

ESTRATEGIAS Y ARGUMENTOS EN EL ESTUDIO DESCRIPTIVO DE LA ASOCIACIÓN USANDO MICROORDENADORES^(*)

BATANERO BERNABEU, M. C., ESTEPA CASTRO, A. y DÍAZ GODINO, J.
Departamento de Didáctica de la Matemática. Universidad de Granada.

(*) Este trabajo forma parte del Proyecto PS88-0104 subvencionado por la Dirección General de Investigación Científica y Técnica (M.E.C. Madrid).

SUMMARY

In this work, problem solving procedures about statistical variables association using microcomputers are analysed. The effect of several didactical variables on these procedures is also studied. The studied population consisted of 18 prospective teachers who had been previously trained for 7 weeks. By studying the students' replies we determined their ideas of the scope and meaning of the term «association» and derived criteria for construction of new didactic situations.

INTRODUCCIÓN

El uso de ordenadores como recurso didáctico en los procesos de enseñanza-aprendizaje de las matemáticas plantea problemas de investigación de reconocido interés e importancia, como se pone de manifiesto en Howson y Kahane (1986). Para el caso de la Estadística, el hecho de que los alumnos dispongan de la herramienta informática para el proceso de datos hace posible ponerles ante situaciones-problemas relativos al análisis de ficheros de datos de aplicaciones reales. De este modo, los programas de cálculo y representación gráfica pueden actuar como amplificadores conceptuales (Lesh 1987) de las nociones y procedimientos estadísticos al permitir el planteamiento de cuestiones más abiertas e inducirá una actividad matemática exploratoria sobre los conjuntos de datos.

En este contexto teórico situamos este trabajo, que analiza los procedimientos empleados por los estudiantes en la resolución de problemas sobre el estudio descriptivo de la asociación entre variables estadísticas incorporadas en ficheros de datos de aplicaciones reales, y el efecto sobre la elección de dichos procedimientos de distintas variables didácticas. Este análisis, junto con el de los argumentos expuestos para apoyar las respuestas

a las preguntas que se les plantea, permite conocer el alcance y sentido dado por estos alumnos al concepto de asociación y deducir criterios para la construcción de nuevas situaciones didácticas.

Desde el punto de vista metodológico usamos la técnica del registro de la interacción alumno-ordenador, lo que nos permite identificar los procedimientos empleados en la resolución de las cuestiones. Sin embargo, se muestra la necesidad de recurrir a las argumentaciones escritas en que basan sus respuestas los alumnos para poder apreciar la corrección y pertinencia de dichos procedimientos y para controlar el sentido de los conocimientos estadísticos adquiridos.

METODOLOGÍA

Experiencia de enseñanza y logicial utilizado

Este trabajo es parte de una investigación sobre enseñanza de la Estadística Descriptiva que se ha desarrollado con un grupo de 18 futuros profesores (estudiantes de

edades entre 19 y 20 años) durante un período de 7 semanas. En este tiempo se han alternado dos horas semanales de exposición teórica y de resolución de ejercicios con papel y lápiz, con otra de práctica en el aula de informática. Los contenidos abarcados en las 21 horas de enseñanza han sido los de estadística descriptiva uni y bivalente y una introducción a la filosofía del análisis exploratorio de datos y a alguno de sus gráficos.

Los programas empleados en la experiencia forman parte del paquete PRODEST que se describe en Batanero y col. (1987). Cada uno de los programas puede analizar el fichero de datos completo, para las variables elegidas por el usuario, o bien seleccionar una parte del mismo, que queda definida especificando el rango de variación de otra variable, la cual actúa como criterio de selección. Los ficheros de datos han sido obtenidos por los alumnos mediante una encuesta a sus compañeros, o bien han sido suministrados por el profesor.

Los programas empleados, además de los de grabación y depuración de ficheros, son los siguientes:

Práctica 1 (9 actividades): programa CONTAJE

Realización de tablas de frecuencias de variables estadísticas cualitativas o discretas, proporcionando la frecuencia absoluta, relativa, acumulada y porcentaje de cada valor de la variable. Realización de diagramas de barras, diagramas acumulativos y gráficos de sectores.

