

UN ESTUDIO EXPERIMENTAL DE LA ESTIMACIÓN DE LA CORRELACIÓN A PARTIR DE DIFERENTES REPRESENTACIONES*

SÁNCHEZ COBO, FRANCISCO T.¹, ESTEPA CASTRO, ANTONIO² y BATANERO BERNABEU, CARMEN³

¹ Universidad de Jaén. Facultad de Humanidades y Ciencias de la Educación. Departamento de Didáctica de las Ciencias. Campus de Las Lagunillas, s/n. Edificio D-2. 23071 Jaén. E-mail: aestepa@ujaen.es

² Universidad de Granada. Facultad de Ciencias de la Educación. Departamento de Didáctica de las Matemáticas. Campus de Cartuja. 18071 Granada. E-mail: batanero@goliat.ugr.es

³ Universidad de Jaén. Escuela Politécnica Superior. Departamento de Matemáticas. Av. de Madrid, s/n. 23071 Jaén. E-mail: fsanchez@ujaen.es

SUMMARY

Correlation is a relevant topic at secondary school and university level. However, most didactical research in this field comes from psychology and only refers to 2x2 contingency tables. On the other hand, representation has fundamental influence on understanding mathematical concepts. In this paper, we take into account the different representations of the correlation: verbal description, numerical table, scatter plot, and correlation coefficient, and we study the different translations between representations. Taking into account the most important task variables in this type of problems, we have studied the accuracy in estimating correlation coefficients, and the students' strategies to solve these tasks. We have identified two factors relating the tasks and strategies. We conclude with some implications for the teaching of the topic.

INTRODUCCIÓN

El concepto de *correlación* es un tema destacado en el currículo estocástico de la enseñanza secundaria y en los primeros cursos universitarios, tanto por su importancia en sí mismo como por su relación con las ideas de *probabilidad* y *función* y por ser requisito previo para comprender conceptos, técnicas y procedimientos estadísticos posteriores en el currículo, como regresión simple y múltiple, análisis de varianza y diversos métodos multivariantes.

Las investigaciones sobre la comprensión de la correlación se inician con los trabajos de Inhelder y Piaget (1955), quienes utilizan tan sólo variables dicotómicas (tablas de contingencia 2x2) y analizan la habilidad de los chicos a partir de 12-13 años para diferenciar la correlación directa, inversa e independencia, así como las estrategias que emplean para establecer sus juicios. A pesar de la sencillez de la tarea propuesta, los trabajos de

Inhelder y Piaget muestran que hasta la etapa de las operaciones formales no se desarrolla la idea de correlación

Los trabajos posteriores en el campo de la psicología se centran particularmente en el análisis de la exactitud en los juicios de correlación en tablas 2x2 (Shaklee y Tucker, 1980; Crocker, 1981; Beyth-Marom, 1982). En otros casos, como Chapman y Chapman (1969), Jennings, Amabile y Ross (1982) y Trolhier y Hamilton (1986), se estudia la influencia que las teorías previas que se poseen sobre la relación entre las variables presentadas en el problema tienen sobre la exactitud de la estimación de la correlación. La conclusión general es que, cuando los datos del problema no coinciden con las expectativas del sujeto, se produce un conflicto cognitivo que afecta a la precisión en la estimación de la correlación.

Aunque la investigación psicológica se ha concentrado, en gran parte, en las tablas de contingencia 2×2 , algunos psicólogos han estudiado la capacidad para estimar la correlación a partir de diagramas de dispersión o desde un conjunto de pares de valores (Erlick y Mills, 1967; Jennings, Amabile y Ross, 1982; Lane et al., 1985). En general, se considera que los adultos tienen una capacidad limitada para estimar la correlación, mejorando cuando ésta es fuerte y positiva.

Dentro de la educación de la matemática, Estepa y Batanero (1996) estudian las estrategias de los estudiantes en los juicios de correlación a partir de nubes de puntos, identificando diversas concepciones erróneas sobre la misma. Truran (1995, 1997) describió la comprensión que de la asociación, regresión y del coeficiente de correlación poseen los estudiantes universitarios. Morris (1997, 1998) y Batanero, Estepa y Godino (1997), Batanero y Godino (1998) y Batanero, Godino y Estepa (1998) estudian las concepciones y comprensión de la correlación por estudiantes universitarios, así como su evolución después de experiencias de enseñanza. En estos trabajos previos se ha observado que los alumnos no aprecian, a veces, la correlación inversa, que tienen un sentido determinista o local de la correlación y que identifican, con frecuencia, la correlación con la causalidad. Como es conocido, asociación y causalidad no siempre son coincidentes, a pesar de que uno de los pasos en la búsqueda de relaciones causales es estudiar la covariación de las variables (Pozo, 1987).

Aunque, como vemos, los estudios sobre este tema son numerosos, los autores se centran en el análisis de las concepciones de los estudiantes, el efecto sobre las estimaciones de variables aisladas o el análisis de la dificultad de tareas particulares. No hemos encontrado un análisis sistemático de las variables en las tareas de traducción de la correlación de una representación a otra, ni un estudio de su efecto sobre la estimación de la correlación y las estrategias empleadas por alumnos que han estudiado el tema de la correlación. En esta investigación abordamos este estudio, ya que lo juzgamos relevante para la educación matemática, a pesar de que desde este campo se ha dedicado escasa atención a la correlación.

REPRESENTACIONES DE LA CORRELACIÓN Y PROCESOS DE TRADUCCIÓN ENTRE ELLAS

Nuestro estudio viene sugerido, asimismo, por el trabajo de Janvier (1978, 1987) en el campo de las funciones y, por tanto, complementa sus resultados para el caso de la dependencia aleatoria. Dicho autor está interesado por las traducciones entre los diversos lenguajes de representación de las funciones, entendiéndolo por «un proceso de traducción, el proceso psicológico involucrado al cambiar de un modo de representación a otro, por ejemplo, de una ecuación a un gráfico» (Janvier, 1987, p. 27). Una de las características del saber matemático es la diversidad de representaciones de un mismo concepto, y

que, en general, ninguna de ellas abarca completamente sus diferentes matices. Kaput (1987) identifica las cinco entidades siguientes en la representación: el mundo representado, el mundo representante, los aspectos del mundo representante que lleva a cabo la representación, los aspectos del mundo representado que se representan, y la correspondencia entre los dos mundos.

Para Duval (1993), los objetos matemáticos nunca deben ser confundidos con la representación que de ellos se hace, pues esto implicaría una pérdida en la comprensión. Pero, si el objeto representado es el que importa, las representaciones son indispensables, ya que los objetos no son directamente accesibles a la percepción. La conceptualización significa, para este autor, la coordinación de diversos registros de representación. Estos registros son complementarios en el sentido de que cada uno de ellos supone una selección de los elementos significativos del contenido que representan. La existencia de varios registros permite la economía en el tratamiento y trabajo matemático. Las transformaciones de un registro de representación a otro no son, a veces, sencillas. Es más, algunos estudiantes ven las diversas representaciones de un concepto como entidades diferentes que aluden, por contra, a conceptos distintos, lo cual dificulta, notablemente, los procesos de traducción (Sanz, 1990).

