

INTERPRETACIÓN DE LOS DIAGRAMAS DE DISPERSIÓN POR ESTUDIANTES DE BACHILLERATO

ESTEPA CASTRO, ANTONIO

Universidad de Jaén. Facultad de Humanidades y Ciencias de la Educación. Departamento de Didáctica de las Ciencias. Campus de las Lagunillas, Jaén
aestepa@ujaen.es

Resumen. En este trabajo se asume que la correlación es un tema importante en el currículo de bachillerato. Se esboza la importancia de los gráficos en la enseñanza de la estadística, fijando la atención en los diagramas de la dispersión, cuyo estudio suele iniciar el tema de dependencia aleatoria entre variables, correlación y regresión en el bachillerato actual. Una muestra de estudiantes de bachillerato ha cumplimentado un cuestionario sobre diagramas de dispersión, con cuyas respuestas se ha caracterizado el significado y la interpretación que estos estudiantes hacen de estos gráficos. Se finaliza con unas implicaciones para la enseñanza y se apuntan cuestiones para otras investigaciones.

Palabras clave. Interpretación de diagramas de dispersión, bachillerato, gráficos estadísticos, correlación y regresión, educación estadística.

The Interpretation of Scatter Plots by Baccalaureate Students

Summary. In this paper, it is assumed that correlation is an important subject in the Baccalaureate curriculum. The importance of graphs in statistic education is outlined; attention is focus on scatter plot diagrams, the study of which usually initiates the subject of random dependency between variables, correlation and regression in the current baccalaureate curriculum. A sample of baccalaureate students was used along with a questionnaire on scatter plots. With the students' answers, we have characterized the meaning and interpretation that these students have of these graphs. We end with implications for education and questions for other research.

Keywords. Interpretation of scatter plots, baccalaureate, statistical diagrams correlation and regression, statistics education.

1. INTRODUCCIÓN

La correlación es un tema importante en el currículo de bachillerato. Dentro de este tema cabe destacar la utilización de los diagramas de dispersión para representar y analizar la relación de dependencia entre dos variables referidas al mismo conjunto de datos y como medio gráfico para introducir su estudio.

Se puede afirmar que el uso de gráficos para analizar datos en estadística se ha visto incrementado en la segunda mitad del siglo xx. A principio de los sesenta, Tukey (1962) afirma que las técnicas gráficas tendrán grandes posibilidades en el futuro del análisis de datos. Posteriormente, Tukey (1977) expondrá la importancia de los gráficos para el análisis exploratorio de datos, proponiendo una nueva filosofía para dicho análisis: el enfoque exploratorio en el análisis de datos, donde las representaciones

gráficas tienen un papel fundamental, por lo que propone algunas nuevas. Todo esto se ha visto favorecido en las últimas décadas la propagación del uso de ordenadores en la enseñanza de la estadística, dada su capacidad de manejar datos y su facilidad para realizar todo tipo de gráficos.

La importancia del uso de gráficos en el análisis estadístico radica en que algunas veces un conjunto de datos puede ser analizado de manera adecuada por métodos gráficos; otras veces, unas adecuadas técnicas gráficas facilitan significativamente el análisis numérico. Las representaciones gráficas tienen un triple objetivo: registrar y almacenar datos de manera compacta, facilitar la comunicación de información, o bien, analizar el conjunto de datos que representan para obtener conoci-

miento de su estructura (Chambers y cols., 1983). Este triple objetivo tiene gran interés en la transmisión de los conocimientos estadísticos, es decir, desde el punto de vista didáctico.

La lectura e interpretación de gráficos no es una tarea trivial; en general, cuando los estudiantes tienen que interpretar o crear un gráfico, deben comprender, al menos, el significado de los ejes y la relación entre las variables (Curcio y Artzt, 1997).

El diagrama de dispersión es la herramienta gráfica más usada, sencilla y potente para analizar la relación que puede existir entre dos variables. En consecuencia, el interés desde el punto de vista didáctico es bien patente, ya que los resúmenes numéricos de los datos no son suficientes para captar todas las características de la relación existente entre las variables. Ahora bien, la interpretación de los diagramas de dispersión no es tarea fácil, pues aunque solamente un diagrama de dispersión nos da una visión completa de la naturaleza de la relación existente entre las variables, también es verdad que una enorme diversidad de estos gráficos, con diferentes configuraciones de la nube de puntos, tiene el mismo coeficiente de correlación (Chambers y cols., 1983).

Un diagrama de dispersión nos sugiere la manera en que se relacionan las dos variables, es decir, si la relación es lineal, cuadrática o de otro tipo. Esto nos da una buena visión de conjunto de la relación entre las dos variables, y nos ayuda a la interpretación de los coeficientes de correlación o el modelo de regresión. Con el gráfico podemos identificar patrones, tendencias o *clusters* en los datos. También nos muestra los valores extremos, que en algunas ocasiones influyen en demasía en el coeficiente de correlación. Un diagrama de dispersión es un resumen útil de un grupo de datos bivariados (dos variables). Con la debida comprensión del gráfico podemos trazar una línea recta que se aproxime a la recta de regresión verdadera.

Cada punto contribuye al diagrama de dispersión; generalmente, la configuración en el plano de los puntos muestra una tendencia, un patrón. Dicho patrón indica el tipo y la intensidad de la relación entre las dos variables. Algunas veces la mayoría de los puntos tienden a agruparse alrededor de una línea recta; si los puntos caen todos sobre dicha recta, la asociación entre las variables será perfecta (dependencia funcional) y, en consecuencia, la correlación lineal será la más alta. Si la recta a la que tienden a agruparse los puntos tiene pendiente positiva, la relación entre las variables será directa (positiva). Si la recta a la que tienden a agruparse los puntos tiene pendiente negativa, la relación entre las variables será inversa (negativa). Si se observa que los puntos se alinean siguiendo una línea curva, la relación entre las variables será no lineal, y, aunque se puede calcular el coeficiente de correlación lineal, en este caso, no será una buena medida de la asociación entre las variables. Si los puntos no muestran una tendencia a agruparse, entonces la relación será muy baja o nula.

Siempre se debe tener en cuenta que correlación no implica causalidad, es decir, la existencia de una fuerte co-

rrelación no implica necesariamente que una de las variables sea causa de la otra. La relación puede ser debida a otras circunstancias.

Los diagramas de dispersión que utilizaremos en este trabajo son simples, en el sentido de que cada punto representa un solo caso, es decir, una sola frecuencia de cada valor de la variable que lo compone.

2. ANTECEDENTES

Desde el punto de vista didáctico, hay investigaciones donde se utilizan diagramas de dispersión, que a continuación comentaremos en los elementos que tengan relación con el presente trabajo. Estepa y Batanero (1995) proponen a estudiantes del último curso de la educación secundaria (18 años de edad), antes de la instrucción, la evaluación del tipo de correlación existente en diagramas de dispersión; estudian la exactitud del juicio de correlación y las estrategias utilizadas por los estudiantes; una de sus conclusiones es que el tipo de correlación es percibido mejor cuanto más intensa es la correlación. Estepa y Batanero (1996) analizan las estrategias que utilizan los estudiantes para realizar juicios de asociación deduciendo concepciones erróneas sobre la asociación estadística (determinista, localista y causalista) que presentan los estudiantes, antes de la instrucción, sobre la correlación existente en diagramas de dispersión y que influyen en el juicio de correlación realizado.

La descripción de la comprensión, por parte de estudiantes universitarios sobre la correlación, coeficiente de correlación y regresión, ha sido realizada por Truran (1997). La comprensión de la correlación por estudiantes universitarios también ha sido estudiada por Estepa (1994), Morris (1997), Batanero, Estepa y Godino (1997), Batanero y Godino (1998) y Batanero, Godino y Estepa (1998), quienes además estudiaron las concepciones sobre la asociación estadística y su evolución como consecuencia de un experimento de enseñanza.

