

## Caracterización del Significado de la Correlación y Regresión en Estudiantes de Educación Secundaria\*

*Antonio Estepa Castro<sup>1</sup>*

**Resumen:** En el presente trabajo se caracteriza el significado que los estudiantes de Educación Secundaria (17 años de edad) dan a la correlación y regresión después de haber recibido instrucción sobre este tema. Se comienza por una revisión de la literatura relacionada, se describe el marco teórico, la muestra y el cuestionario, donde se han propuesto 12 ítems que contienen elementos de significado representativos sobre distribuciones bidimensionales, correlación, covarianza y regresión. Se analizan y discuten las dificultades encontradas y las respuestas correctas y erróneas de los estudiantes, relacionándolas, en su caso, con la literatura referida. Se finaliza con un resumen y conclusiones de los resultados encontrados, útiles, a nuestro entender para planificar la enseñanza y realizar otras investigaciones.

**Palabras Clave:** Educación estadística, elementos de significado, distribuciones bidimensionales, covarianza, correlación y regresión.

## Characterization of the Meaning of the Correlation and Regression in Students of Secondary Education)

**Summary:** In this paper the meaning that the students of Secondary Education (17 years of age) attach to the correlation and regression after having received instruction on this topic, is characterized. The paper begins with a revision of related literature, then the theoretical framework, the sample and the questionnaire are described. The questionnaire contains 12 items showing representative elements of meaning on bidimensional distributions, correlation, covariance and regression. The difficulties and the correct and erroneous answers of the students are analyzed and discussed, some of them

---

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

<sup>1</sup> Profesor Titular da Facultad de Humanidades y Ciências de la Educación. Universidad de Jaén, España. aestepa@ujaen.es.

are related to referred literature. We finalize with a summary and the conclusions of the findings, useful results to plan teaching and to begin with other research.

**Key Words:** Statistical education, bidimensional distributions, elements of meaning, covariance, correlation, and regression.

## Introducción

La amplia presencia de la Estadística en la sociedad actual, tanto en los medios de comunicación social como en los principales campos científicos (Tanur, 1972), demanda a los ciudadanos una formación estadística que les capacite para integrar diversas informaciones estadísticas (tablas, gráficos,...), interpretar resultados de estudios estadísticos, analizar datos, ... En consecuencia, los currícula educativos actualmente comprenden más contenidos estadísticos que hace unas décadas, lo que demanda investigación didáctica sobre los temas estadísticos, que sean de interés para el profesorado cuando planifiquen la enseñanza.

La asociación estadística, llamada por algunos autores covariación y, entendida como la extensión de la idea de correlación a variables cualesquiera, incluso no numéricas, (Hildebrand et al., 1977), es un tema que está presente en todos los cursos introductorios de Estadística, tanto en sus tres campos de problemas (tablas de contingencia, correlación, comparación de muestras) como en alguno de ellos. Es un tema importante, tanto en si mismo, dentro de la estadística descriptiva, como prerrequisito para continuar posteriores estudios estadísticos. Por consiguiente, creemos que es de interés la investigación didáctica sobre correlación y regresión en la enseñanza preuniversitaria.

## Antecedentes

Los primeros trabajos de interés para la enseñanza del tema aparecen en Psicología con el estudio pionero de Inhelder y Piaget (1955), sobre la correlación, estos autores concluyen que la comprensión de la idea de correlación implica la comprensión de la ideas de proporción y probabilidad, además de la capacidad combinatoria.

En Psicología, la idea de que el hombre razona de acuerdo con las leyes formales de la lógica, fue dejando paso a la idea de que el hombre se comporta como un estadístico intuitivo Peterson y Beach (1967), ante la gran cantidad de fenómenos aleatorios con los que se enfrenta en su vida diaria. Los juicios de asociación, (detección de la existencia o no de asociación, tipo de dependencia –directa, inversa, independencia y intensidad, entre variables) son tareas cotidianas en el ser humano actual. Pozo (1987) sitúa las investigaciones psicológicas sobre asociación estadística dentro de las relacionadas con el pensamiento causal.

Las investigaciones psicológicas sobre asociación estadística de interés para la enseñanza del tema las podemos clasificar (Estepa, 1994) en cuatro grandes grupos: a) las que estudian las estrategias utilizadas por los sujetos cuando tienen que realizar un juicio de asociación, es decir, cuando tienen que expresar los resultados muestran que son muchas y variadas, dependen del problema y pocos sujetos utilizan la normativamente correcta (Smedlund, 1963; Nisbett y Ross, 1980; Alloy y Tabachnik, 1984; Pérez Echeverría, 1990; entre otros). b) los que estudian la influencia de las teorías previas en el juicio de asociación; las personas tienen unas teorías previas sobre el contexto de los problemas que tienen que resolver y algunos sujetos las utilizan, aunque contradigan la evidencia de los datos del problema, sobrestimando la correlación (Jennings, Amabile y Ross, 1982). Un concepto relacionado con las teorías previas es el de “*correlación ilusoria*” o percepción de correlación, cuando no hay un hecho empírico que la sustente (Murphy y Medin, 1985); el heurístico de la accesibilidad proporciona una explicación razonable de la correlación ilusoria (Tversky y Kahneman, 1982). c) los trabajos que estudian la exactitud de los juicios de asociación, los resultados muestran que depende de la presentación de los datos y el tipo e intensidad de la asociación en el problema (Erlick y Mills, 1967; Crocker, 1981; Arkes y Harkness, 1983). d) los juicios de asociación también dependen del contexto y de la presentación de la información (Beyth-Marom, 1982; Arkes y Harkness, 1983).

