una tabla o un gráfico ya supone una primera reducción de los datos, pero a veces queremos hallar un único valor representativo de la distribución. En la Figura 2 hemos representado el número promedio de habitantes en cada país, según grupo, usando dos promedios diferentes: media y mediana.

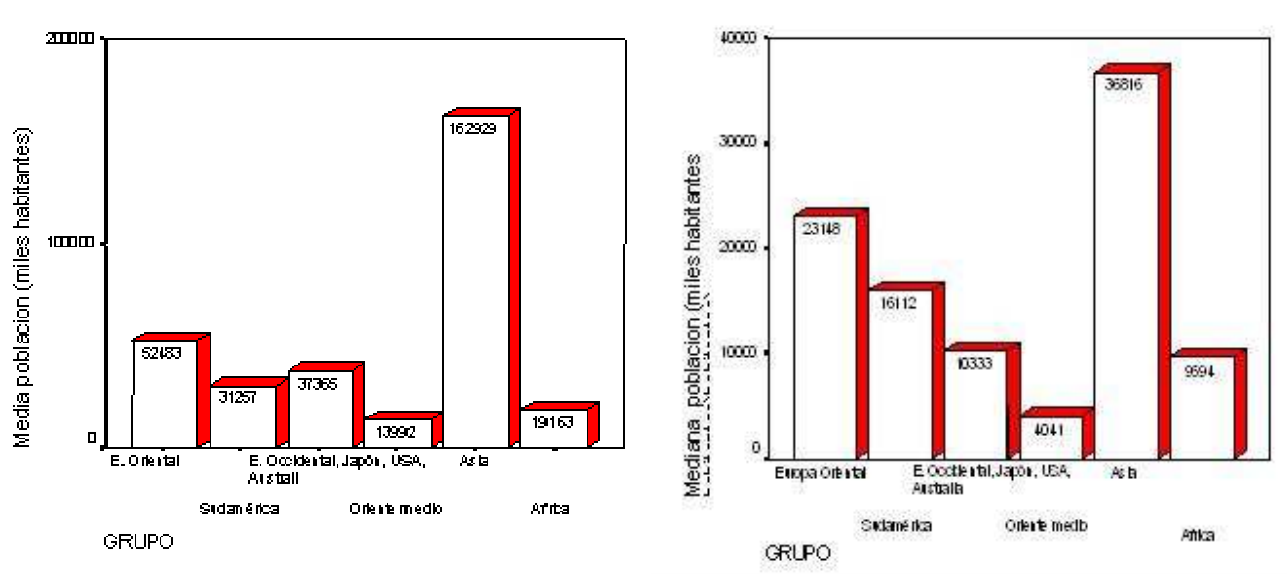


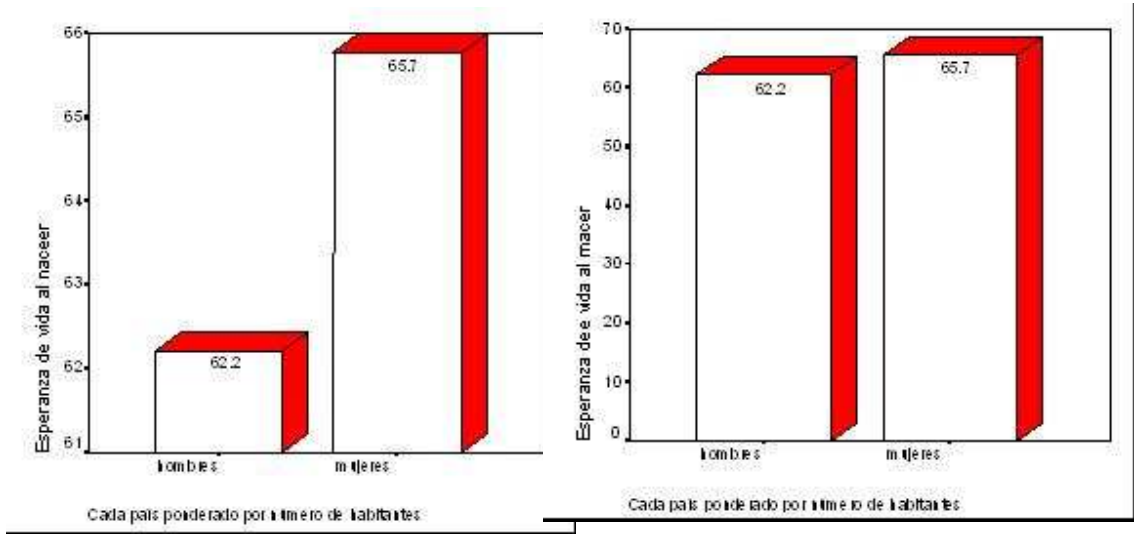
Figura 2: Mediana y media del número de habitantes en los diferentes grupos de países