La traducción o paso de una representación de la correlación a otra que realiza una muestra de estudiantes universitarios es analizada por Sánchez, Estepa y Batanero (2000). Estos autores tienen en cuenta cuatro tipos de representaciones de la correlación: descripción verbal, tabla numérica, diagrama de dispersión y coeficiente de correlación, junto con las variables de tarea en las situaciones presentadas a los estudiantes, esto es, intensidad de la correlación, tipo de covariación, tipo de ajuste (lineal o no lineal), existencia de teorías previas sobre la correlación y tipo de dependencia (directa o inversa). Estudian la exactitud de la estimación del coeficiente de correlación. Concluyen que la estimación del coeficiente de correlación es más precisa cuando se hace a partir de un diagrama de dispersión que cuando se hace a partir de otra representación, aunque dicha precisión depende del tipo de tarea, la intensidad de la correlación, el tipo de covariación y el tipo de dependencia, y no depende del tipo de ajuste ni de la existencia de teorías previas.

Un reciente resumen de las investigaciones sobre covariación, desde el punto de vista didáctico, lo podemos encontrar en Moritz (2004); además, este autor estudia tres estrategias de razonamiento sobre la covariación: *a)* construir una tabla o un diagrama de dispersión a partir de la descripción verbal de un tipo de covariación; *b)* descripción verbal de la covariación mostrada en un diagrama de dispersión; *c)* descripción verbal de la covariación que muestran los datos numéricos de una tabla. Discute sus resultados teniendo en cuenta las dificultades descritas en la literatura de investigación, y las encontradas en este estudio y cómo evolucionan para favorecer el razonamiento sobre covariación; éstas son: *a)* desde considerar valores aislados hasta la tendencia general de la covariación; *b)* desde considerar una sola variable hasta la consideración de la covariación; *c)* desde los juicios basados en las teorías previas a los juicios basados en los datos presentados. Finaliza con unas implicaciones para la enseñanza y futuras investigaciones.

En las investigaciones anteriores, se utilizan diagramas de dispersión y se suele pedir a los sujetos la interpretación de algunos aspectos de los mismos, pero no hemos encontrado ninguna que estudie sistemáticamente la interpretación de los diagramas de dispersión de una muestra representativa de los mismos. Además, la interpretación de diagramas de dispersión es aconsejada por el currículo español y por el de los países de nuestro entorno como, por ejemplo, los Standards (NCTM, 2000). Por otra parte, estamos interesados en estudiar y caracterizar el significado personal de los diagramas de dispersión en una muestra de estudiantes de primer curso de bachillerato, ya que la interpretación de los mismos es una cuestión básica en el estudio del tema, generalmente situada al principio del proceso de estudio. En consecuencia, obtendremos conocimiento sobre el significado personal conseguido en el comienzo del proceso de estudio de la dependencia aleatoria en primer curso de bachillerato sobre la interpretación de diagramas de dispersión, que será de interés para la enseñanza del tema y para otras investigaciones.

3. MARCO TEÓRICO

Adoptamos como marco teórico el enfoque ontológico-semiótico de la cognición matemática (EOS), desarrollado por Godino y sus colaboradores, desde hace más de una década y publicado en diferentes trabajos, en distintos medios de difusión de la Didáctica de la Matemática; una recopilación estructurada de estos trabajos se puede consultar en Godino (2003).

En el EOS se define el significado de un objeto matemático como el sistema de prácticas (operatorias y discursivas) utilizadas en la resolución de problemas matemáticos (Godino y Batanero, 1994). Si el que resuelve el problema es un individuo se estará hablando de significado personal; si el sistema de prácticas es el adoptado en el seno de una institución, estaremos hablando de significado institucional. Se entiende por institución el

conjunto de personas que comparten campos de problemas y sus modos de resolución, socialmente aceptados. La institución es valedora del saber matemático. Ejemplos de instituciones son las personas que desarrollan el currículo de los diferentes niveles educativos (los libros de texto materializan el saber a enseñar); los investigadores matemáticos; otro tipo de instituciones serían las que utilizan el saber matemático (matemática aplicada, instituciones científicas, profesionales o comerciales que precisan de las matemáticas).

En un cierto nivel de enseñanza, comparando el significado institucional con el significado personal alcanzado por los alumnos, podremos caracterizar los aprendizajes de los alumnos sobre un objeto matemático determinado (Godino, 2003). En investigación didáctica se debe delimitar en primer lugar el significado institucional. El significado institucional de la correlación ha sido caracterizado en trabajos anteriores sobre asociación estadística, de los que podemos destacar, entre otros, Estepa (1994), Estepa y Sánchez (1998), Sánchez, Estepa y Batanero (2000). Este significado institucional, obtenido en nuestras investigaciones previas, lo hemos tenido en cuenta a la hora de construir el cuestionario.

4. OBJETIVOS

Generalmente al comienzo de la enseñanza de la correlación, se suele enfrentar a los estudiantes a tareas de interpretación de diagramas de dispersión. En consecuencia, el estudio que presentamos está dirigido a caracterizar el significado personal sobre diagramas de dispersión de los estudiantes de la muestra bajo estudio, es decir, caracterizar la capacidad para interpretar una muestra variada de diagramas de dispersión, que tiene una muestra de estudiantes del curso primero de bachillerato, después de estudiar el tema de dependencia no funcional. El estudio se centra en la estimación del coeficiente de correlación, la discriminación entre la dependencia funcional o aleatoria que muestran los diagramas de dispersión propuestos y la estimación gráfica de la recta de regresión a la que se pueden ajustar los puntos de los diagramas de dispersión presentados. Realizado el estudio se obtienen unas implicaciones para la planificación de la enseñanza y para futuras investigaciones.

5. MUESTRA

La muestra ha sido intencional, formada por grupos naturales de estudiantes de un instituto de bachillerato de Úbeda, provincia de Jaén (España), y está constituida por 75 estudiantes de cuatro grupos del primer curso de bachillerato de las modalidades de Ciencias de la Naturaleza y Salud (46 estudiantes: 19 chicos y 27 chicas) y de Humanidades y Ciencias Sociales (29 estudiantes: 17 chicos y 12 chicas) a los que se les había impartido el tema de dependencia no funcional, en el que se desarrollaron los contenidos estadísticos nube de puntos, estudio del grado de relación entre dos variables, representación

y análisis de la nube de puntos, correlación y regresión lineal, mediante una metodología tradicional, utilizando libros de texto de este nivel de enseñanza. Aunque el currículo de matemáticas para estas modalidades de bachillerato es distinto, los contenidos estadísticos relativos al presente estudio se han impartido completamente en ambas modalidades y de forma similar.

Como la muestra está formada por grupos naturales de estudiantes, nuestros resultados no se pueden extrapolar a ninguna población; creemos que se obtendrían resultados similares en otras muestras de características análogas.

6. CUESTIONARIO

El cuestionario fue cumplimentado por los estudiantes de la muestra al final del curso escolar 2005-2006. Los datos de este trabajo se han extraído de dos cuestionarios (Anexo 1), de un cuestionario más amplio sobre la dependencia aleatoria, correlación y regresión.

Con la primera cuestión (tomada de Estepa [1994] y actualizada) intentamos evaluar la capacidad de los estudiantes de interpretar los puntos aislados del diagrama de dispersión, requisito imprescindible para una correcta interpretación de la nube de puntos. En consecuencia, en esta cuestión realizamos dos preguntas: en la primera, dada la abscisa se pregunta por la ordenada y, en la segunda, dada la ordenada se pregunta por la abscisa. Aunque el gráfico está claro, es posible que algunos estudiantes se desvíen en una o dos unidades al leer la abscisa o la ordenada.