Godino y Batanero (1994) señalan tres componentes diferenciados en el significado de un concepto matemático, siendo el sistema de representaciones uno de estos tres componentes, a los que se añadirían el campo de problemas de donde surge el concepto y el conjunto de propiedades intensionales asociados al mismo. Estos tres tipos de elementos están relacionados y, precisamente, las dificultades de comprensión de la asociación surgen, a veces, porque los alumnos no son capaces de poner en relación los diferentes elementos del significado del concepto (Batanero, Estepa y Godino, 1998). De ello se deduce la importancia del estudio de la comprensión de las diferentes representaciones y del cambio de representaciones por parte de los estudiantes.

METODOLOGÍA DE LA INVESTIGACIÓN

Población y muestra

Nuestro objetivo era caracterizar el conocimiento de los alumnos al finalizar la enseñanza sobre correlación en los primeros cursos universitarios. La población objetivo estuvo constituida por los estudiantes de las carreras de la diplomatura en Empresariales y la diplomatura en Enfermería de la Universidad de Jaén. La muestra de alumnos que cumplimentó el cuestionario estaba compuesta por 193 estudiantes de estas carreras, 104 (37 hombres y 67 mujeres) de la diplomatura en empresariales y 89 (20 hombres y 69 mujeres) de la diplomatura en enfermería. Por tratarse del primer curso universitario de estas diplomaturas, la edad media de la muestra era de 20 años, teniendo el 87,9 % de la muestra 22 años o menos.

La forma de acceso a la universidad de esta muestra de alumnos ha sido la siguiente: 46 alumnos (24'0 %) de la opción A de COU, científico-tecnológica (ciencias puras); 42 alumnos (21'9 %) de la opción C de COU, ciencias sociales (letras mixtas); 41 alumnos (21'4 %) de la opción B de COU, biosanitaria (ciencias mixtas); 21 alumnos (10'9 %) de formación profesional II, administrativa y comercial; 23 alumnos (12'0 %) de las distintas especialidades de formación profesional II rama sanitaria; 14 alumnos (7'2 %) de las diversas especialidades del bachillerato de la reforma (LOGSE); 5 alumnos (2'6 %) de procedencia distinta a las anteriores, provienen de otras carreras o del acceso para mayores de 25 años.

En cuanto a los estudios de estadística en cursos anteriores nos hemos encontrando que 117 alumnos (60'6 %) no habían estudiado estadística en cursos anteriores; 32 alumnos (16'6 %) la habían estudiado en alguna de las especialidades de formación profesional; 16 alumnos (8'3 %) en COU; 14 alumnos (7'2 %) en bachillerato; 14 alumnos (7'2 %) en otros curso, como el curso anterior –eran repetidores– o bien en otros estudios cursados.

Descripción del cuestionario

En este trabajo consideraremos cuatro formas de representar la correlación entre dos variables cuantitativas: *a) la descripción verbal*, cuando describimos una distribución bivariada mediante el lenguaje natural; *b) tabla de valores*, o presentación de un conjunto de pares de valores numéricos de una distribución bivariada; *c) diagrama de dispersión*, cuando un conjunto de pares de valores de una distribución bivariada se presentan mediante un diagrama cartesiano, y *d) coeficiente de correlación*, cuando se da el coeficiente de correlación existente entre dos modalidades de una distribución bivariada. Estudiaremos el paso de una representación a otra a partir de las respuestas escritas en un cuestionario que se presenta como anexo y que consiste en seis tareas diferentes, cada una con cinco apartados, que denominaremos *subtareas*.

En esta sección analizaremos los diferentes procesos de traducción que han sido incluidos en el instrumento de evaluación. Hacemos notar que la investigación de Sánchez Cobo (1999) muestra que este tipo de tareas apenas se contemplan en la enseñanza universitaria, pues se supone una habilidad ya adquirida por el alumno.

La tarea 1 del cuestionario, que se presenta en apéndice, consiste en pasar de la descripción verbal a un diagrama de dispersión coherente con la misma. Esta prueba no la hemos encontrado, con el enunciado presente, en la literatura de investigación. Para resolver la tarea, el alumno debe valorar, en primer lugar, el tipo de correlación existente y, con posterioridad, dibujar un diagrama de dispersión aproximado, o podría traducir, primero, de descripción verbal a tabla y realizar, a continuación, la representación gráfica. Otra posibilidad es hacer un diagrama de dispersión aproximado sin una verdadera estimación de los valores particulares de los puntos.

La tarea 2 consiste en pasar de la descripción verbal a un valor aproximado del coeficiente de correlación y ha sido tomada de Jennings, Amabile y Ross (1982), quienes la utilizaron para estudiar la precisión de la estimación del coeficiente de correlación a partir de la descripción de pares de variables y a partir de datos representados en forma numérica (pares de puntos). Estos autores suponían que la descripción verbal activa las teorías previas y esto puede influir en la estimación de la correlación. Hemos modificado algunas de las situaciones empleadas por los investigadores citados, para obtener contextos más próximos al alumnado español. Los valores correspondientes del coeficiente de correlación se han obtenido de conjuntos de datos reales sobre las variables presentadas aportados en la investigación de Jennings y sus colaboradores (1982) o de anuarios estadísticos.

La tarea 3 consiste en pasar de la tabla de valores al coeficiente de correlación; incluye ítems usados en la investigación de Estepa (1994), y otros de Jennings, Amabile y Ross (1982). Como en esta situación se parte de un marco numérico, analizaremos si los estudiantes tienen en cuenta la tendencia de la variación conjunta para determinar el signo de la correlación y qué métodos emplean para dar una estimación de la intensidad de la misma. Nuevamente deseamos estudiar el peso que las teorías previas tienen en los juicios de asociación. Es por ello que, en concreto, en los apartados *b* y *c*, se han planteado escenarios en los cuales se da correlación ilusoria (Chapman y Chapman, 1969; Hamilton y Rose, 1980; Tversky y Kahneman, 1982; Murphy y Medin, 1985; Jong, Merckelbach y Arntz, 1995; Tomarken, Sutton y Mineka, 1995), pues los datos contradicen las creencias de los alumnos.

La tarea 4 consiste en pasar del diagrama de dispersión al coeficiente de correlación, siendo una modificación respecto a los utilizados en la investigación de Estepa (1994). Consideramos *a priori* que los métodos empleados abarcarían un espectro que va desde el puramente intuitivo, pasando por la valoración de la forma del diagrama de dispersión y llegando al ajuste más o menos intuitivo de las rectas de regresión.

La tarea 5, con el enunciado actual, no la hemos encontrado en la literatura de investigación. A partir del valor numérico del coeficiente de correlación, los sujetos deben indicar una pareja de variables cuya covariación conjunta se adapte, convenientemente, al mismo, lo cual nos será útil para analizar la interpretación, que hacen los alumnos, del coeficiente de correlación. Una cuestión que se estudiará es si el signo de la dependencia ofrecida por las variables que el alumno asocia coincide o no con el del coeficiente de correlación dado.

Tampoco hemos encontrado dentro de la literatura de investigación la tarea 6, en la que se solicita a los alumnos que, a partir del valor numérico del coeficiente de correlación, dibuje un diagrama de dispersión que se ajuste, razonablemente, a dicho parámetro estadístico.

De modo general, el cuestionario que hemos construido es una prueba bastante completa, pues incluye tareas de

interpretación, estimación y trazado y considera además otras variables relevantes que describimos a continuación.