La clase puede dividirse en grupos para calcular estos promedios, así como la moda, y para explicar lo que representa cada uno de estos promedios y elegir en cada grupo el que mejor lo representa, argumentando la elección. Se puede pedir a los alumnos que señalen las principales diferencias entre los dos gráficos y que decidan cuál de los dos promedios acentúa más las diferencias explicando la razón.

Otra pregunta para los alumnos es qué representa el valor obtenido al calcular la media de la esperanza media de vida al nacer en estos 97 países y cómo habría que hacer para calcular la esperanza media de vida al nacer en hombres y mujeres, si no tenemos en cuenta el país de nacimiento. Con ejemplos sencillos se podría conducir a los alumnos a la idea de media ponderada y hacerles ver la necesidad de tener en cuenta el número de hombres y mujeres de cada país, para calcular el promedio.

Nosotros hemos calculado la esperanza media de vida global en hombres y mujeres, ponderando los datos de cada país por su número de habitantes (suponiendo un número aproximadamente igual de hombres y mujeres). En la Figura 3 hemos representado la esperanza media de vida en hombres y mujeres con dos escalas diferentes. En esta gráfica se pedirá a los alumnos comparar estos dos gráficos e indicar si parecen o no adecuados para representar la diferencia entre la esperanza media de vida de mujeres y hombres. Uno de los dos gráficos ha sido obtenido directamente del ordenador, mientras que el otro ha sido manipulado. Los alumnos deben averiguar cuál ha sido manipulado.

Figura 3. Esperanza de vida media en hombres y mujeres



Actividad 4. La tasa de natalidad. En esta actividad estudiaremos la distribución de las tasas de natalidad. Podemos plantear preguntas tales como: ¿Por qué en este caso conviene agrupar en intervalos? ¿Qué representa la frecuencia dentro de un intervalos? ¿Cuántos intervalos conviene usar en la tabla de frecuencias? ¿Cómo representaría gráficamente estos datos? ¿Es simétrica la distribución? ¿Cómo cambia la forma al variar el número de intervalos? ¿Y si usamos intervalos de distinta amplitud? ¿Qué representa y cómo representaría la frecuencia acumulada? ¿Qué posición ocupa mi país respecto a la tasa de natalidad? ¿Hay algunos países atípicos respecto a la tasa de natalidad? Para contestar esta y otras preguntas similares podemos usar el gráfico del tallo y hojas (Figura 4).

Figura 4: Gráfico de tallo y hojas: Tasa de natalidad

```

0|99
1|001111222223333333444444
1|556778
2|011222334
2|677888899
3|00111122234
3|5568899
4|011122224444
4|55566677788888
5|0012
    
```

Actividad 5. Diferencias en la tasa de natalidad Una vez estudiada globalmente la tasa de natalidad podemos dividir a la clase en grupos para estudiar si la tasa de natalidad es la misma en los diferentes grupos de países. Los alumnos obtendrían las gráficas presentada en la Figura 5. Se puede investigar qué grupos de países tienen valores atípicos para la tasa de natalidad y qué países tienen una tasa de natalidad atípica respecto a su grupo así como analizar por qué al realizar un gráfico global en todo el fichero no aparecen valores atípicos y sí aparecen dentro de los grupos. Los alumnos pueden listar todas las diferencias observadas en los diversos grupos de países.