Con las respuestas a la cuestión 2 (Anexo 1) se pretende caracterizar el significado que los diagramas de dispersión tienen para los estudiantes del primer curso de bachillerato, una vez concluida la enseñanza del tema en un curso introductorio. Para ello construimos esta cuestión con una muestra representativa de diagramas de dispersión que abarcara los contenidos básicos que queríamos estudiar, esto es, la discriminación de la dependencia (funcional o aleatoria, lineal o no lineal), tipo de dependencia (independencia, directa e inversa), intensidad de la dependencia (asignación de un coeficiente de correlación a un diagrama de dispersión) y ajuste de una línea recta a la nube de puntos. En concreto esta cuestión consta de nueve diagramas de dispersión, numerados del 1 al 9. Tres muestran dependencia funcional, dos lineal (Gráficos 6 y 7) y uno no lineal (Gráfico 9). Los seis gráficos restantes presentan dependencia aleatoria, uno (Gráfico 1) muestra independencia, otro (Gráfico 2) dependencia aleatoria: no lineal, y cuatro (3, 4, 5, 8) dependencia aleatoria lineal con más o menos intensidad, dos de dependencia directa (3, 8) y otros dos de dependencia inversa (4, 5).

En esta cuestión se realizan cuatro preguntas, que comentaremos a continuación.

En la primera se pide la estimación del coeficiente de

correlación de cada uno de los 9 diagramas de dispersión. En la literatura de investigación, comentada anteriormente, hemos encontrado que la estimación del coeficiente de correlación no es tarea fácil para los estudiantes (por ejemplo, en Sánchez, Estepa y Batanero [2000], el error de estimación del coeficiente de correlación por estudiantes universitarios fue de $\pm 0,3$), creemos que es debido a que este coeficiente es adimensional y que los valores que puede tomar no son proporcionales. En consecuencia, hemos optado por dar, en el enunciado, los 9 coeficientes de correlación, teniendo el estudiante que discriminar entre ellos, con lo que la estimación efectuada tiene este referente. Por tanto, dado un diagrama de dispersión, el estudiante debe discernir si la correlación es directa o inversa, a continuación, su intensidad y, teniendo en cuenta los otros diagramas de dispersión, elegir el coeficiente de correlación más apropiado entre los 9 ofrecidos. En el análisis de resultados veremos que algunos estudiantes (muy pocos) responden un coeficiente de correlación no coincidente con uno de los 9 dados.

La segunda y tercera preguntas están referidas a que el estudiante discrimine si la dependencia es funcional o aleatoria en cada gráfico, para ello debe observar si los puntos dados en el gráfico pueden pertenecer todos a una línea, recta o curva, o por el contrario, no.

En la cuarta pregunta se pide al estudiante que ajuste una línea recta a la nube de puntos ofrecida en el gráfico. El estudiante debe observar la tendencia del patrón que siguen los puntos y ajustar la mejor línea que la aproxime, trazándola.

El análisis de datos se ha realizado con el paquete SPSS 12.0 para Windows.

En general el porcentaje de no respuestas oscila entre el 10% y el 20%, llegando en algunos casos a superar el 40%. Este índice de no respuestas nos indica que la tarea ha tenido cierta dificultad para esta muestra de estudiantes. Con la intención de hacer más comprensibles los resultados, hemos optado por calcular también el porcentaje de aciertos respecto a los que han respondido.

7. LECTURA DE LOS PUNTOS DE UN DIAGRAMA DE DISPERSIÓN

Un prerequisite esencial para la interpretación de un diagrama de dispersión es la lectura correcta de los puntos que lo conforman, ya que obviamente, si el estudiante no lee los puntos correctamente, no los interpretará de manera adecuada y el presente estudio quedaría invalidado (Estepa, 1994). Por este motivo hemos incluido la primera cuestión (Anexo I). Esta cuestión tiene dos preguntas, se pregunta por una coordenada y el estudiante debe responder con la otra. En la tabla 7.1 se dan las respuestas y los porcentajes a cada una de las preguntas.

7.1. Lectura de la ordenada, dada la abscisa

La respuesta correcta a la pregunta *a*) es 13. Como la interpretación del diagrama de dispersión siempre se hace valorando de manera aproximada las coordenadas de los puntos, se puede tomar como correcta la respuesta 14, dada su proximidad al 13; en consecuencia, el 80% de los estudiantes determina la ordenada del punto de abscisa dada. Las demás respuestas no son adecuadas y tienen frecuencia baja. Respecto al número de estudiantes que han respondido a la cuestión, el porcentaje de respuestas correctas es de 92,3%.

Tabla 7.1

Frecuencia y porcentaje de las respuestas a la lectura de las coordenadas de un punto del diagrama de dispersión.

a) Lectura de la ordenada, dada la abscisa		b) Lectura de la abscisa, dada la ordenada	
Respuesta	Frecuencia (%)	Respuesta	Frecuencia (%)
13 (correcta)	58 (77,3)	2 y 6	1 (1,3)
14	2 (2,7)	2 y 7	4 (5,3)
Otras	5 (6,7)	2 y 8 (correcta)	46 (61,3)
No responde	10 (13,3)	2 y 9	2 (2,7)
TOTAL	75 (100,0)	Otras	9 (12,0)
		No responde	13 (17,3)
		TOTAL	75 (99,9)

7.2. Lectura de la abscisa, dada la ordenada

Las respuestas correctas a la pregunta *b*) de la cuestión 1 (Anexo 1) son 2 y 8. En esta pregunta, hay tres estudiantes más que en la anterior que no han respondido, también hay más diversidad de respuestas; además en la respuesta se deben dar dos coordenadas, en lugar de una. Todo esto muestra una mayor dificultad de esta pregunta respecto de la anterior.

Si tomamos como correctas las respuestas 2 y 6, 2 y 7, 2 y 9 y 3 y 8 por su proximidad a la correcta 2 y 8, tenemos que el total de respuestas correctas sería de 54, lo que significa un 72% de respuestas correctas. Respecto a los estudiantes que han respondido a la pregunta, el porcentaje de respuestas correctas es de 87,09%.

Como conclusión podemos decir que aproximadamente un 90% de los estudiantes que han respondido a ambas preguntas interpretan correctamente las coordenadas de los puntos de un diagrama de dispersión y, en consecuencia, dicha lectura no será obstáculo para las respuestas a la segunda cuestión del anexo 1. Resultados similares se obtuvieron en Estepa (1994), Estepa y Batanero (1996) y Moritz (2004).

8. INTERPRETACIÓN DE LOS DIAGRAMAS DE DISPERSIÓN

Como hemos indicado anteriormente, la interpretación de un diagrama de dispersión es una de las primeras tareas que se les suele plantear a los estudiantes cuando se enfrentan al estudio de la dependencia aleatoria. Para caracterizar el significado que los estudiantes tienen sobre los diagramas de dispersión incluimos en el cuestionario la cuestión 2 del anexo 1, en la que se pide al estudiante, para cada uno de los 9 diagramas de dispersión propuestos, que estime el coeficiente de correlación, discrimine si la dependencia es funcional o aleatoria y ajuste una línea recta de manera adecuada a los puntos de los diagramas de dispersión ofrecidos. A continuación analizaremos las respuestas a cada una de las preguntas.

8.1. Dificultad de la tarea y tipo de dependencia en la estimación del coeficiente de correlación en los diagramas de dispersión

En la tabla 8.1 podemos observar que la dificultad de la estimación del coeficiente de correlación en los diagramas de dispersión presentados a los estudiantes es moderada. El número de no respuestas oscila entre 8 y 14, habiendo respondido todos los estudiantes a algunas preguntas, lo que nos indica que la tarea era asequible a los estudiantes de la muestra, ya que la han respondido entre el 81,3 y el 89,3% de los estudiantes. El gráfico que más respuestas ha tenido ha sido el gráfico 7, que presenta una dependencia funcional inversa. Los gráficos que menos respuestas han tenido han sido el 5 (dependencia aleatoria inversa) y el 9 (dependencia funcional no lineal); este último ha podido resultar un diagrama de dispersión «extraño» a algunos estudiantes.