Otras variables tenidas en cuenta

Cuando los alumnos encuentran una asociación importante entre dos variables, tenderán a pensar que una de ellas provoca o es causa de la variación de la otra. Por todo ello, hemos tenido en cuenta los *tipos de covariación* descritos por Barbancho (1973):

– *Dependencia causal unilateral*: Cuando la ocurrencia de una variable *X* (*causa*) influye en la ocurrencia de *Y* (*efecto*), pero no al contrario, como en la subtareas t2d.

– *Interdependencia*: Cuando la ocurrencia de una variable *X* influye en la ocurrencia de una variable *Y*, y viceversa; como, por ejemplo, en la subtarea t1a.

– *Dependencia indirecta*: Dos variables pueden mostrar cierta covariación debido a la variación de una tercera variable que está correlacionada con ambas, produciendo una asociación aparente, como en la subtarea t2e.

– *Concordancia*: Correlación producida por la ordenación de un conjunto de datos por dos personas de forma independiente; por ejemplo, la subtarea t4b.

– *Covariación casual*: Cuando parece que en la covariación de dos variables hay cierta sincronía, lo cual podría interpretarse como la existencia de asociación entre ambas; sin embargo, ésta es casual o accidental, como en la subtarea t1e.

Otras variables tenidas en cuenta han sido la *intensidad* y el *signo* de la correlación, si la relación es o no *lineal* y si las *teorías previas* sobre las variables dadas por el contexto coinciden o no con la correlación dada en los datos del problema. La combinación específica de estas diversas variables en cada ítem del cuestionario se recoge en la tabla I.

ANÁLISIS DE RESULTADOS

A continuación, exponemos, detalladamente, el análisis que hemos desarrollado con los resultados obtenidos. Este análisis se distribuye en los siguientes apartados:

- La precisión en la estimación del coeficiente de correlación.
- Análisis de correspondencias entre tareas y estrategias empleadas para estimar la correlación.
- Identificación de situaciones reales en las que se presente un valor dado del coeficiente de correlación.

LA PRECISIÓN DE LA ESTIMACIÓN DEL COEFICIENTE DE CORRELACIÓN

Definición de variable dependiente

Una vez recogidos los datos del cuestionario, se han analizado, en primer lugar, los valores estimados del coeficiente de correlación (tareas 2, 3 y 4) o del coeficiente de correlación obtenido de la nube de puntos que construyen los alumnos (tareas 1 y 6). Al comparar estos valores con el coeficiente de correlación normativo de cada subtarea hemos obtenido el valor absoluto de los errores en la estimación del coeficiente de correlación. Este error, en valor absoluto, será la variable dependiente, siendo analizada en este apartado para estudiar la posible influencia sobre la misma de las variables de tarea incluidas en el cuestionario.

Tanto las estimaciones de los alumnos del coeficiente de correlación como el coeficiente de correlación normativo se han multiplicado por 100, con lo que los datos que se refieren al coeficiente de correlación varían entre -100 y +100; en consecuencia, los errores absolutos varían entre 0 y +100. Como resumen de esta variable, en la tabla I se muestra la media correspondiente a cada subtarea y las variables de tarea de cada una de ellas.

Efecto de la intensidad de la correlación y el tipo de tarea

A partir de esta variable dependiente se llevó a cabo un análisis de varianza de medidas repetidas, utilizando el paquete SPSS, tomando como factores la intensidad de la correlación en la tarea propuesta y el tipo de tarea, ya que el diseño del cuestionario permite un diseño factorial completo para estas dos variables.

De este análisis obtuvimos un efecto significativo tanto respecto a la intensidad de correlación ($F = 16,417$; $p = 0,0009$) como al tipo de tarea ($F = 3,122$, $p = 0,015$). Estos resultados se deducen, asimismo de las tablas II y III, en las que presentamos los valores medios del error para cada nivel de estos dos factores, así como el error típico. Podemos observar que las tareas que dan menor valor absoluto de error son la 4 (estimación del coeficiente de correlación a partir de una nube de puntos) y su inversa, la tarea 6 (construir una nube de puntos a partir del coeficiente de correlación). Respecto a los intervalos de intensidad se puede observar que, conforme la intensidad de la asociación crece en valor absoluto, el error absoluto en la estimación del coeficiente de correlación disminuye, lo que corrobora lo expresado en otros análisis previos y en investigaciones anteriores como la de Crocker (1981).

Efecto del tipo de covariación

Asimismo se llevó a cabo un análisis de varianzas de medidas repetidas respecto al tipo de covariación en la tarea propuesta y el tipo de tarea, ya que el diseño del cuestionario permite un diseño factorial completo para

INVESTIGACIÓN DIDÁCTICA

Tabla I
Media y error típico del error absoluto en la estimación y variables de tarea en los diferentes ítems del cuestionario.

Ítem	Media	Error típico	Intensidad (valor absoluto)					Tipo de ajuste		Tipo de covariación				Teorías previas		Dependencia		
			0 - 0,1	0,1 - 0,35	0,35 - 0,65	0,65 - 0,9	0,9 - 1	Lineal	No lineal	Dep. causal Unila.	Interdependencia	Dep. Indirecta	Concordancia	Espúrea	Coincide	No coincide	+	-
1A	14,2	1,23				X			X		X							X
1B	10,12	2,01					X		X	X							X	
1C	34,60	1,34			X			X				X					X	
1D	57,09	1,88		X				X				X						X
1E	42,19	3,63	X						X				X					
2A	23,42	2,93					X	X			X						X	
2B	31,15	8,27		X				X					X					X
2C	21,58	5,33	X						X					X			X	
2D	37,47	1,19			X				X	X							X	
2E	28,21	2,63				X		X				X						X
3A	36,37	3,3					X		X	X					X			X
3B	39,71	4,66	X						X		X					X	X	
3C	32,31	2,86			X			X					X			X	X	
3D	19,39	2,0				X		X				X			X		X	
3E	31,17	3,92		X				X				X			X		X	
4A	30,85	4,37	X					X			X					X	X	
4B	19,93	1,52			X				X			X			X		X	
4C	12,83	1,55					X	X		X						X	X	
4D	24,6	2,46		X				X					X	X				X
4E	37,71	7,63				X			X			X			X			X
6A	33,24	2,19		X													X	
6B	14,54	2,58					X											X
6C	39,36	2,82	X															X
6D	19,17	1,14				X											X	
6E	26,84	1,65			X													X

Tabla II
Media y error típico del error absoluto en la estimación, en función del tipo de tarea.

Tarea	1	2	3	4	6
Media	31,64	28,37	31,79	25,19	26,63
Error típico	0,92	2,65	2,26	2,20	1,16

Tabla III
Media y error típico del error absoluto en la estimación, en función de la intensidad de la correlación.

Intensidad	0-0,1	0,1-0,35	0,35-0,65	0,65- 0,9	0,9-1
Media	34,74	35,45	30,23	23,74	19,46
Error típico	2,64	2,77	0,9	1,67	1,24

estas dos variables. De nuevo observamos un efecto significativo de la tarea y respecto al tipo de covariación ($F = 6,297$, $p = 0,009$). En la tabla IV se presentan los valores medios correspondientes.