Práctica 2 (9 actividades): programa HISTO

Cálculo de tablas de frecuencias de variables estadísticas continuas, agrupando en intervalos, de los cuales se puede elegir el extremo inferior y la amplitud. Proporciona los extremos y marcas de clase de los intervalos, frecuencias absolutas, relativas, acumuladas y porcentajes, así como los correspondientes histogramas, polígonos de frecuencias y polígonos acumulativos.

Práctica 3 (7 actividades): programa TRONCO

Realización del gráfico del tronco (stem and leaf).

Práctica 4 (6 actividades): programa MEDIANA:

Cálculo de los estadísticos: máximo, mínimo, recorrido, mediana, cuartiles, recorrido intercuartílico. Dibujo del gráfico de la caja y la curva empírica de distribución. Cálculo de percentiles y rango de percentiles por petición del usuario, en la forma en que se hace habitualmente para datos no agrupados.

Práctica 5 (7 actividades): programa ESTADIS

Cálculo de los estadísticos siguientes: máximo, mínimo, recorrido, media, varianza, desviación típica, coeficiente de variación, coeficiente de asimetría y coeficiente de curtosis.

Práctica 6 (9 actividades): programa TABLAS

Preparación de tablas de contingencia incluyendo porcentajes respecto filas, columnas y total y cálculo del estadístico Chi-cuadrado. Posibilidad de agrupación en filas y columnas.

Práctica 7 (10 actividades): programa REGRESIO

Cálculo de valores mínimos, máximos, medias y varianzas y del coeficiente de correlación lineal entre dos variables, la ecuación de la recta de regresión y la representación gráfica del correspondiente diagrama de dispersión y de la recta de regresión ajustada.

Tipos de actividades prácticas realizadas

En cada una de las prácticas, las actividades realizadas han sido la determinación de la respuesta a una serie de preguntas referidas a una situación problemática de análisis de un conjunto de datos. La resolución de las cuestiones implica el empleo de uno o varios de los programas de cálculo estadístico que se han descrito anteriormente, que el alumno puede elegir a voluntad, debiendo también decidir los distintos parámetros que hay que emplear, entre los posibles, para cada uno de estos programas. Podemos clasificar los problemas propuestos en estas actividades prácticas en los tipos siguientes:

- a) Análisis de una única variable en el fichero completo.
- b) Análisis de una variable en una parte del fichero
- c) Estudio conjunto de dos variables
- d) Comparación de muestras independientes.
- e) Comparación de muestras relacionadas.
- f) Entre varias soluciones posibles a un problema dado elegir la más adecuada, por ejemplo, elegir la medida de posición central, en función de la distribución de la variable.

g) Comprobar experimentalmente algunas propiedades matemáticas de las distribuciones y sus estadísticos, como la convergencia del polígono de frecuencias a la gráfica de la función de densidad.

h) Sobre un conjunto dado de datos, plantear cuestiones que puedan resolverse mediante el empleo de un procedimiento estadístico.

De esta clasificación puede observarse que en la enseñanza de la estadística basada en el uso del ordenador se refuerza bastante el estudio de la asociación entre variables, ya que los problemas clasificados en los apartados c), d) y e) se refieren a este concepto estadístico. Aunque las preguntas del tipo c) no podrían plantearse hasta la quinta clase práctica, las otras dos categorías de problemas pueden ser utilizadas desde la primera, ya que incluso al estudiar las tablas de frecuencias y gráficos de variables discretas resulta de gran interés para los alumnos la comparación de variables y muestras.

Además de esta clasificación, dentro de los apartados a) hasta el g), los problemas planteados difieren entre sí por el contenido estadístico presente en el mismo, si éste es implícito o explícito, y la apertura del problema. Aunque, como se ha dicho, muchas de las cuestiones planteadas son abiertas, en el sentido de que pueden ser contestadas mediante el empleo de diversos procedimientos estadísticos, esta apertura crece a medida que aumenta el número de clases prácticas realizadas, al aprender el alumno en cada una de estas clases el empleo de un nuevo programa de cálculo que aumenta sus posibilidades de solución.

Por último, es esencial el tipo de variable estadística incluida en el problema: dicotómica, cualitativa o cuantitativa, así como ciertas características de su distribución: necesidad de agrupamiento, ser o no simétrica, poseer valores atípicos en caso de asimetría y bimodalidad, ya que estos aspectos condicionan el análisis estadístico que debe efectuarse.