En cuanto al tipo de dependencia, globalmente, la media del número de estudiantes que responden correctamente es muy similar: 54 estudiantes para la independencia, 54,75 para la correlación positiva y 55 para la correlación negativa, con lo que podemos afirmar que, aproximadamente, el 73% de los estudiantes de la muestra y el 86,6% de los que han respondido han dado un signo de la correlación correcto. Si estudiamos separadamente la correlación positiva y la negativa, tenemos que las respuestas correctas dependen del tipo del gráfico presentado. Así, para la dependencia directa, el gráfico 3 es el que más respuestas correctas ha tenido; este tipo de gráfico suele aparecer con más frecuencia en los libros de texto; además corrobora los resultados de otras investigaciones según las cuales la correlación fuerte positiva es la más fácil de detectar (p. e. Estepa y Batanero, 1995). El gráfico 8, también de correlación directa, tiene una diferencia con el anterior de 14,7% de respuestas correctas, diferencia acusada, aunque existe diferencia de intensidad de correlación entre los gráficos.

Para la dependencia inversa, los gráficos 2 y 4 son los que han tenido mayor número de respuestas correctas, y el gráfico 5 el que mayor respuestas incorrectas ha tenido, siendo en este caso la diferencia entre respuestas correctas e incorrectas más pequeña.

Tabla 8.1

Frecuencia y porcentaje del tipo de dependencia asignado por los estudiantes en la estimación del coeficiente de correlación de los 9 diagramas de dispersión.

RESPUESTA	GRÁFICO 1	GRÁFICO 2	GRÁFICO 3	GRÁFICO 4	GRÁFICO 5	GRÁFICO 6	GRÁFICO 7	GRÁFICO 8	GRÁFICO 9
	Frec. (%)								
Independencia	* 54 (72,0)	1 (1,3)	0 (0,0)	0 (0,0)	2 (2,7)	4 (5,3)	0 (0,0)	2 (2,7)	3 (4,0)
Correlación positiva	9 (12,0)	6 (8,0)	* 60 (80,0)	7 (9,3)	9 (12,0)	* 57 (76,0)	9 (12,0)	* 49 (65,3)	* 53 (70,7)
Correlación negativa	2 (2,7)	* 55 (73,3)	2 (2,7)	* 57 (76,0)	* 50 (66,7)	4 (5,3)	* 58 (77,3)	11 (14,7)	5 (6,7)
No responde	10 (13,3)	13 (17,3)	13 (17,3)	11 (14,7)	14 (18,7)	10 (13,3)	8 (10,7)	13 (17,3)	14 (18,7)
TOTAL	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)

* Respuesta correcta

Tabla 8.2

Frecuencia y porcentaje de las estimaciones del coeficiente de correlación en los gráficos que muestran independencia o correlación perfecta.

GRÁFICO 1, COEFICIENTE DE CORRELACIÓN 0,05		GRÁFICO 6, COEFICIENTE DE CORRELACIÓN +1,00		GRÁFICO 7, COEFICIENTE DE CORRELACIÓN -1,00	
Respuesta	Frec. (%)	Respuesta	Frec. (%)	Respuesta	Frec. (%)
-1,00	2 (2,7)	-1,00	4 (5,3)	-1,00 (correcta)	52 (69,3)
0,05 (correcta)	54 (72,0)	0,05	4 (5,3)	-0,98	1 (1,3)
0,65	1 (1,3)	0,90	1 (1,3)	-0,62	5 (6,7)
0,71	1 (1,3)	0,93	1 (1,3)	0,90	1 (1,3)
0,93	7 (9,3)	Otras	2 (2,7)	0,93	1 (1,3)
No responde	10 (13,3)	1,00 (correcta)	53 (70,7)	1,00	7 (9,3)
TOTAL	75 (99,9)	No responde	10 (13,3)	No responde	8 (10,7)
		TOTAL	75 (99,9)	TOTAL	75 (99,9)

Otro punto a destacar es que, aunque el gráfico 3 y 4 tienen forma similar y el signo de la correlación distinto, las respuestas erróneas en el gráfico 4 son superiores a las del gráfico 3, lo que confirma los resultados de investigaciones anteriores, que hemos comentado, sobre la interpretación de la dependencia inversa que es más difícil que la directa para los estudiantes (Estepa, 1994).

8.2. Estimación del coeficiente de correlación en diagramas de dispersión con independencia entre las variables, o bien con correlación perfecta (gráficos 1, 6 y 7)

Agrupamos estos tres tipos de gráficos por tener las correlaciones extremas, es decir, en términos absolutos sus coeficientes de correlación están en los extremos, mien-

tras que en la dependencia aleatoria el valor del coeficiente de correlación se sitúa entre estos extremos; además, son los que han tenido mayor número de respuestas y, en consecuencia, los estudiantes de la muestra los han considerado más asequibles para dar una respuesta. En la tabla 8.2 se dan los resultados de las respuestas a los tres gráficos.

Gráfico 1, coeficiente de correlación 0,05 (independencia). El 72% de los estudiantes de la muestra dan una respuesta correcta; si consideramos solamente los estudiantes que han respondido, este porcentaje se eleva a 83,08%. Nueve estudiantes dan una correlación positiva fuerte y dos asignan al gráfico una correlación negativa perfecta. En definitiva, once estudiantes, uno de cada seis, asignan una fuerte correlación a un diagrama de dispersión que muestra independencia.

Gráfico 6, coeficiente de correlación +1. Al tratarse de un diagrama de dispersión con correlación perfecta, solamente se puede admitir como correcta la respuesta 1 que ha sido expresada por 53 estudiantes, 70,7% del total de la muestra, el 81,54% de los estudiantes que han respondido. En cuanto a las respuestas incorrectas, cuatro estudiantes han respondido correlación negativa perfecta (-1); estos mismos estudiantes asignan al gráfico 7 un coeficiente de correlación 1, de lo que se deduce que aunque mantienen la intensidad de la correlación equivocan el signo, confunden la dependencia directa con la inversa. Los cuatro estudiantes que responden independencia (0,05) indican correlación negativa fuerte en el gráfico 7. Los cuatro estudiantes que responden correlación positiva fuerte no han admitido la dependencia funcional.

Gráfico 7, coeficiente de correlación -1. La respuesta correcta la dan 52 estudiantes que representan el 77,61% de los estudiantes que han dado respuesta. Entre las respuestas incorrectas, tenemos siete que dan una correlación perfecta, pero positiva, entre los que están los ya comentados en el gráfico 6 que reconocen la dependencia funcional, pero cambian el signo, seis dan una correlación negativa no perfecta (-0,98, -0,62). Dos estudiantes, responden una correlación fuerte positiva (0,90, 0,93). Ningún estudiante ha respondido independencia, lo que nos indica que en este punto nuestros resultados contradicen los de Morris (1997), ya que en su estudio el 15% de su muestra creía que una correlación negativa perfecta indicaba independencia.

8.3. Estimación del coeficiente de correlación en diagramas de dispersión con correlación positiva no perfecta (gráficos 3, 9 y 8)

En la tabla 8.3 damos las frecuencias y porcentajes de las respuestas a los gráficos de correlación positiva no perfecta.