Los tipos de covariación en los que el error absoluto en la estimación del coeficiente de correlación es más bajo son la dependencia causal unilateral e interdependencia, que son las que con más frecuencia aparecen en ejercicios y ejemplos en los libros de texto (Sánchez Cobo 1996; Estepa y Sánchez Cobo, 1998).

Efecto del tipo de ajuste, teorías previas y tipo de dependencia

Para analizar el efecto del tipo de ajuste sobre el error de estimación del coeficiente de correlación, se tomó como variable dependiente el error medio cometido por cada alumno en el conjunto de tareas según tipo (lineal o no lineal). El valor medio del error absoluto en la estimación del coeficiente de correlación, cuando el ajuste es lineal, es 31,52, con un error típico de 1,39. Cuando el ajuste es no lineal, la media del error es 28,98, con error típico de 1,11. También se ha obtenido un coeficiente de correlación entre las dos estimaciones de 0,53, siendo significativo, lo que indica que las estimaciones de los alumnos en los dos tipos de tareas están correlacionadas y esto nos autoriza a utilizar la prueba T para muestras relacionadas. Se obtiene una diferencia muy pequeña en valor absoluto, aunque significativa, debido al tamaño de la muestra ($p = 0,042$). Para efectos prácticos podemos aceptar la igualdad de las dos medias, lo que significa que no existe diferencia entre los errores absolutos de la estimación del coeficiente de correlación cuando la relación presentada en la tarea es lineal o no lineal.

El mismo tipo de análisis se repitió para comparar los errores en las estimaciones del coeficiente de correlación

de los alumnos en las tareas en que los datos coinciden o no coinciden con sus teorías previas, obteniéndose una media de 29,09 con un error típico de 1,57 para cuando las teorías previas están a favor; cuando las teorías previas están en contra se obtiene una media de 29,07, con un error típico de 1,91. Entre las dos existe un coeficiente de correlación significativo de 0,57, por lo que realizamos una comparación de muestras relacionadas mediante la prueba T cuyos resultados fueron no significativos; por tanto, no podemos decir que haya diferencia en el error de estimación del coeficiente de correlación según haya teorías previas a favor o en contra. Sin embargo, hay que tener en cuenta que esta variable de tarea solamente se ha podido aplicar en las tareas 3 y 4.

En cuanto al tipo de dependencia, el error medio de estimación, para la dependencia directa, fue 26,43, con error típico de 1,15 y, para la inversa, de 31,7 con error típico de 1,08. Además hemos obtenido un coeficiente de correlación de 0,62. También hemos realizado una comparación de muestras relacionadas aplicando la prueba T para muestras, obteniéndose un valor $t = -5,377$, significativo ($p = 0,0009$), lo que sugiere que el valor absoluto del error es más pequeño cuando se trata de una dependencia directa.

ANÁLISIS DE CORRESPONDENCIAS ENTRE TAREAS Y ESTRATEGIAS EMPLEADAS PARA ESTIMAR LA CORRELACIÓN

Descripción de las estrategias

Un segundo análisis efectuado fue el estudio de las estrategias que los alumnos emplean para resolver las tareas propuestas, las cuales han sido clasificadas en las siguientes categorías:

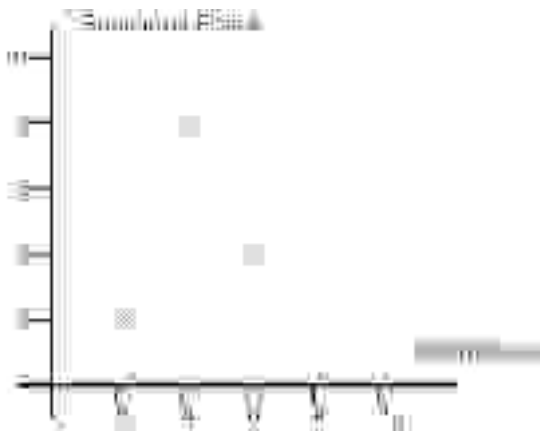
Tabla IV
Media y error típico del error absoluto en la estimación, en función del tipo de covariación.

Tipo de covariación	Dependencia causal unilateral	Interdependencia	Concordancia	Dependencia casual	Dependencia indirecta
Media	24,38	26,88	29,29	30,38	35,93
Error típico	1,10	2,03	2,8	2,35	1,83

- *E1.* Cuando la estrategia no se explicita. Las siguientes respuestas son ejemplos de esta estrategia: «Varía con respecto al número de trabajadores.» (sujeto 26, tarea 1a); «La relación entre pintores y tiempo es inversa.» (sujeto 147, tarea 1a); o «Existe dependencia entre las calificaciones de uno y otro profesora.» (sujeto 31, tarea 1c).

- *E2.* Los alumnos hacen referencia a alguna característica del diagrama de dispersión o lo dibujan. Así, el sujeto 2 indica, al responder a la tarea 1, apartado *a*, que es una «función decreciente», y el sujeto 185 responde, a la tarea 1, apartado *e*, diciendo que «no es lineal». El sujeto 6 responde a la tarea 3, apartado *c*, dibujando una nube de puntos, la cual se muestra en la figura 1, e indicando « $r = 0,5$ ».

Figura 1



- *E3.* Los alumnos se apoyan en algún hecho del marco numérico implícito; por ejemplo, el sujeto 16 responde, a la tarea 3; apartado *d*, indicando: « $r = 0,8$. Tiene relación una variable con otra, a medida que aumenta una variable, la otra aumenta con un incremento cada vez menor.» El sujeto 113, al contestar a la tarea 1, apartado *a*, indica: «Cuantos más pintores haya, menos tiempo tardarán en pintar la habitación.» El sujeto 28 dice, tarea 4, apartado *a*: « $r = 0,2$. A mayor número de

empates, menor clasificación». Este alumno, a pesar de que manifiesta que la dependencia es inversa, valora el coeficiente de correlación con un signo positivo.

- *E4.* Los alumnos utilizan el marco numérico y gráfico en forma conjunta. Así, por ejemplo, el sujeto 11, al contestar a la tarea 1, apartado *e*, manifiesta: «Lineal negativa. Cuanto más edad le corresponde al alumno, menor serán las calificaciones y, a menor edad, mayor serán las calificaciones.» El sujeto 185, al responder a la tarea 1, apartado *d*, dice: «Cuanto más latitud haya, menor temperatura habrá (↑ latitud ↓ temperatura, relación indirecta decreciente).» Parece como si los sujetos de la muestra no son sensibles a que los hechos referidos al marco numérico implícito —«cuanto más latitud haya menor temperatura habrá»— son equivalentes a los referidos al marco gráfico— «relación indirecta decreciente»— y, por lo tanto, lo que ofrecen es información redundante.

- *E5.* Los alumnos sustentan sus argumentos en las teorías previas que poseen sobre el contexto del problema planteado. Así, el sujeto 20 responde, a la tarea 1, apartado *b*, de la siguiente manera: «El tiempo depende de la memoria de cada persona y de la concentración que cada uno ponga. Todos no tenemos la memoria desarrollada igual; por lo tanto, el tiempo es independiente del número de datos.» El sujeto 85 responde, a la tarea 2 apartado *d*, en los siguientes términos: « $r = -0,5$. Los estudiantes hacen poco ejercicio y, por tanto, en una prueba de rendimiento físico no acabarían en buenas posiciones a no ser que haya alguno que pertenezca a algún club deportivo.» El sujeto 10 responde a la tarea 4 apartado *e*: « $r = -0,8$. Inversa a más natalidad, podrán comer menos cantidad de proteínas, esto es lo que pasa en los países tercermundistas».