Evaluación final del aprendizaje

El proceso de aprendizaje se ha observado mediante la grabación de las interacciones alumno-ordenador en todas las actividades. Además se ha realizado una evaluación final de cada alumno, también grabada, que consistió en la respuesta a una serie de cuestiones, usando el ordenador, sobre el análisis de un fichero relativo a datos deportivos reales relativos a 60 alumnos de octavo de EGB de un centro de Educación General Básica.

Las variables estadísticas incluidas en este fichero fueron las siguientes: sexo del alumno, si habitualmente practica o no algún deporte fuera del horario escolar, el número de pulsaciones por minuto en reposo y después de efectuar 30 flexiones, el tiempo en segundos que tardó en recorrer 30 metros en una prueba efectuada en Septiembre y en otra efectuada en Diciembre, el peso en kilos, la talla en centímetros y la edad en años. Junto con las preguntas del examen se proporcionó a los estudiantes una descripción del fichero, los datos impresos y una hoja de codificación del mismo, con objeto de que pudiesen apreciar el tipo de variables incluidas en el conjunto de datos, su rango de valores, la necesidad o no de agrupación, presencia de valores atípicos, etc.

En dicha evaluación se planteó a los alumnos preguntas de los diferentes tipos a los que hemos aludido en la clasificación efectuada de las prácticas y referidas a los diversos contenidos estadísticos. El análisis que efectuamos a continuación se refiere únicamente a las cuestiones sobre asociación, que fueron las siguientes:

1. ¿Hay diferencia en la práctica de deporte entre chicos y chicas?
2. ¿Hay relación entre las variables «número de pulsaciones en reposo» y «practica algún deporte»?
3. ¿Hay relación entre las variables «pulsaciones en reposo» y «pulsaciones después de 30 flexiones»? ¿De qué tipo?

4. ¿Ha disminuido el tiempo que tardan los alumnos en recorrer 30 metros de Septiembre a Diciembre?

Hay que hacer notar que cada una de estas preguntas puede ser contestada utilizando varios de los programas de cálculo disponibles. Por ejemplo, el primero trata de la posible asociación entre el sexo y la práctica de deporte. Las dos variables son cualitativas dicotómicas y, por tanto, podría resolverse de las siguientes maneras:

- Con el programa **CONTAJE**: Obteniendo las tablas de frecuencias separando chicos de chicas y comparando después las frecuencias relativas, los gráficos de barras o sectores o la moda de las distribuciones. Ello supondría estudiar separadamente la variable «práctica de deporte» en dos muestras independientes, empleando para ello la posibilidad de selección de casos del paquete estadístico.

- Con el programa **TABLAS**: Obteniendo la tabla de contingencia y comparando las distribuciones condicionales, con lo que se efectuaría un estudio conjunto de las dos variables.

- Con el programa **REGRESIO**: Las variables son cualitativas, pero al ser dicotómicas y tomar cada una sólo dos valores (1,2) se puede utilizar el coeficiente de correlación para ver la asociación que existe entre ellas. Por el mismo motivo anterior se podría utilizar el programa **ESTADIS** y comparar las medias.

En la Tabla 1 se muestra para cada uno de estos problemas las opciones del paquete que se pueden utilizar en su solución. Con esta tabla puede observarse la apertura de los problemas y compararse estas posibles soluciones con las seguidas efectivamente por los alumnos; ello permite, además, determinar las variables didácticas presentes en estos problemas.

Tabla 1
Opciones del paquete utilizables en la resolución de los ejercicios sobre asociación.

Opciones	Ejercicio núm. :			
	1	2	3	4
CONTAJE	X			X
HISTO		X		X
TRONCO		X		X
MEDIANA		X		X
ESTADIS	X	X		X
TABLAS	X	X	X	X
REGRESIO	X	X	X	X

Análisis de los datos

Además de los registros citados de interacción alumno-ordenador se recogieron las respuestas y argumentaciones escritas para las distintas cuestiones en fichas de trabajo entregadas a los alumnos.