Gráfico 3, coeficiente de correlación 0,93. Ya hemos visto en la literatura de investigación consultada (Sánchez, Estepa y Batanero, 2000) que el error de estimación del coeficiente de correlación es aproximadamente de $\pm 0,3$. Por otra parte, aunque el valor correcto es el de mayor frecuencia, 0,90 está muy próximo al valor correcto, de ahí su alta frecuencia; además, la correlación es positiva y alta, distinta de la perfecta; en consecuencia, se pueden considerar correctas las respuestas 0,71, 0,90, 0,93, 0,98, con lo que las respuestas correctas ascenderían a 55, es decir, el 73,3% de toda la muestra, o bien, 88,71% de los alumnos que han respondido a la cuestión. De acuerdo con la consideración anterior las respuestas incorrectas serían siete, dos de las cuales han confundido el signo de la correlación (-0,98), aunque mantienen la intensidad y cinco que han asignado correlación positiva perfecta (1,00). Estos cinco estudiantes han asignado al gráfico 4 un coeficiente de correlación de -1, lo que nos muestra que para estos cinco estudiantes, cuando los puntos se agrupan intensamente a lo largo de una recta, el coeficiente de correlación tomará el valor de ± 1 , aunque todos los puntos de la nube no estén sobre la recta.

Gráfico 9, coeficiente de correlación 0,90. La correlación es positiva alta y, como hemos indicado antes, se puede considerar una buena estimación de las respuestas 0,71, 0,90 y 0,93, con lo que podemos considerar que 52 estudiantes, el 69,3% de la muestra, responde correctamente, lo que equivale al 85,24% de los que han respondido a la pregunta. Además, se puede observar que la respuesta correcta es la de mayor frecuencia y que entre los valores más cercanos se reparten las frecuencias consideradas como correctas. Los cinco estudiantes que responden -0,62 confunden el signo de la correlación, y el estudiante que responde 1,0 estimamos que ha creído que, al ser la relación perfecta entre las variables dependencia funcional no lineal, el coeficiente de correlación lineal debería ser 1,0. Algo similar pero con la independencia deben haber pensado los tres estudiantes que responden 0,05, es decir, la relación no es lineal, luego debe haber independencia; no olvidemos que solamente han estudiado la correlación lineal y el ajuste lineal.

Tabla 8.3

Frecuencia y porcentaje de las estimaciones del coeficiente de correlación en los gráficos que muestran correlación positiva no perfecta.

GRÁFICO 3, COEFICIENTE DE CORRELACIÓN 0,93		GRÁFICO 9, COEFICIENTE DE CORRELACIÓN 0,90		GRÁFICO 8, COEFICIENTE DE CORRELACIÓN 0,71	
Respuesta	Frec. (%)	Respuesta	Frec. (%)	Respuesta	Frec. (%)
-0,98	2 (2,7)	-0,62	5 (6,7)	-0,98	3 (4,0)
0,71	2 (2,7)	0,05	3 (4,0)	-0,90	5 (6,7)
0,90	22 (29,3)	0,71	16 (21,3)	-0,62	3 (4,0)
0,93 (correcta)	30 (40,0)	0,90 (correcta)	19 (25,3)	0,05	2 (2,7)
0,98	1 (1,3)	0,93	17 (22,7)	0,71 (correcta)	30 (40,0)
1,00	5 (6,7)	1,0	1 (1,3)	0,90	15 (20,0)
No responde	13 (17,3)	No responde	14 (18,7)	0,93	3 (4,0)
TOTAL	75 (100,0)	TOTAL	75 (100,00)	1,00	1 (1,3)
				No responde	13 (17,3)
				TOTAL	75 (99,9)

Gráfico 8, coeficiente de correlación 0,71. La correlación es fuerte y, como hemos comentado anteriormente, se pueden considerar correctas las respuestas 0,71, 0,90 y 0,93; además, se puede observar que la respuesta correcta es la que tiene mayor frecuencia y que 0,90 tiene una frecuencia muy superior a 0,93; en consecuencia, el número de respuestas correctas sería de 48, el 64,9% de la muestra de estudiantes correspondiente al 77,42% de los estudiantes que han dado respuesta. En las respuestas incorrectas se pueden destacar las que han dado una correlación fuerte negativa (once respuestas), que suponen un 14,7% de los estudiantes de la muestra; de éstos, la mayoría (nueve) han respondido un coeficiente de correlación positivo en el gráfico 5, lo que significa que han conceptualizado el signo de la correlación de manera errónea, intercambiándolo.

8.4. Estimación del coeficiente de correlación en diagramas de dispersión con correlación negativa no perfecta (gráficos 4, 2 y 5)

En la tabla 8.4 damos las frecuencias y porcentajes de las respuestas a los gráficos de correlación negativa no perfecta.

Gráfico 4, coeficiente de correlación -0,98. La respuesta correcta tiene la máxima frecuencia; la siguiente de máxima frecuencia es la más cercana a este valor (en esto no contamos -0,93 porque en el enunciado sólo se proporcionaba 0,93). La correlación de este gráfico es fuerte y negativa; como ya hemos comentado anteriormente, se pueden considerar acertadas las respuestas -0,93, -0,90, e incluso -0,62, con lo que el número de respuestas aceptables sería de 52, que representan un 69,3% de los estudiantes de esta muestra y un 81,25% de los estudiantes que han respondido la pregunta. Las respuestas incorrectas son las cinco

que han respondido -1. Como ya hemos comentado; en la sección 8.3, estos cinco estudiantes asignan 1 al gráfico 3, lo que indica que estos estudiantes confunden la correlación aleatoria fuerte con la dependencia funcional. También son incorrectas las que consideran que la correlación es fuerte y positiva, respuestas 0,71, 0,90, 0,93 y 0,98, en total siete respuestas (9,3%), pues en el gráfico dado la correlación es fuerte, pero negativa, o sea, confunden el signo de la correlación.

Gráfico 2, coeficiente de correlación -0,90. La respuesta correcta es la que mayor número de respuestas ha obtenido, siendo la siguiente en frecuencia la más cercana; aunque -0,62 ha obtenido una frecuencia alta de respuestas, posiblemente las respuestas se reparten entre estos tres valores, dada la forma de la nube de puntos. La correlación es negativa y fuerte y, como hemos comentado en otros gráficos similares, podemos considerar como correctas las respuestas -0,98 y -0,62, con lo que el porcentaje de respuestas correctas sería de 73,3% de los estudiantes de la muestra, o bien el 88,71% de los estudiantes que han respondido. Las demás respuestas erróneas son de dos tipos: los estudiantes que han estimado una correlación positiva y fuerte (0,71, 0,90, 0,98) que son en total seis (8,0% de la muestra y 9,67% de los que han respondido) y un estudiante que ha optado por la independencia.

Gráfico 5, coeficiente de correlación -0,62. La respuesta correcta es la que ha obtenido mayor frecuencia. Como hemos comentado en los primeros gráficos analizados, por su proximidad, se pueden considerar correctas las respuestas -0,90 (segunda en frecuencia) y -0,98, con lo que tendríamos un total de respuestas aceptables de 48, que representan el 64,0% del total de la muestra de estudiantes y el 78,69% de los estudiantes que han respondido. En las respuestas inco-

Tabla 8.4

Frecuencia y porcentaje de las estimaciones del coeficiente de correlación en los gráficos que muestran correlación positiva no perfecta.

GRÁFICO 4, COEFICIENTE DE CORRELACIÓN -0,98		GRÁFICO 2, COEFICIENTE DE CORRELACIÓN -0,90		GRÁFICO 5, COEFICIENTE DE CORRELACIÓN -0,62	
Respuesta	Frec. (%)	Respuesta	Frec. (%)	Respuesta	Frec. (%)
-1,00	5 (6,7)	-0,98	19 (25,3)	-1,00	2 (2,7)
-0,98 (correcta)	30 (40,0)	-0,90 (correcta)	21 (28,0)	-0,98	4 (5,3)
-0,93	1 (1,3)	-0,62	15 (20,0)	-0,90	17 (22,7)
-0,90	18 (24,0)	0,05	1 (1,3)	-0,62 (correcta)	27 (36,0)
-0,62	3 (4,0)	0,71	4 (5,3)	0,00	1 (1,3)
0,71	1 (1,3)	0,90	1 (1,3)	0,05	1 (1,3)
0,90	1 (1,3)	0,98	1 (1,3)	0,62	1 (1,3)
0,93	2 (2,7)	No responde	13 (17,3)	0,71	3 (4,0)
0,98	3 (4,0)	TOTAL	75 (99,8)	0,90	3 (4,0)
No responde	11 (14,7)			0,93	2 (2,7)
TOTAL	75 (100,0)			No responde	14 (18,7)
				TOTAL	75 (100,0)

rectas destacan los nueve estudiantes (12% de la muestra y 19,7% de los que han respondido a esta pregunta) que le han asignado una correlación positiva alta; de estos nueve estudiantes, seis asignan una correlación negativa alta al gráfico 8 (gráfico similar con dependencia directa), lo que indica que han conceptualizado erróneamente el signo de la correlación.