- *OE.* Los alumnos usan otros conceptos, como la *proporcionalidad* o la *dependencia funcional*. Éste es el caso del sujeto 39, quien, al responder a la tarea 1, apartado *a*, dice: «Estas variables tienen relación funcional, por tanto, $r = 1$.» El sujeto 37, al responder a la tarea 1 apartado *b*, indica: «Si para [1 sola] persona aprender 10 números le cuesta 2 minutos, para aprender 50 le costará 10 minutos.» O el sujeto 94 contesta a la misma pregunta exponiendo: «En memorizar la lista máxima se tarda el tiempo máximo y se comprueba que hay una relación entre las variables de 5.» Ambos alumnos exhiben una concepción proporcional de la asociación. En

particular, este último concluye que la constante de proporcionalidad sería 5.

Para cada una de las subtareas de las tareas 1 a 4, ambas inclusive, se han clasificado las respuestas de los alumnos según estas estrategias, obteniéndose la tabla V, de la que se ha tomado únicamente en consideración las respuestas claras. Las frecuencias esperadas de esta tabla de contingencia verifican las condiciones de aplicación del estadístico c^2 , ya que ninguna de ellas es menor que 1, siendo menos del 20 % de ellas menores que 5.

Para analizar las posibles relaciones existentes entre las estrategias utilizadas y las diferentes tareas propuestas, hemos utilizado un análisis de correspondencias de esta tabla, que permite visualizar las relaciones entre las filas y las columnas de una tabla de contingencia (Nishisato, 1980; Greenacre, 1984; Greenacre y Hastie, 1987; Lacasta y Brousseau, 1995). Éste es uno de los métodos multivariantes factoriales que, a partir de la distancia c^2 definida sobre las dos nubes de puntos (fila y columna) que se obtienen mediante las frecuencias condicionales respecto a filas y columnas en la tabla de contingencia permite determinar la dimensión del espacio vectorial definido por dichas distribuciones de puntos.

Tabla V
Frecuencias absolutas de las estrategias observadas en las tareas 1, 2, 3 y 4.

Tareas	Estrategias						Total
	E1	E2	E3	E4	E5	OE	
t1a	2	3	60	3	5	13	86
t1b	2	5	52	13	4	7	83
t1c	7	1	3	2	16	49	78
t1d	2	6	7	45	9	4	73
t1e	6	2	2	45	16	12	83
t2a	4	22	0	9	0	25	60
t2b	11	43	0	0	1	20	75
t2c	6	59	0	0	0	6	71
t2d	4	3	0	36	0	30	73
t2e	2	16	2	3	12	25	60
t3a	7	4	4	5	37	5	62
t3b	8	4	2	3	0	18	35
t3c	29	4	5	2	0	32	72
t3d	14	7	2	5	0	25	53
t3e	21	2	1	1	0	40	65
t4a	7	4	3	15	0	1	30
t4b	1	0	20	1	0	0	22
t4c	7	0	23	0	0	1	31
t4d	5	18	2	1	0	5	31
t4e	2	4	3	0	18	0	27
Total	147	207	191	189	118	318	1170

Al aplicar el análisis de correspondencias, obtuvimos un valor de c^2 igual 1796,37, con 96 gl, que es estadísticamente significativo (p -valor = 0,000); en consecuencia, hay asociación entre las filas y las columnas, es decir, entre las tareas y la estrategia utilizada para dar la respuesta. Los tres primeros autovalores explican el 34,4 %, 25,6 % y 18,5 % de la inercia (variabilidad en la tabla), respectivamente, por lo que deducimos que nos encontramos ante un fenómeno no unidimensional, ya que no existe un solo autovalor que acapare toda la inercia; la diferencia en porcentaje de explicación de la inercia entre los dos primeros tampoco es muy grande. A continuación interpretamos los dos primeros factores.

Para facilitar la interpretación, se ha creído conveniente introducir, como filas suplementarias, las variables de tarea que se incluyeron en el diseño. Estas filas suplementarias no intervienen en la determinación de los ejes factoriales. Sin embargo, una vez delimitados éstos, se pueden proyectar sobre ellos para facilitar la interpretación. Debido a la propiedad baricéntrica, cada fila suplementaria se proyecta sobre la «media» de todas las columnas que la definen; es decir, podemos interpretar estas filas como el comportamiento promedio de las tareas caracterizadas por un cierto valor de una de las variables de tarea. Por ello, podremos relacionar los valores de las variables de tarea con las estrategias utilizadas por los alumnos.

Factor 1: Oposición entre razonamiento numérico y gráfico

El primer factor, que explicó el 34,4 % de la inercia, opone algunas subtareas de las tareas 1 y 4. En concreto opone las tareas t1a ($x = -1,346, r = 0,856$); t1b ($x = -1,214, r = 0,903$); t4b ($x = -1,882, r = 0,866$); t4c ($x = -1,442, r = 0,748$), a las siguientes: [t2a ($x = 0,607, r = 0,642$); t2b ($x = 0,800, r = 0,495$); t2c ($x = 0,965, r = 0,312$); t2e ($x = 0,398, r = 0,338$)]. Respecto a las estrategias, en la parte negativa del eje aparece la estrategia 3 ($x = -0,516, r = 0,885$) asociada a las subtareas t1a, t1b, t4b, t4c, mientras que, en la parte positiva, aparece la estrategia 2 ($x = 0,779, r = 0,348$), asociada a algunas subtareas de la tarea 2 (t2a, t2b, t2c, t2e). Esta diferenciación respecto a estas dos estrategias, en las que sólo se usa uno de los marcos numérico o gráfico, se puede observar en la tabla V.

Así, para construir un diagrama de dispersión a partir de la descripción verbal o para estimar el coeficiente de correlación desde una nube de puntos –tareas 1 y 4 respectivamente–, los alumnos han usado preferentemente razonamientos numéricos, examinando la variabilidad de ambas características y comparando los valores numéricos hipotéticos de las mismas. Por el contrario, para estimar el coeficiente de correlación de dos variables presentadas mediante su descripción verbal (tarea 2), los alumnos necesitan apoyarse en alguna característica del marco gráfico. De este modo, parece que los alumnos, con sus argumentos, quieren complementar la información dada por el problema o bien por su solución.

Tabla VI
Frecuencia y porcentaje*, respecto al total de soluciones, de las características de las soluciones aportadas en la tarea 5.

	T5a	T5b	T5c	T5d	T5e
V. bidimensional consistente	134 91,2	93 83,8	94 75,8	106 82,8	99 81,8
Coincidencia del signo	125 85,0	46 41,4	84 67,7	69 53,9	77 63,6
Dependencia funcional	30 20,4	5 4,5	6 4,8	9 7,0	6 5,0
Dependencia aleatoria	107 72,8	79 71,8	46 37,1	92 71,9	87 72,5
Independencia	10 6,8	26 23,6	72 58,1	27 21,1	27 22,5
Total	147	110	124	128	120

* Como las características de las soluciones no son excluyentes, la suma de los porcentajes en filas y en columnas no suele coincidir con 100.