Figura 5. Distribución de la tasa de natalidad en los diferentes grupos de países

Gráfico del tallo y hojas: Países del Grupo 1

1|1
 1|22333
 1|445
 1|7
 HI|24.7

Gráfico del tallo y hojas: Países del Grupo 2

1|8
 2|03
 2|77889
 3|224
 HI|46.6

Gráfico del tallo y hojas: Países del Grupo 3

0|99
 1|1111
 1|222
 1|3333
 1|4444
 1|5
 1|6

Gráfico del tallo y hojas: Países del Grupo 4

2|22
 2|689
 3|1
 3|8
 4|222
 4|5

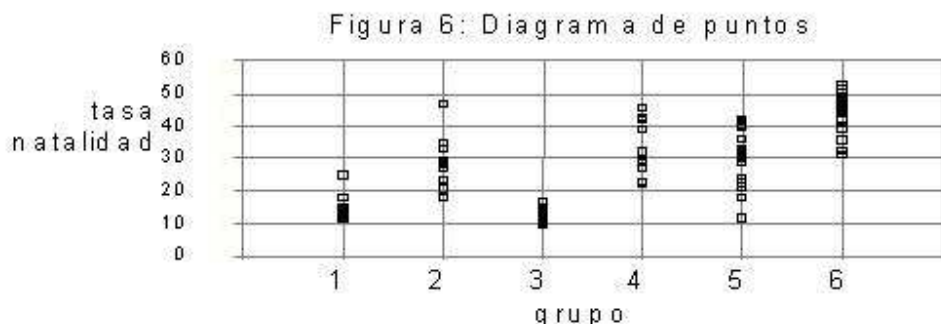
Gráfico del tallo y hojas: Países del Grupo 5

1|1
 1|7
 2|1123
 2|8
 3|00113
 3|69
 4|012

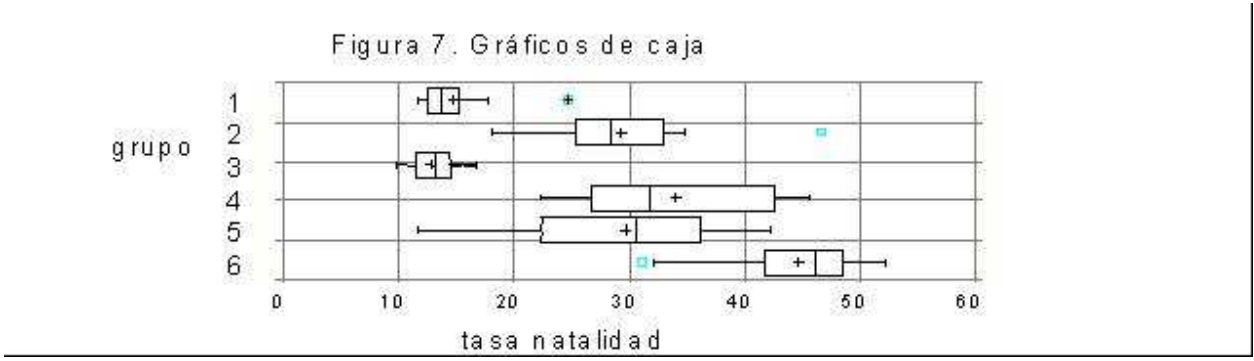
Gráfico del tallo y hojas: Países del Grupo 6

LO|31.1
 3|
 3|
 3|2
 3|55
 3|
 3|89
 4|1
 4|
 4|444455
 4|66777
 4|88888
 5|001
 5|2

El gráfico de tallo y hojas ha sido útil para analizar las diferencias en la tasa de natalidad, pero también podemos usar otras representaciones, como diagramas de puntos y gráficos de caja. Divididos en grupos, los alumnos pueden tratar de hallar representaciones alternativas que pongan de manifiesto las diferencias en la tasa de natalidad. Por ejemplo, podemos observar las Figuras 6 y 7 y comparar con los gráficos del tronco de la Figura 5, señalando las ventajas que tiene cada una de las representaciones gráficas y cómo podríamos cambiar cada gráfica para resaltar más (menos)



las diferencias.