En los tres gráficos anteriores de correlación negativa como en los tres de correlación positiva no perfecta (3, 8 y 9) se puede observar que la respuesta correcta es la moda y que las frecuencias de los valores próximos a la respuesta correcta son mayores conforme están más cerca de la respuesta correcta, lo que nos indica el carácter aproximado de la estimación del valor del coeficiente de correlación.

8.5. Dependencia funcional versus dependencia aleatoria

La capacidad para discriminar si un diagrama de dispersión presenta dependencia funcional o aleatoria es importante en la interpretación de dicho diagrama. En el cuestionario se incluyeron, en la cuestión 2, las preguntas b) y c) relativas a este tema. En la tabla 8.5 podemos observar el alto índice de no respuestas, que oscila entre el 40% para el gráfico 6 y 7 y el 48% para el gráfico 2. Esto nos da un alto índice de dificultad para esta tarea, en apariencia sencilla.

Aproximadamente el 60% de los estudiantes de la muestra responden cuando la dependencia es funcional. La dependencia funcional lineal es detectada por prácticamente todos los estudiantes que responden (Gráficos 6 y 7), la dependencia funcional no lineal es detectada por el 48% de los estudiantes de la muestra, es decir, por el 81,8% de los estudiantes que responden, habiendo un número significativo de estudiantes, ocho (10,7% de la muestra y 18,2% de los que dan respuesta), que responde

dependencia aleatoria, cuando es funcional no lineal. Los estudiantes de la muestra solamente han estudiado regresión lineal, lo que puede haberles producido la creencia de que la dependencia en los diagramas de dispersión solamente puede ser dependencia funcional lineal o dependencia aleatoria.

Las respuestas en los diagramas de dispersión con dependencia aleatoria han oscilado entre el 56 y el 52% de los estudiantes de la muestra.

La dependencia aleatoria es aceptada para el gráfico 1 por la totalidad de los estudiantes que han respondido, para el gráfico 5 y 8, por aproximadamente el 80% de los estudiantes que han respondido. Sin embargo, cerca del 60% de los estudiantes que han respondido no aceptan la dependencia aleatoria para los gráficos 2, 3 y 4, lo que supone una carencia de importancia, ya que gráficos similares al 3 y 4 suelen incluirse en los libros de texto. Parece ser que estos estudiantes, cuando los puntos de la nube están cercanos, marcando una fuerte tendencia, con una correlación fuerte, pero no perfecta, consideran que existe dependencia funcional, dejando para la dependencia aleatoria los gráficos cuyos puntos están más dispersos en la nube como en los gráficos 1, 5 y 8.

8.6. Dibujar una línea que se ajuste a los puntos del diagrama de dispersión

Truran (1997) encontró dificultades en la asunción de linealidad y aconseja investigación de lo lineal en la asociación; además, el trazado de una recta que aproxime la dependencia en un diagrama de dispersión es otra de las competencias importantes en la formación de los estudiantes de este nivel; por estos motivos se formuló esta pregunta que ha registrado un alto índice de dificultad, destacando el gráfico 1 (independencia) con un 53,4% de no respuestas; los estudiantes pueden haber pensado que si existe independencia no tiene sentido la recta de

Tabla 8.5
Frecuencia y porcentaje de las respuestas dadas a dependencia funcional o aleatoria en los diagramas de dispersión.

RESPUESTA	GRÁFICO 1	GRÁFICO 2	GRÁFICO 3	GRÁFICO 4	GRÁFICO 5	GRÁFICO 6	GRÁFICO 7	GRÁFICO 8	GRÁFICO 9
	Frec. (%)								
Dependencia aleatoria	41* (54,7)	16* (21,3)	14* (18,7)	14* (18,7)	34* (45,3)	0 (0,0)	1 (1,3)	33* (44,0)	8 (10,7)
Dependencia funcional	0 (0)	23 (30,7)	26 (34,7)	26 (34,7)	8 (10,7)	45* (60,0)	44* (58,7)	9 (12,0)	36* (48,0)
No responde	34 (45,3)	36 (48,0)	35 (46,6)	35 (46,6)	33 (44,0)	30 (40,0)	30 (40,0)	33 (44,0)	31 (41,3)
Total	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)

* Respuesta correcta

regresión. Los demás gráficos han tenido un índice de no respuestas que oscila entre el 34,7% para los gráficos 6 y 7 (dependencia funcional lineal) al 40% para el gráfico 8 (dependencia aleatoria no muy alta). Los resultados de las respuestas están en la tabla 8.6.

En cuanto a la corrección de las respuestas, podemos destacar que en los diagramas de dispersión de dependencia funcional lineal (Gráficos 6 y 7) todos los estudiantes que han respondido han trazado la línea correctamente; los gráficos 3, 4, 5 y 8, dos de dependencia aleatoria directa y dos de inversa, han tenido muy baja tasa de respuestas incorrectas. Los gráficos 1, 2 y 9 son los que han tenido una tasa de respuestas incorrectas más alta. El gráfico 1, que muestra independencia, aparece casi como un círculo, lo que puede haber provocado que el estudiante trace la recta de forma inadecuada, o bien, como ya hemos comentado, que no tenga sentido la recta de regresión. El gráfico 2, al tener una forma «muy uniforme», puede haber provocado las nueve respuestas incorrectas, algo similar ocurre con el gráfico 9. Por consiguiente, podemos inferir que los diagramas de dispersión cuya forma no es aproximadamente una elipse (la elipse convencional mostrada en la mayoría de los libros de texto), o bien, muestran dependencia funcional no lineal, presentan mayor dificultad a los estudiantes para trazar una línea recta que ajuste los datos.

9. CONCLUSIONES

Al final de la sección 2 y en la sección 4 nos propusimos caracterizar el significado personal que una muestra de estudiantes de primer curso del actual bachillerato da a los diagramas de dispersión después de recibir la enseñanza del tema. A continuación resumimos y discutimos los resultados del estudio realizado.

En primer lugar, podemos afirmar que, aproximadamente el 90% de los estudiantes que han respondido, leen correctamente las coordenadas de los puntos del diagrama

de dispersión; en consecuencia los errores que cometen los estudiantes en las respuestas a la cuestión 2 no pueden ser atribuidos a una mala lectura de los puntos.

La asignación de un coeficiente de correlación a un diagrama de dispersión tiene una dificultad moderada, puesto que el porcentaje de no respuesta oscila entre, 13,3% y el 18,7% de los estudiantes de la muestra. No ocurre lo mismo con la discriminación entre dependencia funcional y aleatoria pues, en este caso, el porcentaje de no respuestas ha sido del 44%. Sin embargo, la ausencia de respuestas en promedio ha estado próxima al 40% en el ajuste de una recta a la nube de puntos, excepto para el gráfico 1 que ha sido del 43,4%; en consecuencia, la dificultad de las preguntas depende del contenido de éstas y del gráfico propuesto.

La detección de la independencia, dependencia directa e inversa depende del tipo de gráfico, aunque no hemos observado grandes diferencias entre gráficos; en general, aproximadamente el 73% de los estudiantes de la muestra distinguen estos tipos de dependencia.