En cuanto a las variables de tarea, podemos observar asociación con la estrategia E3 (parte negativa del eje) de la correlación positiva ($x = -0,450$, $r = 0,723$) y el tipo de ajuste no lineal ($x = -0,589$, $r = 0,891$), dependencia causal unilateral (dcausal $x = -0,630$, $r = 0,764$) interdependencia (interd $x = -0,533$, $r = 0,510$) y fuerte intensidad de la correlación, pues los intervalos de más alta intensidad de la correlación aparecen asociados a la parte negativa del eje I4 ($x = -0,460$, $r = 0,547$), I5 ($x = 0,337$, $r = 0,632$). En resumen, los argumentos numéricos de los alumnos han sido usados preferentemente en ítems con correlaciones intensas y positivas, en caso de dependencia causal e interdependencia y ajuste no lineal.

Factor 2: Teorías previas

El segundo factor, que explica un 25,6 % de la variabilidad, opone las subtareas t2b ($x = 0,728$, $r = 0,419$), t2c ($x = 1,042$, $r = 0,364$), t4d ($x = 0,762$, $r = 0,469$) a las t1d ($x = -0,946$, $r = 0,554$), y t1e ($x = -1,051$, $r = 0,840$), en la parte negativa. Se caracteriza por el uso de la estrategia E2 ($x = 0,768$, $r = 0,338$) en la parte positiva del eje, que se asocia a las subtareas t2b, t2c y t4d, frente a las estrategias E4 ($x = -0,975$, $r = 0,560$) y E5 ($x = -0,910$, $r = 0,302$), que aparecen en la parte negativa, y se asocian, por tanto, a las subtareas t1d y t1e y a las subtareas t3a y t4e, siendo en éstas donde, con mayor frecuencia, los alumnos se apoyan en sus teorías previas con el fin de reforzar sus argumentaciones.

En cuanto a las variables de tarea, las correlaciones con éstas son muy bajas. Hay, sin embargo, una pequeña asociación de la dependencia indirecta y concordancia con la parte negativa del eje, que es el tipo de covariación que estas subtareas presentan en el diseño del experimento. Coinciden estos resultados con nuestros trabajos previos, ya que, en los casos de concordancia y dependencia indirecta, se produjeron con mayor frecuencia la influencia de las teorías previas.

IDENTIFICACIÓN DE SITUACIONES REALES EN LAS QUE SE PRESENTE UN VALOR DADO DEL COEFICIENTE DE CORRELACIÓN

Se requiere un análisis diferente para las actividades de traducción de un valor del coeficiente de correlación a la representación de la descripción verbal –variable estadística bidimensional–, es decir, las correspondientes a la tarea 5. Para ello hemos tomado en consideración si el alumno describe una variable estadística bidimensional consistente, si el signo de la correlación de la variable bidimensional sugerida por el alumno coincide o no con el signo de la tarea y si su dependencia es funcional, aleatoria o no hay tal relación.

Como puede advertirse en la tabla VI en todas las subtareas de la tarea 5, los alumnos han respondido, de forma mayoritaria, con una variable estadística bidi-

mensional pertinente. En cuanto al signo, excepto en la tarea *b* (coeficiente de correlación $-0,3$), en más de la mitad de las soluciones, coincide la variable estadística bidimensional propuesta por el alumno con el correspondiente del coeficiente de correlación dado en la tarea, aumentando el porcentaje de aciertos en el caso de una relación positiva. Finalmente, si examinamos el tipo de dependencia, vemos que la dependencia funcional, que es la que se presenta en la subtarea *a*, sólo ha sido detectada por el 20,4 % de estudiantes y la independencia, que es la correspondiente a la subtarea *c*, por el 58,1%. Cuando la dependencia es aleatoria –subtareas *b*, *d* y *e*–, las variables estadísticas bidimensionales dadas por los alumnos concuerdan, en un alto porcentaje –por encima del 70 % en los tres casos–, con lo solicitado en la subtarea.

CONCLUSIONES

En este trabajo hemos analizado la capacidad de traducir diferentes representaciones de la correlación en alumnos universitarios una vez terminada la enseñanza del tema. Nuestros análisis muestran que este tipo de tareas no siempre resulta sencillo y que no todos los estudiantes actúan de modo similar al enfrentarse a esta clase de problemas.

En general, los alumnos han mostrado una buena capacidad de estimación de la correlación, ya que la media global de la diferencia entre los valores absolutos de los coeficientes de correlación normativos y los estimados por los alumnos ha sido 28,73 (en una escala 0-100). Esta precisión en la estimación ha dependido del tipo de tarea, la intensidad de la correlación, el tipo de covariación, el tipo de dependencia, aunque no del tipo de ajuste y de la existencia de teorías previas.

La estimación es más precisa al estimar el coeficiente de correlación a partir de un diagrama de dispersión y la tarea inversa, debido seguramente, a que los diagramas de dispersión son una de las representaciones que los alumnos han usado con frecuencia en la enseñanza. Los errores son mayores al construir una nube de puntos a partir de una descripción verbal y estimar el coeficiente

de correlación desde una tabla de valores numéricos, tareas que no han sido familiares a los alumnos.

La precisión de las estimaciones se hace mayor cuando la correlación es más intensa. Esto sucede, igualmente, tanto en los casos que exhiben una dependencia causal unilateral como en los que presentan una dependencia positiva. Los resultados obtenidos confirman y extienden los de investigaciones anteriores, al haberse considerado, en nuestro diseño, un mayor número de variables de tarea que, además, han sido controladas mediante el diseño experimental. También ampliamos los estudios previos mediante el análisis de las estrategias y la determinación de factores en las mismas a partir del análisis de correspondencias.

En cuanto a la capacidad de proponer situaciones factibles a un coeficiente de correlación dado, los alumnos, en su mayoría, proponen variables bidimensionales consistentes. El 63,5 % de los alumnos propone variables cuya correlación tiene el mismo signo que el dado en el problema y se observan dificultades en la identificación de la dependencia funcional y, en menor medida, de la independencia.

Todos estos resultados ponen de manifiesto la complejidad del concepto de *correlación* y de la comprensión del coeficiente de correlación, incluso al finalizar la enseñanza. Puesto que este tema se sugiere en los nuevos currículos de secundaria, sería preciso reflexionar sobre las situaciones didácticas adecuadas para estos alumnos. Probablemente las tareas que hemos propuesto en nuestro cuestionario podrían emplearse para diseñar tareas significativas encaminadas a lograr la comprensión gradual de la correlación y el coeficiente de correlación.

NOTA

* Esta investigación está realizada en el marco de los Proyectos de Investigación PB97-0851 y PB96-1411, subvencionados por la Secretaría de Estado de Universidades, Investigación y Desarrollo (SEUID)(I+D).

REFERENCIAS BIBLIOGRÁFICAS

- BARBANCHO, A.G. (1973). *Estadística elemental moderna*. Barcelona: Ariel.
- BATANERO, C., ESTEPA, A. y GODINO, J.D. (1997). Evolution of students' understanding of statistical association in a

computer based teaching environment, en Garfield, J.B. y Burrill, G. (eds.). *Research on the Role of Technology in Teaching and Learning Statistics. IASE Round Table Conference Papers*, pp. 191-205. Voorburg: International Statistical Institute.