Actividad 6. Esperanza de vida en hombres y mujeres.: En la actividad anterior hemos comparado una variable en distintos subconjuntos de países, pero a veces tiene también sentido comparar dos variables diferentes en el total de los datos. Se dice, por ejemplo, que las mujeres forman el sexo débil. Sin embargo muchas chicas estarían en contra de esta opinión cuando comparen en algunos países la esperanza de vida al nacer en hombres y mujeres. Podríamos investigar si la distribución de la esperanza de vida en el total de países es igual en hombres y mujeres analizando para ello los gráficos de las figuras 8, 9 y 10 y contestar preguntas como las siguientes: ¿Hay mayor variabilidad entre países en la esperanza de vida en hombres o en mujeres? ¿En qué porcentaje de países la esperanza de vida de hombres (mujeres) es mayor de 58 años? ¿y de 68 años? ¿Cuál es el valor de la esperanza de vida de modo que el 70 por ciento de países tiene una esperanza de vida mayor? ¿Es igual en hombres y mujeres? ¿Cómo podrías disimular la diferencia de esperanzas de vida en hombres y mujeres en cada gráfico?

Figura 8: Gráficos de caja. Distribución de la esperanza de vida en hombres y mujeres

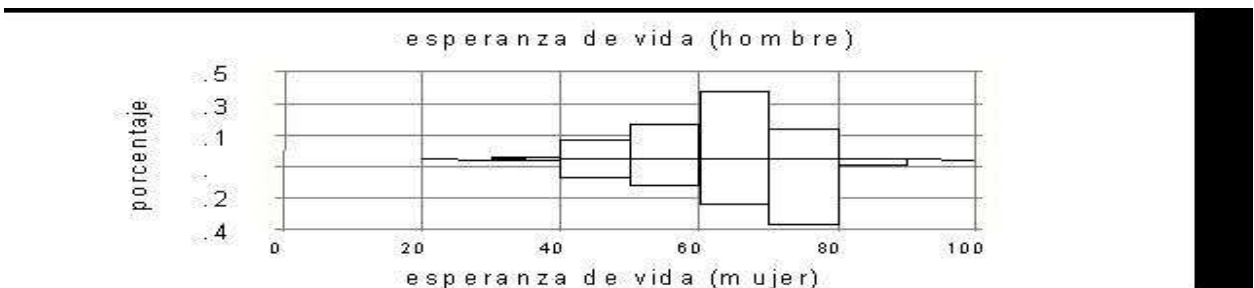
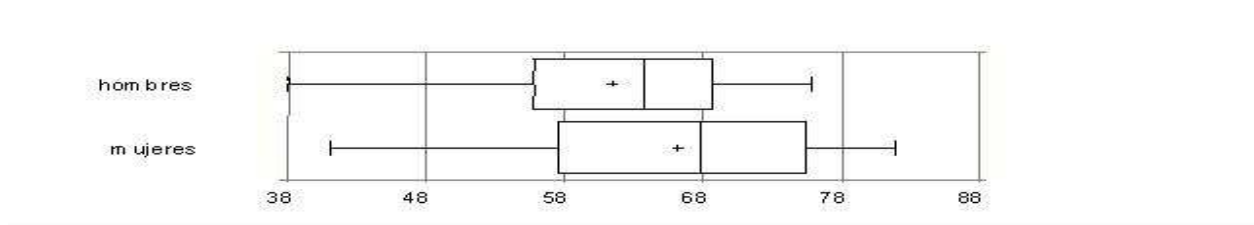
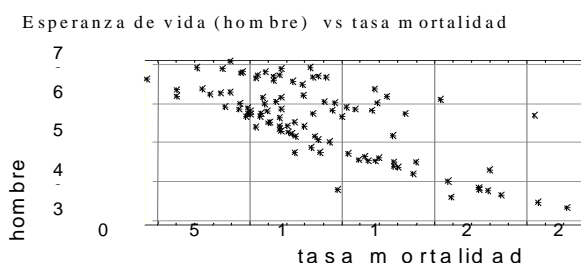
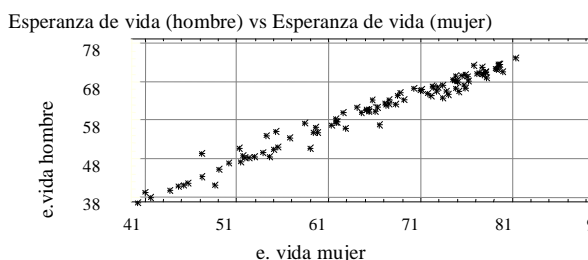
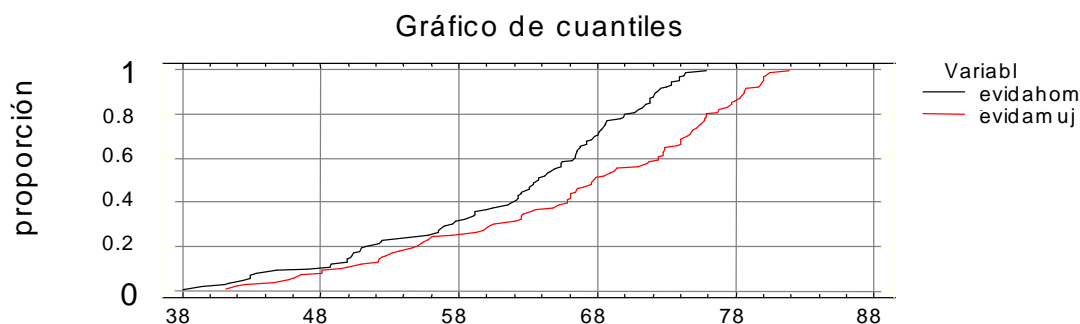