Con objeto de obtener una codificación de la información recogida se ha aplicado a estos instrumentos de recogida de datos un análisis de contenido (Weber 1985). Esta técnica está basada en la definición de unidades de análisis sobre los datos observacionales y en la identificación de variables y categorías sobre dichas unidades, para luego proceder al análisis estadístico de las variables obtenidas de este modo. En nuestro caso, al tratar de responder a las cuestiones planteadas en la evaluación, el alumno utiliza uno de los programas disponibles y puede ocurrir que en la primera ejecución resuelva el problema o que se produzca un error de ejecución o de otro tipo, y deba intentarlo de nuevo.

De este modo, se ha tomado el *intento* de solución como unidad de análisis, obteniéndose en total 120 unidades de análisis para las 4 cuestiones propuestas y los 18 alumnos, lo que proporciona un número medio de 1,66 intentos por alumno y pregunta. Hay que tener en cuenta también que, como veremos, algunos de los problemas son resueltos por varios alumnos por más de un método, de lo que se deduce que han conseguido un manejo suficiente de los programas.

Para cada una de las unidades de análisis estudiaremos las siguientes variables: 1) programa empleado en la resolución del problema y 2) argumento en que el alumno basa su respuesta.

RESULTADOS Y DISCUSIÓN

Procedimientos de análisis de datos

En la Tabla 2 se presenta la clasificación de los 120 intentos por problema y programa utilizado. Un primer hecho que puede observarse es que el número total de veces que se usan el conjunto de programas para una misma cuestión es, en general, mayor que el número de alumnos (18). Aparte de los errores de ejecución que han supuesto el 6,1 por ciento (no incluidos en la tabla 2), en algunos casos el alumno debe elegir una solución entre otras

posibles (por ejemplo, la amplitud del intervalo), por lo que lógicamente ha de hacer varios ensayos con un mismo programa. En otros casos no es así, pero el alumno confirma la solución del problema, una vez hallada, por otro de los métodos posibles, por lo que emplea más de un programa para dar su respuesta.

Observamos que el procedimiento empleado para estudiar la asociación depende no sólo del tipo de variable analizada. Así, los ejercicios 3 y 4, referidos a la diferencia entre muestras relacionadas son resueltos en forma distinta. En el primer caso, plantear la pregunta en términos de *relación* entre las dos variables induce al empleo casi exclusivo de la regresión. En el segundo, la palabra *diferencia* provoca el estudio separado de cada variable por diversos métodos, principalmente por medio de las tablas de frecuencias (HISTO) o de los estadísticos de orden (MEDIANA).

Éste no es, sin embargo, el único factor observado. Dentro de una pregunta planteada en los mismos términos, es esencial el tipo de variable estadística presente en la tarea. Por ejemplo, en el problema 1, la pregunta se plantea en términos de *diferencias*, pero, en este caso, todos los alumnos lo resuelven estudiando separadamente las tablas de frecuencias de las variables (CONTAJE), aunque un número importante de ellos, además, confirma sus resultados empleando el estudio conjunto (TABLAS). Asimismo, la pregunta 2, que se plantea en términos de *relación*, es resuelta, en primer lugar, con el de la regresión y después mediante la tabla de contingencia.

Argumentaciones correctas

En los párrafos anteriores hemos analizado los procedimientos de resolución seguidos por los alumnos. Este análisis no es, sin embargo, suficiente para asegurar la corrección de las respuestas: por un lado, hemos visto que cada pregunta podría resolverse con varios procedimientos, pero lo esencial es la información utilizada de los mismos y la interpretación dada a esta información. Por este motivo, se han analizado las *argumentaciones* ofrecidas por los estudiantes, que se relacionan en la tabla 3. Los tipos de argumentos hallados en las preguntas de asociación se han clasificado en la forma siguiente:

Frecuencias absolutas: Razonamientos basados en la frecuencia absoluta de las distribuciones univariantes o de las condicionales o marginales de una tabla de contingencia.

Frecuencias relativas: Proporciones o porcentajes en distribuciones univariantes o en las condicionales o marginales de una tabla de contingencia.

Percentiles o rangos de percentiles: Frecuencias acumuladas ascendentes o descendentes y argumentos basados en los estadísticos de orden, a excepción de la mediana.

Valores centrales: Comparación de medias, modas o medianas de las distribuciones.

Tabla 2
Frecuencia de intentos según cuestión y programa usado.