- BATANERO, C., ESTEPA, A. y GODINO, J.D. (1998). La construcción del significado de la asociación mediante actividades de análisis de datos: reflexiones sobre el papel del ordenador en la enseñanza de la estadística. *II Seminario de la Sociedad Española en Educación Matemática*. Pamplona.
- 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*, 2, pp. 1017-1024. Voorburg: International Statistical Institute.
- BATANERO, C., GODINO, J.D. y ESTEPA, A. (1998). Building the meaning of association through data analysis activities. Research Forum. *22 Conference of the International Group for the Psychology of Mathematics Education*. Stellembosch, Sudáfrica.
- BEYTH-MAROM, R. (1982). Perception of correlation reexamined. *Memory and Cognition*, 10(6), pp. 511-519.
- CROCKER, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin*, 90(2), pp. 272-292.
- CHAPMAN, L.J. y CHAPMAN, J.P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 74, pp. 271-280.
- DUVAL, R. (1993). *Semiosis et Noesis. Lecturas en Didáctica de la Matemática: Escuela Francesa*. México: Sección de Matemática Educativa del CINVESTAV-IPN.
- ERLICK, D.E. y MILLS, R.G. (1967). Perceptual quantification of conditional dependency. *Journal of Experimental Psychology*, 73(1), pp. 9-14.
- 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. Departamento de Didáctica de la Matemática. Universidad de Granada.
- ESTEPA, A. y BATANERO, C. (1996). Judgments of correlation in scatterplots: students' intuitive strategies and preconceptions. *Hiroshima Journal of Mathematics Education*, 4, pp. 25-41.
- ESTEPA, A. y SÁNCHEZ COBO, F.T. (1998). Correlation and regression in secondary school text books, en Pereira-Mendoza, L., Seu, L., Wee, T. y Wong, W. (eds.). *Proceedings of the Fifth International Conference on Teaching of Statistics*, 2, pp. 671-676. Voorburg (Netherlands): International Statistical Institute.
- GREENACRE, M.J. (1984). *Theory and applications of correspondence analysis*. Londres: Academic Press.
- GREENACRE, M.J. y HASTIE, T. (1987). The geometric interpretations of correspondence analysis. *Journal of the American Statistical Association*, 82, 398, pp. 437-447.
- 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.
- HAMILTON, D.L. y ROSE, T.R. (1980). Illusory correlation and the maintenance of stereotypic beliefs. *Journal of Personality and Social Psychology*, 39, pp. 832-845.
- INHELDER, B. y PIAGET, J. (1955). *De la logique de l'enfant à la logique de l'adolescent*. París: Presses Universitaires de France (Trad. cast. 1985.) Barcelona: Paidós.
- JANVIER, C. (1978). *The interpretation of complex cartesian graph representing situations, studies and teaching experiments*. Tesis doctoral. Universidad de Québec.
- JANVIER, C. (1987). Procesos de traducción en educación matemática, en Janvier, C. (ed.). *Problems of representation in the teaching and learning of mathematics*, pp. 27-32. Londres: LEA Publ. Trad. de Coriat Benarroch, M.
- JENNINGS, D.L., AMABILE, T.M. y ROSS, L. (1982). Informal covariation assessment: Data-based versus theory-based judgments, en Kahneman, D. Slovic, P. y Tversky, A. (eds.). *Judgment under uncertainty: Heuristics and biases*, pp. 211-230. Nueva York: Cambridge University Press.
- JONG, P.J. DE, MERCKELBACH, H. y ARNTZ, A. (1995). Covariation bias in phobic women: The relationship between a priori expectancy, on line expectancy, autonomic responding and a posteriori contingency judgment. *Journal of Abnormal Psychology*, 104(1), pp. 55-62.
- KAPUT, J.J. (1987). Representation systems and mathematics, en Janvier, C. (ed.). *Problems of representation in the teaching and learning of mathematics*, pp. 19-26. Hillsdale: LEA.
- LACASTA, E. y BROUSSEAU, G. (1995). Utilisation de la contingence par l'analyse factorielle. Traitement d'un cas: Le graphique, en Gras, R. (ed.). *Méthodes d'analyses statistiques multidimensionnelles en didactique des mathématiques*, pp. 53-90. Rennes: ARDM.
- LANE, D.M., ANDERSON, C.A. y KELLAM, K.L. (1985). Judging the relatedness of variables: The psychophysics of covariation detection. *Journal of Experimental Psychology. Perception and Performance*, 11(5), pp. 640-649.
- MORRIS, E.J. (1997). An investigation of students' conceptions and procedural skills in the statistical topic correlation. *Centre for Information Technology in Education*, 230. The Open University.
- MORRIS, E.J. (1998). Link: The principled design of a computer assisted learning program for correlation, en Pereira-Mendoza, L., Seu Kea, L., Wee Kee, T. y Wong, W. (eds.). *Proceedings of the Fifth Conference on Teaching Statistics*, 2, pp. 1033-1040. Voorburg: International Statistical Institute.
- MURPHY, G.L. y MEDIN, D.L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92(3), pp. 289-316.
- NISHISATO, S. (1980). *Analysis of categorical data: Dual scaling and its applications*. Toronto: University of Toronto Press.
- POZO, J.I. (1987). *Aprendizaje de la ciencia y pensamiento causal*. Madrid: Visor.
- SÁNCHEZ COBO, F.T. (1996). *Análisis de la exposición teórica y de los ejercicios de correlación y regresión en los textos de bachillerato*. Memoria de tercer ciclo. Departamento de Didáctica de la Matemática. Universidad de Granada.
- SÁNCHEZ COBO, F.T. (1999). *Significado de la correlación y regresión para los estudiantes universitarios*. Tesis doctoral. Departamento de Didáctica de la Matemática. Universidad de Granada.
- SANZ, I. (1990). Comunicación, lenguaje y matemáticas, en Linares, S. y Sánchez, M.V. (eds.). *Teoría y práctica en educación matemática*, pp. 173-235. Sevilla: Alfar.
- SHAKLEE, H. y TUCKER, D. (1980). A rule analysis: judgments of covariation between events. *Memory and Cognition*, 8(5), pp. 459-467.
- TOMARKEN, A.J., SUTTON, S.K. y MINEKA, S. (1995). Fear-relevant illusory correlations: What types of associations

- promote judgmental bias. *Journal of Abnormal Psychology*, 104(2), pp. 312-326.
- TROLIER, T.K. y HAMILTON, D.L. (1986). Variables influencing judgments of correlational relations. *Journal of Personality and Social Psychology*, 50(5), pp. 879-888.
- TRURAN, J.M. (1995). Some undergraduates' understanding of the meaning of a correlation coefficient, en Atweh, B. y Flavel, S. (eds.). *MERGA 18: Galtha*, pp. 524-529. Proceedings of the Eighteenth Annual Conference of the Mathematics Education Research Group of Australasia (MERGA). Northern Territory University, Darwin, Australia.
- TRURAN, J.M. (1997). Understanding of association and regression by first year economics students from two different countries as revealed in responses to the same examination questions, en Garfield, J.B. y Truran, J.M. (eds.). *Research Papers on Stochastics Educations from 1997*, pp. 205-212. Department Educational Psychology University of Minnesota.
- TVERSKY, A. y KAHNEMAN, D. (1982). Judgment under uncertainty: heuristics and biases, en Kahneman, D. Slovic, P. y Tversky, A. (eds.). *Judgment under uncertainty: Heuristics and biases*, pp. 3-20. Nueva York: Cambridge University Press.