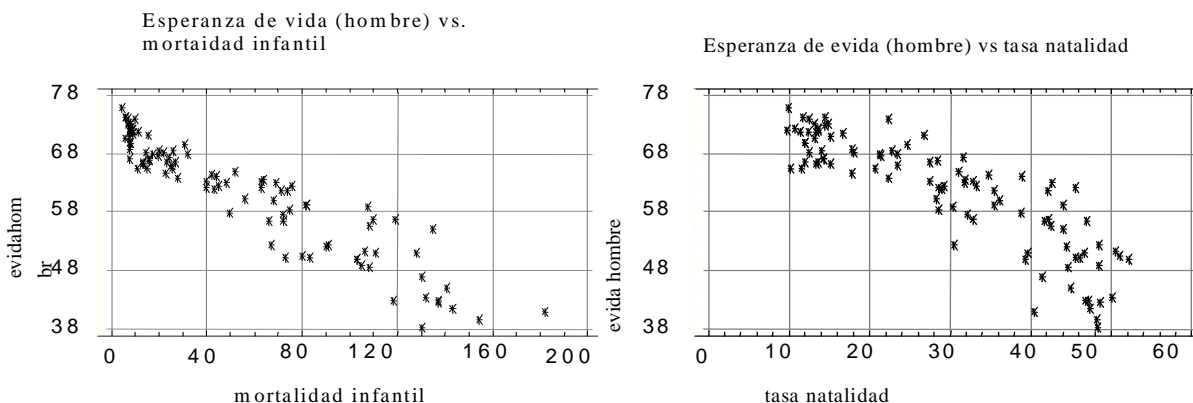
Figura 9: Histogramas. Distribución de la esperanza de vida en hombres y mujeres

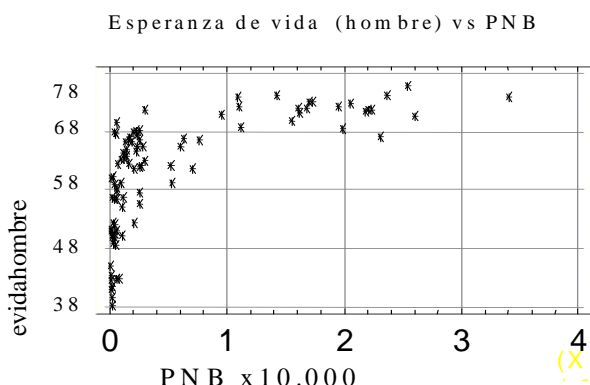
Figura 10: Gráfico de cuantiles. Distribución acumulativa de la esperanza de vida en hombres y mujeres



Actividad 7. Relación de la esperanza de vida con otras variables. En la Figura 11 hemos representado los diagramas de dispersión de la esperanza de vida del hombre en cada país, en función de diversas variables. Podemos proponer a los alumnos las siguientes preguntas: ¿Cuál de las variables influye en la esperanza de vida del hombre? ¿Cuáles son influidas por la esperanza de vida? ¿Cuáles hacen crecer / disminuir la esperanza de vida? ¿Cuál de ella sirve mejor para predecir la esperanza de vida? ¿En qué casos la relación es directa/ indirecta? ¿En qué casos la relación es causal? ¿Es debida a otras variables? ¿Podríamos en alguno de los casos hallar una función matemática para predecir, aproximadamente la esperanza de vida del hombre a partir de la otra variable? ¿Qué tipo de función?