Procedimiento	Cuestión				Total
	1	2	3	4	
CONTAJE	19			2	21
HISTO	2			7	9
TRONCO				1	1
MEDIANA				6	6
ESTADIS				4	4
TABLAS	9	23	2	3	37
REGRESIO		20	20	2	42
Total	30	43	22	25	120

Coefficiente de correlación: (Valor absoluto y signo).

Nube de puntos: Forma creciente o decreciente de la nube, forma aproximada de línea recta, mayor o menor dispersión de los puntos de la nube.

Otros argumentos: En el estudio general se han tenido, además, en cuenta, los siguientes tipos de razonamientos, que en el caso de la asociación de variables apenas han sido empleados por los alumnos: *Gráficos* (forma, presencia de varias modas, diferencias entre gráficos, etc.), *medidas de dispersión, medidas de forma*. También se ha considerado la realización de cálculos adicionales, ya que algunos alumnos han encontrado insuficiente, en algunos casos, la información proporcionada por los programas.

El número de argumentos es mayor que el número de estudiantes, porque a veces utilizan dos o más para el mismo problema. En las dos primeras preguntas en las que interviene una variable dicotómica, los argumentos más utilizados correctamente son los relacionados con frecuencias: relativas, absolutas y rangos de percentiles, a pesar de que en la segunda cuestión casi todos los alumnos han empleado el programa REGRESIO, además de la tabla de contingencia.

Independientemente de que la pregunta haya inducido a estudiar las variables en forma conjunta (segunda cuestión) o no (primera), parece deducirse de la argumentación que la asociación de una variable con otra dicotómica se interpreta preferentemente como diferencia en las distribuciones de frecuencias condicionales. Tam-

bién concluimos que el uso del procedimiento REGRESIO en la cuestión segunda se ha utilizado a modo de confirmación visual de la respuesta, ya que sólo unos pocos alumnos han empleado un argumento sobre el valor del coeficiente de correlación.

La tendencia a basar sus argumentos en valores numéricos se presenta de nuevo en la tercera cuestión, en la que las razones basadas en el coeficiente de correlación superan a las basadas en la *forma de la nube de puntos*. Ambos tipos de argumentos ponen de relieve que la idea de relación entre variables cuantitativas es sinónima de la de dependencia, aunque aleatoria.

Al analizar los argumentos empleados en la última cuestión, observamos de nuevo la influencia de la forma en que está hecha la pregunta, ya que pocos alumnos identifican la asociación existente entre las dos variables cuantitativas. Muy pocos argumentos se refieren a ello. La mayor parte de argumentos correctos se refieren a los valores centrales. Aunque la mitad de los alumnos han empleado el programa HISTO, pocos han comparado las frecuencias relativas de las dos distribuciones. Por ello pensamos que no han llegado a identificar el problema como uno de estudio de asociación de una variable cuantitativa respecto a una dicotómica, sino más bien como de diferencias de dos variables cualesquiera.

Argumentaciones incorrectas

Para finalizar el análisis comentaremos los razonamientos erróneos expuestos por los alumnos. Si bien su número es muy escaso, consideramos de interés este comentario para poner de relieve los puntos en los que los alumnos encuentran mayor dificultad.

La principal fuente de error en la tabla de contingencia ha sido la interpretación de las distribuciones de frecuencias condicionales, al suministrar el programa TABLAS tanto los porcentajes respecto a filas como respecto a columnas, como podemos observar en la siguiente respuesta a la primera cuestión, en la que el alumno ha obtenido la tabla de contingencia mostrada en la figura 1:

Tabla 3

Frecuencias de argumentos en las cuestiones sobre asociación.

Cuestión	Argumentos correctos	Argumentos incorrectos	Total
1	Frec. relativa	15	29
	Frec. absoluta	7	
	Otros argumentos	5	
	Total cuestión 1:	27	
2	Frec. relativa	7	29
	Percentil/rango	6	
	Frec. absoluta	4	
	Dispersión	3	
	Correlación	3	
	Otros argumentos	4	
	Total cuestión 2:	27	
3	Correlación	13	27
	Nube de puntos	6	
	Otras	3	
	Total cuestión 3:	22	
4	Valor central	11	21
	Otras	7	
	Total cuestión 4:	18	

Figura 1

Tabla de contingencia de «Práctica de deporte por sexo».