[Artículo recibido en junio de 1999 y aceptado en febrero de 2000.]

ANEXO
CUESTIONARIO

Tarea 1. Dadas las siguientes parejas de variables, dibujad un diagrama de dispersión que contenga 10 puntos y que muestre razonablemente su variación conjunta. (Nota: Se dio al alumno los ejes dibujados con escalas adecuadas a los datos del problema.)

- a) Número de pintores pintando una habitación y tiempo en horas para acabar el trabajo.
- b) Longitud de una lista de números y tiempo empleado por una persona en memorizarla.
- c) Calificaciones de un mismo examen por dos profesores de un tribunal de oposiciones.
- d) La latitud de capitales europeas y la temperatura que hace en ellas un día determinado.
- e) Las calificaciones en estadística y la edad del alumno.

Tarea 2. Dadas las siguientes parejas de variables, indicad un valor razonable para el coeficiente de correlación que expresaría el tipo de relación entre las variables (directa, inversa o independencia) y su intensidad (fuerte o débil).

- a) Altura y envergadura –distancia entre los extremos de los brazos puestos en cruz– de los estudiantes de la diplomatura en empresariales: $r =$
- b) Ordenación por grado de timidez de los estudiantes de la Universidad de Jaén y ordenación por número de ciudades diferentes que han visitado: $r =$
- c) Grado de ambición y estatura de los estudiantes de la Universidad de Jaén: $r =$
- d) Tiempo semanal dedicado por los estudiantes a actividades atléticas y la posición que obtienen en una prueba de rendimiento físico: $r =$
- e) Número de días de lluvia y número de horas de sol registradas durante un año por un observatorio de las diferentes comunidades autónomas: $r =$

Tarea 3. Para cada una de las siguientes tablas de datos, estimar un valor razonable del coeficiente de correlación que muestre el tipo de relación entre las variables (directa, inversa o independencia) y su intensidad (fuerte o débil). (Nota: Se pidió a los alumnos no efectuar cálculos numéricos.)

- a) Tiempo en meses desde que se prepara un medicamento y porcentaje de efectividad para una cierta enfermedad: $r =$

Tiempo en meses	1	2	3	4	5
% de efectividad	90	75	42	30	21

- b) Calificaciones de 10 alumnos de COU en los exámenes de matemáticas y física: $r =$

Matemáticas	1	2	2	3	4	4	5	6	7	7
Física	3	7	6	2	2	7	4	5	3	5

- c) Calificaciones de 10 alumnos de COU en los exámenes de matemáticas y educación física: $r =$

Matemáticas	2	3	4	4	5	6	7	8	9	10
Educación física	2	5	7	8	5	4	6	5	5	9

- d) Valores de la presión sanguínea antes y después de haber efectuado un cierto tratamiento médico a un grupo de 10 mujeres: $r =$

Mujer	Presión sanguínea en cada mujer									
	A	B	C	D	E	F	G	H	I	J
Antes tratamiento	115	112	107	119	115	138	126	105	104	115
Después tratamiento	128	115	106	128	122	145	132	109	102	117

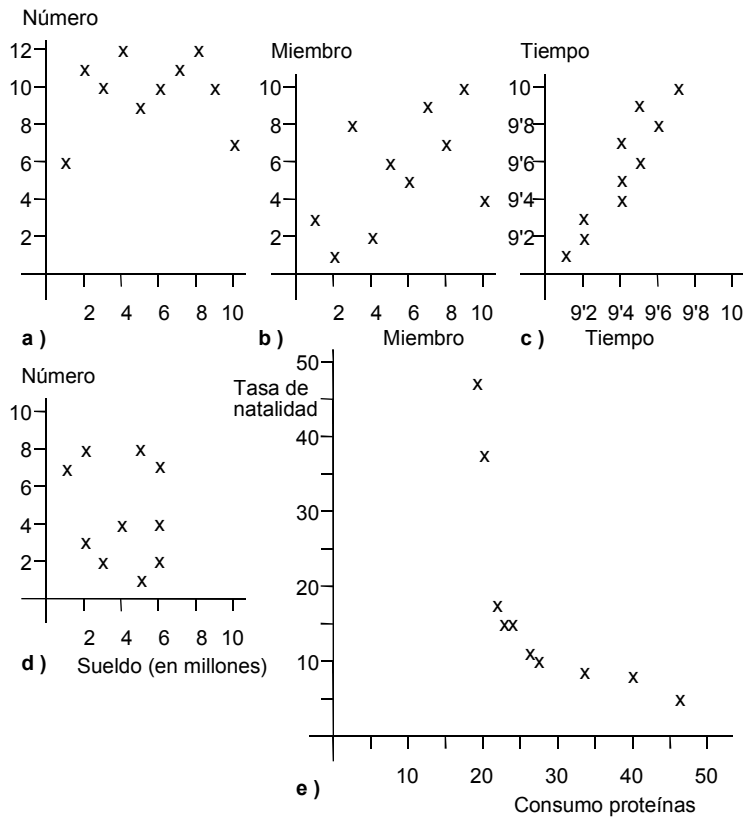
- e) Ordenación dada por dos entrenadores a 10 atletas según su estado físico: $r =$

INVESTIGACIÓN DIDÁCTICA

Atletas	A	B	C	D	E	F	G	H	I	J
Entrenador A	1	2	3	4	5	6	7	8	9	10
Entrenador B	2	8	1	3	9	10	4	5	7	6

Tarea 4. Dadas las siguientes gráficas que representan la variación conjunta de variables, estimad el valor de su coeficiente de correlación teniendo en cuenta el tipo de relación (directa, inversa o independenciana) y la intensidad de la misma. (Nota: Se pidió a los alumnos que no efectuasen cálculos numéricos.)

- a) Puesto ocupado por los 10 primeros equipos de 1a. división de la liga de fútbol en la temporada 1987-88 y los partidos empatados: $r =$
- b) Puntuaciones otorgadas por los miembros A y B de un tribunal a 10 proyectos presentados: $r =$
- c) Tiempo, en segundos, de 10 atletas en correr 100 m lisos en septiembre y diciembre: $r =$
- d) Sueldo, en millones de pesetas, de los empleados de una empresa y número de libros que leen al cabo de un año: $r =$
- e) Tasa de natalidad y consumo diario de proteínas animales en 10 países: $r =$



Tarea 5. Dados los siguientes valores del coeficiente de correlación lineal, describid dos variables para las cuales sea razonable obtener este coeficiente de correlación en función del tipo de dependencia entre las variables (directa o inversa) y la intensidad de las mismas.

- a) $r = 1$ b) $r = -0,3$ c) $r = 0,05$ d) $r = -0,8$ e) $r = 0,5$

Tarea 6. Dados los siguientes valores del coeficiente de correlación entre dos variables X e Y, dibujad un diagrama de dispersión, con 10 puntos, que se adapte razonablemente a ellos. (Nota: Se dio al alumno los ejes dibujados con escalas adecuadas a los datos del problema.)

- a) $r = 0,25$ b) $r = -1$ c) $r = -0,01$ d) $r = 0,7$ e) $r = -0,4$