El 70% de los estudiantes reconocen la dependencia funcional con su signo adecuado y el 5,3% de los estudiantes reconocen la dependencia funcional pero cambian el signo.

En los gráficos con correlación positiva, un promedio de 52 estudiantes (69,3%) dan una respuesta correcta a la estimación del coeficiente de correlación. En la dependencia aleatoria positiva, el número de respuestas correctas varía paulatinamente desde el 73,3% para el gráfico de más intensidad de correlación, 0,93, hasta el 64,9% en el gráfico de menor intensidad de correlación 0,71, lo que corrobora resultados de investigaciones anteriores según los cuales la correlación fuerte positiva es más fácil de detectar por principiantes; cuanto más fuerte es la correlación más fácil es de detectar. En consonancia con esto también hemos encontrado que en los gráficos con dependencia aleatoria negativa, el porcentaje medio de respuestas correctas es similar al de los de la correlación

Tabla 8.6
Frecuencia y porcentaje de las respuestas dadas al ajuste de una línea en los diagramas de dispersión.

RESPUESTA	GRÁFICO 1	GRÁFICO 2	GRÁFICO 3	GRÁFICO 4	GRÁFICO 5	GRÁFICO 6	GRÁFICO 7	GRÁFICO 8	GRÁFICO 9
	Frec. (%)								
Correcta	25 (33,3)	37 (49,3)	44 (58,6)	46 (61,4)	43 (57,3)	49 (65,3)	49 (65,3)	42 (56,0)	36 (48,0)
Incorrecta	10 (13,3)	9 (12,0)	2 (2,7)	1 (1,3)	2 (2,7)	0 (0,0)	0 (0,0)	3 (4,0)	10 (13,3)
No responde	40 (53,4)	29 (38,7)	29 (38,7)	28 (37,3)	30 (40,0)	26 (34,7)	26 (34,7)	30 (40,0)	29 (38,7)
Total	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)	75 (100,0)

positiva, pero en este caso no desciende paulatinamente el porcentaje de respuestas correctas con la intensidad de la correlación.

La dependencia funcional lineal es más fácil de detectar para estos estudiantes que la dependencia funcional no lineal, aunque en ambos casos el porcentaje de respuestas correctas es menor que en las preguntas anteriores.

La dependencia aleatoria fuerte es interpretada por algunos estudiantes como dependencia funcional; este tipo de confusión es de importancia y se debe tener en cuenta en el proceso de enseñanza.

El ajuste correcto de una línea recta a una nube de puntos depende del gráfico, siendo más fácil para estos estudiantes el caso de dependencia funcional lineal y el más difícil, el de independencia.

A lo largo del trabajo hemos podido observar que tareas en apariencia sencillas, como las que aquí hemos planteado a los estudiantes, tienen una dificultad que depende de la tarea planteada y el tipo de gráfico sobre la que se plantea. Algunos de nuestros resultados coinciden con investigaciones anteriores, como hemos indicado en el texto; cuando no se indica coincidencia, los resultados obtenidos son aportes de esta investigación.

Creemos que hemos cumplido el objetivo que nos planteábamos al principio de caracterizar el significado personal que una muestra de estudiantes de bachillerato da a la interpretación de una muestra significativa de diagramas de dispersión; nuestros resultados confirman esta afirmación. Dicha caracterización creemos que es de interés para la planificación de la enseñanza y para futuras investigaciones.

Como hemos dicho en la descripción de la muestra, la enseñanza recibida por esta muestra de estudiantes ha sido de tipo tradicional, con explicaciones del profesor, uso del libro de texto y realización de actividades de refuerzo por parte de los estudiantes. Sería conveniente en próximas investigaciones hacer un estudio de la enseñanza realizada y continuar la indagación sobre la asimilación de otros elementos del tema de dependencia aleatoria en bachillerato, como cálculo e interpretación del coeficiente de correlación, ajuste de la nube de puntos a una recta e interpretación de esta recta y sus elementos.

NOTA

Esta investigación está realizada en el marco del proyecto de investigación BSO2003-06331, subvencionado por la Secretaría de Estado de Política Científica y Tecnológica. Ministerio de Educación y Ciencia. Madrid.

REFERENCIAS BIBLIOGRÁFICAS

- BATANERO, C., ESTEPA, A. y GODINO, J.D. (1997). Students' understanding of Statistical association in a computer environments, en Garfield, J. y Burrill, G. (eds.). *Research on the Role of Technology in Teaching and Learning Statistics*. International Statistical Institute, pp. 191-205. Voorburg (Holanda): International Statistical Institute.
- BATANERO, C. y GODINO, J.D. (1998). Understanding graphical and numerical representations of statistical association in a computer environment, en Pereira-Mendoza, L., Seu Kea, L., Wee Kee, T. y Wong, W. (eds.). *Proceedings of the Fifth Conference on Teaching Statistics*, vol. 2, pp. 1017-1024. Voorburg (Holanda): International Statistical Institute.
- BATANERO, C., GODINO, J. y ESTEPA, A. (1998). Building the meaning of statistical association through data analysis activities (Research Forum), en Olivier, A. y Newstead, K. (eds.). *Proceedings of the 22nd Conference of the International Group for the Psychology of Mathematics Education*, 1, pp. 221-236. Stellenbosh (South Africa): University of Stellenbosh.
- CURCIO, F.R. y ARTZT, A.F. (1997). Assessing Students' Statistical Problem-Solving Behaviours in a Small-Group Setting, en Gal, I. y Garfield, J.B. (eds.). *The assessment Challenge in Statistics Education*. Amsterdam (Holanda): International Statistical Institute and IOS Press.
- CHAMBERS, J.M., CLEVELAND, W.S., KLEINER, B. y TUKEY, P.A. (1983). *Graphical Methods for Data Analysis*. Nueva Jersey: Bell Telephone Laboratories Incorporated.
- ESTEPA, A. (1994). «Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores». Tesis doctoral. Barcelona: ETD Micropublicaciones, S.L.
- ESTEPA, A. y BATANERO, C. (1995). Judgments of association in scatterplots, en Garfield, J. (ed.). *Research Papers from the Fourth International Conference on Teaching Statistics*, pp. 117-124. Minnesota: Universidad de Minnesota: The International Study Group for Research on Learning Probability and Statistics.
- ESTEPA, A. y BATANERO, C. (1996). Judgments of correlation in scatterplots: An empirical study of students' intuitive strategies and preconceptions. *Hiroshima Journal of Mathematics Education*, 4, pp. 25-41.
- ESTEPA, A. y SÁNCHEZ, F.T. (1998). Correlation and regression in secondary school text books, en Pereira-Mendoza, L., Seu, L., Wee, T. y Wong, W.K. (eds.). *Proceedings of the Fifth International Conference on Teaching of Statistics*, 2, pp. 671-676. Voorburg (Holanda): International Statistical Institute Permanent Office.
- GODINO, J.D. (2003). *Teoría de las Funciones Semióticas. Un enfoque ontológico-semiótico de la cognición e instrucción matemática*. Universidad de Granada. Departamento de Didáctica de la Matemática. Consultado en (2-marzo-07): <<http://www.ugr.es/~jgodino/funciones-semioticas/monografiatfs.pdf>>.
- GODINO, J.D. y BATANERO, C. (1994). Significado institucional y personal de los objetos matemáticos. *Recherches en Didactique des Mathématiques*, 14(3), pp. 325-355.
- MORITZ, J. (2004). Reasoning about covariation, en Ben-Zvi, D. y Garfield, J. (eds.). *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*, pp. 227-255. Dordrecht (Holanda): Kluwer Academic Publishers.
- MORRIS, E.J. (1997). *An investigation of students' conceptions and procedural skills in the statistical topics correlation*. Londres, The Open University: Centre for Information Technology in Education.
- NCTM (2000). *Principles and Standards for School Mathematics*. Reston, Virginia: National Council of Teachers of Mathematics, en <<http://standards.nctm.org/>>.
- SÁNCHEZ, F.T., ESTEPA, A. y BATANERO, C. (2000). Un estudio experimental de la estimación de la correlación a partir de diferentes representaciones. *Enseñanza de las Ciencias*, 18(2), pp. 297-310.
- TRURAN, J. (1997). Understanding of association and regression by first year economics students from two different countries as revealed in responses to the same examination question, en Garfield, J.B. y Truran, J.M. (eds.). *Research Papers on Stochastic Education from 1997*, pp. 205-212. Minnesota: University of Minnesota, Department Educational Psychology.
- TUKEY, J.W. (1962). The future of Data Analysis. *Annals of Mathematical Statistics*, 33, pp. 1-67.
- TUKEY, J.W. (1977). *Exploratory Data Analysis*. Nueva York: Addison-Wesley.