Figura 11. Relación de la esperanza de vida del hombre con otras variables





Algunas reflexiones didácticas

En una enseñanza tradicional nos centramos en la resolución de problemas acotados con unos datos proporcionados por el profesor. En el enfoque exploratorio de la enseñanza de la estadística los alumnos pueden investigar y obtener sus propios datos o completar los dados por el profesor. Los ficheros contenidos en Internet, los anuarios estadísticos, la prensa, o los datos recogidos por los alumnos (mediante medición, encuesta u observación) pueden ser la base para plantear problemas que motiven a los alumnos y les haga interesarse por el aprendizaje. Los ordenadores actuales brindan muchas posibilidades para la enseñanza y el aprendizaje de la estadística (Godino, 1995; Biehler, 1997), desde paquetes profesionales o paquetes didácticos a las hojas electrónicas que se suministran habitualmente como parte del software de uso general. El análisis exploratorio de datos puede también implementarse con las calculadoras gráficas o dividiendo la clase en equipos para repartir el trabajo de cálculo y la realización de los gráficos.

Un punto muy importante es la discusión de las variables y el problema de la medición. Es preciso concienciar a los alumnos de la dificultad que reviste el proceso de categorización o de medición, porque la realidad es siempre más compleja que nuestros métodos para estudiarla. La toma de conciencia sobre la complejidad del proceso de elaboración de las estadísticas demográficas o económicas es un paso importante para valorar el trabajo del estadístico y fomentar la cooperación en censos y encuestas. El trabajo con un fichero completo, en lugar de centrarse en variables aisladas supone el inicio de una filosofía multivariante donde cada variable cobra su importancia o bien es explicada en función del resto y donde el alumno puede tratar de comprobar sus conjeturas con la incorporación de nuevas variables al estudio.

El primer paso en el análisis es el estudio de cada variable, la tabulación y representación gráfica. Algunos investigadores han analizado los diferentes niveles de comprensión de las gráficas (Curcio, 1989), y las dificultades de los alumnos en la elaboración de las mismas o la selección de un gráfico adecuado, debido a la diferente información que aportan las diversas gráficas estadísticas (Li y Shen, 1992). Deberíamos también fomentar en los alumnos un sentido gráfico que les haga ser críticos frente a los posibles gráficos tendenciosos que con frecuencia encontramos en los medios de comunicación.

En este fichero las unidades estadísticas son agregados, lo que tiene repercusión en la interpretación de los promedios de las variables; por ejemplo la media (simple) de todas las esperanzas de vida al nacer en los distintos países no es igual a la esperanza global de vida, sino que esta tiene que ser calculada como una media ponderada en función del número de hombres / mujeres de cada país. Esto puede servir de reflexión sobre las diferentes unidades estadísticas que pueden usarse en un estudio, la necesidad que a veces tenemos de trabajar con valores aproximados y la forma de combinar estudios parciales para obtener índices globales, así como para reflexionar sobre las propiedades de los promedios y la comprensión de las mismas por parte

de los estudiantes (Tormo, 1995; Batanero, 2000). Esta misma observación debe hacerse en la interpretación de otros estadísticos, como los de dispersión o los coeficientes de correlación.

Junto con las representaciones gráficas tradicionales, como diagramas de barras o histogramas, en análisis exploratorio de datos aparecen nuevas representaciones como el diagrama de tallo y hojas o los gráficos de cajas, cuya potencia exploratoria se acentúa con el paso de una a otra representación, así como la selección de partes del fichero para realizar estudios comparativos, por ejemplo al comparar las variables en los distintos grupos de países. El interés no sólo se centra en las tendencias, sino en la variabilidad, así como el estudio de los valores atípicos. Vemos también como la idea de distribución siempre es relativa a un colectivo; por eso un valor puede ser atípico dentro de un subconjunto de datos y no serlo en el global.