	SEXO		TOTAL
	H	V	
PRACTICA	15	13	28 (58,3)
DEPORTE	53,6	46,4	
	75,0	46,4	
NO PRACTICA	5	15	20 (41,6)
DEPORTE	25,0	75,0	
	25,0	53,6	
TOTAL	20 (41,7)	28 (58,3)	48 (100,0)

Si nos vamos a la tabla de contingencia, vemos que, en el total, un 53,6 % de las chicas practican deporte y un 46,4% de los chicos y, sin embargo, hay un 25% de las chicas que no lo practican.

Como podemos ver, no sólo se trata de un error de identificación de porcentajes, ya que no es consciente, al menos en ese momento, de que la suma de porcentajes ha de ser igual a 100.

Al interpretar la nube de puntos algunos alumnos piensan que la relación debe ser de tipo aditivo, debiendo aumentar aproximadamente lo mismo en todos los casos, no siendo conscientes del factor de proporcionalidad marcado por la pendiente de la recta:

No existe relación: a pocas pulsaciones antes, pocas pulsaciones después...

También la existencia de otras variables parece que excluye la idea de dependencia:

Las pulsaciones en reposo y después de 30 flexiones no guardan entre ellas ninguna relación, dependiendo de si el individuo practica o no deporte...

Alguno piensa erróneamente que la dependencia debe inducir un cambio de las frecuencias relativas de casos por encima y debajo de la media:

Vemos también que el número de personas que tienen muchas pulsaciones después de las 30 flexiones es aproximadamente igual al que tiene pocas pulsaciones después de las 30 flexiones...

En otros casos, los alumnos encuentran respuestas contradictorias con los diferentes procedimientos, no siendo conscientes de que se debe llegar a una única conclusión:

Mirando la tabla de contingencia vemos que los que practican deporte tienen mayor número de pulsaciones que los que no y viendo el coeficiente de correlación, que es casi cero, encontramos que no hay relación.

Por último, se observa la presencia de sistemas de creencias arraigadas y la no identificación del problema como de índole estadística, hecho descrito por autores como Kahnemann y col. (1982), en argumentos como los siguientes:

No existe, ya que no son variables relacionables, aunque el número de pulsaciones es inferior en los que practican deporte. No, en realidad la rentabilidad de los ejercicios en la clase de educación física no consistirá en que los niños lo hagan sino en cómo lo hagan cuando lo hagan.

El resto de las respuestas erróneas pueden clasificarse en alguno de los casos anteriores.

CONCLUSIONES

El análisis de los resultados ha puesto de manifiesto que un proceso de enseñanza de nociones estadística basado en la resolución de problemas realistas de análisis de datos, hecho posible al disponer del recurso informático, es insuficiente para que los estudiantes adquieran todo el sentido y verdadero alcance de las nociones estadísticas de asociación e independencia. El diseño de situaciones didácticas, basadas en el uso de logicales, precisa tener en cuenta la naturaleza de las distintas variables didácticas, así como atribuir un papel relevante al profesor como gestor de las situaciones, especialmente para explicitar los matices del saber matemático incorporado en los problemas.

REFERENCIAS BIBLIOGRÁFICAS

BATANERO, M. C., DÍAZ GODINO, J. y ESTEPA CASTRO, A., 1987. Un paquete didáctico de programas para el laboratorio de estadística, en I. Brincones y col. (ed.). *Actas del I Simposio Internacional de Educación e Informática*. (ICE de la Universidad Autónoma: Madrid), pp. 380-388.

HOWSON, A. B. y KAHANE, J. P., 1986. *The influence of computers and informatics on mathematics and its teaching*. (Cambridge University Press: New York).

LESH, R., 1987. The evolution of problem representation in the presence of powerful conceptual amplifiers, in C. Janvier (ed). *Problems of representation in the teaching and learning of mathematics*. (Lawrence Erlbaum Assoc.: London).

KAHNEMAN, D., SLOVIC, P. y TVERSKY, A., (eds.) 1982. *Judgements under uncertainty: heuristics and biases*. (Cambridge University Press: New York).

WEBER, R. P., 1985. *Basic Content Analysis*. (Sage University Press: London).