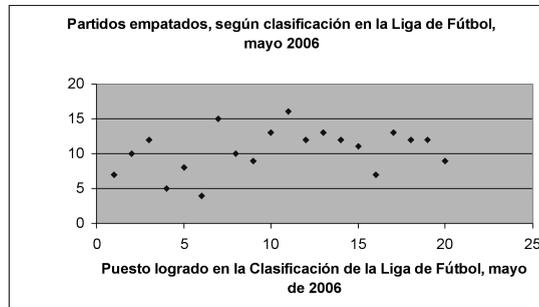
[Artículo recibido en abril de 2007 y aceptado en enero de 2008]

ANEXO 1

Cuestionario utilizado en este estudio.

1. En el siguiente diagrama hemos representado el puesto ocupado por cada uno de los 20 equipos de primera división de la liga de fútbol en la temporada 2005-2006 y los partidos empatados. (Un equipo de fútbol consigue tres puntos cuando gana un partido, 1 punto cuando empata y 0 puntos cuando pierde. La clasificación final en la liga se realiza según el número total de puntos obtenido. El que más puntos consigue obtiene el primer puesto –puesto 1–, el que menos puntos consigue obtiene el último puesto –puesto 20–).

- a) ¿Cuántos partidos empató el equipo clasificado en décimo lugar? _____
- b) Los equipos que empataron 10 partidos, ¿en qué puesto quedaron clasificados? _____



2. Los siguientes gráficos, diagramas de dispersión o nube de puntos tienen por coeficientes de correlación: -1; -0,98; -0,90; -0,62; 0,05; 0,71; 0,90; 0,93; 1

Se pide:

- a) Escriba debajo de cada gráfico el coeficiente de correlación que crea que le corresponde
- b) Escribir una «F» dentro de los gráficos que muestran una dependencia funcional
- c) Escribir una «A» dentro de los gráficos que muestran una dependencia aleatoria
- d) Dibujar una línea recta, en cada gráfico, que aproxime lo mejor posible los puntos representados

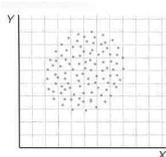


Gráfico 1

Coefficiente de correlación = _____

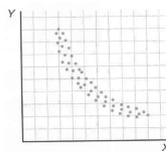


Gráfico 2

Coefficiente de correlación = _____

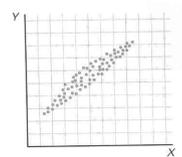


Gráfico 3

Coefficiente de correlación = _____

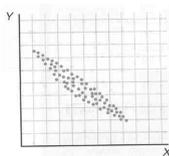


Gráfico 4

Coefficiente de correlación = _____

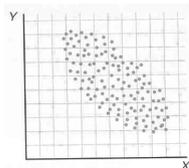


Gráfico 5

Coefficiente de correlación = _____

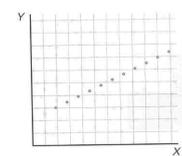


Gráfico 6

Coefficiente de correlación = _____

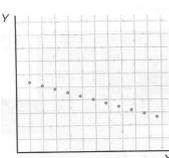


Gráfico 7

Coefficiente de correlación = _____

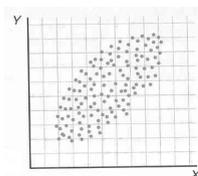


Gráfico 8

Coefficiente de correlación = _____

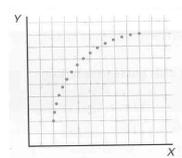


Gráfico 9

Coefficiente de correlación = _____

The Interpretation of Scatter Plots by Baccalaureate Students

ESTEPA CASTRO, ANTONIO

Universidad de Jaén. Facultad de Humanidades y Ciencias de la Educación. Departamento de Didáctica de las Ciencias.
Campus de las Lagunillas, Jaén
aestepa@ujaen.es

Abstract

Nowadays, correlation is an important topic in Secondary Education. Teaching correlation usually begins with proposing the interpretation of scatter plots to the students and obtaining useful information on diverse types of these graphics, because these graphs are the graphical tool most-used to analyze the relationship that can exist between two variables referring to the same sample. Consequently, given the interest of the topic, we have decided to do research on the interpretation of scatter plots by students currently studying for the Spanish Baccalaureate.

We have done a review of previous educational research into scatter plots and have not found any research where the baccalaureate students interpret a representative sample of scatter plots. Prior to this, we conducted an epistemological study of correlation. Consequently, we propose the following objective: characterizing the personal meaning that students currently studying for the Spanish Baccalaureate give the correlation when they interpret a representative sample of scatter plots. We focus the study on the estimation of correlation coefficients in scatter plots, the discrimination between functional dependency and random dependency and the estimation of a regressional straight line that adjusts to the points of the scatter diagram presented in the items.

The theoretical framework used is the semiotic-ontological approach of mathematical cognition (EOS), which distinguishes between personal and institutional meaning; by comparing the personal meaning obtained by the students with the institutional meaning of reference, it is possible to characterize the students' learning.

The questionnaire has two questions and it is part of another, used in another, broader research study. The capacity of the students to read the points on the scatter diagram is evaluated with the first question. 9 scatter plots with 9 coefficients of correlation are proposed in the second question. The students are required to: a) assign a correlation coefficient to each scatter plot, b) say if the dependency shown in the scatter plots is functional or random, c) draw a straight line that approximates the points represented in each scatter plot as best as possible.

This questionnaire was answered by a total sample of 75 students from four groups in their first year of baccalaureate

including 46 students (19 boys and 27 girls) in the Natural Sciences and Health section and 29 students (17 boys and 12 girls) in the Humanities and Social Sciences section. These students have received traditional-type instruction which involves explanations by the professor, the text book and the accomplishment of tasks on the topic under study.

Next, we summarize the results obtained.

a) 90% of the students in the sample correctly read the point coordinates on scatter plots, an essential requirement for them to be able to interpret a scatter plot correctly.

b) Difficulty of the tasks: Assigning a coefficient of correlation to a scatter plot is of moderate difficulty, the index of no answers range from 13.3% to 18.7%; the discrimination of functional or random dependence showed an index of non answers of around 44% and the adjustment of a straight line to the scatter plot shows an absence of answers of around 40%. Consequently, the difficulty of the task depends on the content of the question and the proposed graph.

c) The discrimination between direct and inverse dependence depends on the type of graph; 73% of the students discriminate it correctly.

d) The estimation of the correlation coefficient depends on the intensity of the correlation.

e) The linear functional dependence is easier to detect than no linear dependence.

f) Some students confuse the strong random dependence with the functional dependence.

g) The correct adjustment of a straight line to a scatter plot depends on the graph presented.

Some of our results agree with the results from previous research, others are unique to this research. We emphasized that apparently simple tasks, as those presented here, are difficult, depending on the tasks and the graph presented.

Let us think that our results are of interest to teaching the topic and to continuing to do research on other elements involving correlation and regression and on teaching the topic.