El trabajo con análisis exploratorio de datos refuerza también algunos objetivos sugeridos para la educación matemática en los nuevos currículos de secundaria, como el trabajo con problemas abiertos, el uso de sistemas múltiples de representación, la introducción al trabajo con ordenadores o calculadoras gráficas y la conexión de las matemáticas con otras áreas del currículo (Shaughnessy, Garfield y Greer, 1997). Pero el razonamiento estadístico va más allá del conocimiento matemático y de la comprensión de los conceptos y procedimientos. La modelización, la valoración de la bondad del ajuste de los modelos a la realidad, la formulación de cuestiones, la interpretación y síntesis de los resultados, la elaboración de informes son también componentes esenciales de las capacidades que queremos desarrollar en nuestros alumnos.

Al analizar las relaciones entre dos variables numéricas los alumnos deben de extender la idea de dependencia funcional a dependencia aleatoria y diferenciar sus tipos (lineal o no; directa e inversa) así como graduar al menos intuitivamente la intensidad de la relación. Es importante también diferenciar correlación y causalidad y analizar los distintos tipos de relaciones que pueden llevar a la existencia de correlación: dependencia causal; interdependencia, dependencia indirecta, concordancia y correlación espúrea (Estepa, 1995). Una vez detectada la correlación, el interés se centra en la búsqueda de modelos que puedan predecir las variables explicadas en función de las variables explicativas, lo que de nuevo conecta con otro contenido del currículo de matemáticas: las funciones.

Finalmente es preciso reconocer que la enseñanza de la estadística es una tarea delicada, debido a la dificultad que revisten los conceptos estadísticos incluso los más sencillos (Batanero y cols., 1994).

Referencias:

- Batanero, C. (2000). Significado y comprensión de las medidas de posición central. Ponencia presentada en el "*Encontro sobre Ensino e Aprendizagem da Estatística*". Faculdade de Ciências de Lisboa, 3-4 Fevereiro 200. [Recuperable en <http://www.ugr.es/local/batanero>]
- Estepa, A. y Batanero, C. (1996). Judgments of correlation in scatter plots: students' intuitive strategies and preconceptions. *Hiroshima Journal of Mathematics Education*, 4, 21-41.
- Batanero, C., Estepa, A. y Godino, J. D. (1992). Análisis exploratorio de datos: Sus posibilidades en la enseñanza secundaria. *Suma*, 9, 25-31.
- Batanero, C., Godino, J. D., Vallecillos, A., Green, D. R. y Holmes, P. (1994). Errors and difficulties in understanding statistical concepts. *International Journal of Mathematics Education in Science and Technology*, 25(4), 527-547.
- Biehler, R. (1997). Software for learning and for doing statistics. *International Statistical Review*, 65(2), 167-190.
- Day, A. (ed.) (1992), *The Annual Register 1992*, 234, London:Longmans.

- Curcio, F. R. (1989). *Developing graph comprehension*. Reston, VA: N.C.T.M.
- Estepa, A. (1995). Las tablas de contingencia y su enseñanza. ¿Qué podemos aprender de las investigaciones realizadas? *UNO*, 3, 89-100.
- Estepa, A., Batanero, C. y Sánchez, F. T. (1999). Students' intuitive strategies in judging association when comparing two samples. *Hiroshima Journal of Mathematics Education*, 7, 17-30.
- Godino, J. D. (1995). ¿Qué aportan los ordenadores a la enseñanza y aprendizaje de la estadística? *UNO*, 5, 45-55.
- Li, K. Y. y Shen, S. M. (1992). Students' weaknesses in statistical projects. *Teaching Statistics*, 14 (1), 2-8.
- Tormo, C. (1995). Dificultades del alumnado respecto a la media aritmética. *UNO*, 5, 29-36.
- Rouncenfield (1995). The statistics of poverty and inequality. *Journal of Statistics Education*, 3(2).
- Tukey, J. W. (1962). The future of data analysis. *Annals of Mathematical Statistics*, 33, 1-67.
- Tukey, J. W. (1970). *Exploratory data analysis*. New York: Addison Wesley.
- U.N.E.S.C.O. *1990 Demographic Year Book*. New York: United Nations.
- Vallecillos, A., y Batanero, C. (1997). Conceptos activados en el contraste de hipótesis estadísticas y su comprensión por estudiantes universitarios. *Recherches en Didactique des Mathématiques*, 17(1), 29-48.
- Shaughnessy, J. M., Garfield, J., & Greer, B. (1996). Data handling. En A. Bishop et al. (Eds.), *International handbook of mathematics education* (v.1, pp. 205-237). Dordrecht: Kluwer, A. P.