Hasta hace muy pocos años apenas existían investigaciones desde el punto de vista didáctico sobre asociación estadística. En las últimas dos décadas se han realizado varias investigaciones sobre asociación estadística que podemos agrupar de la siguiente manera: a) las que estudian las estrategias de resolución y los concepciones que manifiestan los estudiantes sobre la asociación estadística; b) las

que estudian experimentos de enseñanza sobre la asociación estadística; c) las que estudian el modo en que los estudiantes pasan de una representación de la asociación (descripción verbal, diagrama de dispersión, tabla numérica y mediante el coeficiente de correlación) a otro tipo de representación, en la literatura de educación matemática se suelen llamar *proceso de traducción* (Janvier, 1987a). Según varios autores citados por Janvier (1987b), los procesos de traducción tienen interés para la resolución de problemas, la modelización matemática y para lograr competencias matemáticas. A continuación realizaremos una breve reseña de cada grupo de estas investigaciones.

a) Las que estudian las estrategias de resolución y concepciones. Una forma natural de iniciar la investigación sobre la asociación estadística desde el punto de vista didáctico es estudiar las estrategias que utilizan los estudiantes en la resolución de problemas de asociación. En Estepa (1994), Estepa y Batanero (1995, 1996), Batanero, Estepa, Godino y Green (1996) y Estepa, Batanero y Sánchez (1999), se estudian las estrategias que utilizan los estudiantes cuando resuelven problemas de asociación de los tres tipos existentes: tablas de contingencia, correlación entre variables continuas, comparación de muestras. También se estudian las argumentaciones de los estudiantes cuando resuelven estos tipos de problemas y el juicio de asociación expresado; se identifican los teoremas en acto (Vergnaud, 1990) que utilizan los estudiantes para emitir su juicio de asociación (Estepa y Batanero, 1996), obteniendo finalmente cuatro concepciones sobre la asociación estadística: *Concepción determinista*, cuando el estudiante admite un solo valor de la variable dependiente para un valor dado de la variable independiente, algunos estudiantes cuando observan que a un valor de la variable independiente le corresponde más de un valor de la variable dependiente, declaran independencia; *concepción causal*, cuando el estudiante exige la existencia de relación causal para admitir la existencia de asociación; *concepción local*, cuando el estudiante basa su juicio de asociación en una parte de los datos presentados, el juicio de asociación dependerá de la asociación que muestre la parte de los datos considerada; *concepción unidireccional*, cuando el estudiante solo admite como asociación la asociación directa; la inversa la considera como independencia o como directa. Morris (1997) estudia las concepciones de los estudiantes sobre la asociación y encuentra las concepciones causalista y unidireccional.

b) Se han realizado experimentos de enseñanza de la asociación (Estepa, 1994; Batanero, Estepa y Godino, 1997; Batanero, Godino y Estepa, 1998; Batanero y Godino, 1998; Batanero, Godino y Estepa, 2001), donde el tema de asociación estaba incluido en un curso de estadística descriptiva utilizando ordenadores, siguiendo el enfoque del análisis exploratorio de datos y analizando datos reales. Estos experimentos de enseñanza fueron cuidadosamente diseñados e implementados, en ellos se estudian la evolución de las concepciones de los estudiantes, llegando a la conclusión de que la concepción local de la asociación es superada por todos los estudiantes, las concepciones determinista y unidireccional es superada por muchos estudiantes, sin embargo la concepción causal permanece en muchos estudiantes después de la enseñanza (Batanero, Estepa y Godino, 1997). También se estudió el proceso de aprendizaje de una pareja de estudiante, encontrando nueve actos de comprensión (en el sentido de Sierpinska, 1991) necesarios para que los estudiantes adquieran el concepto de asociación estadística.

c) El paso de un modo de representación a otra o procesos de traducción de la correlación lo estudian Sánchez, Estepa y Batanero (2000) y Moritz (2004). Sánchez, Estepa y Batanero (2000) realizan su estudio con estudiantes del primer curso universitario. Controlan las variables de tarea en los tipos de representaciones presentadas, esto es, la 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 y las estrategias utilizadas por los estudiantes para realizar los juicios de asociación en las tareas propuestas; mediante un análisis de correspondencias y teniendo en cuenta las variables de tarea, identifican dos factores que relacionan las estrategias utilizadas en cada tarea propuesta, estos son: el razonamiento numérico frente al gráfico y el uso de las teorías previas para realizar el juicio de asociación. 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.

Moritz (2004), realiza un resumen de las investigaciones sobre covariación, desde el punto de vista didáctico. Estudia tres procesos

de traducción: 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 muestra los datos numéricos de una tabla. Discute sus resultados teniendo en cuenta las dificultades descritas en la literatura de investigación, y encontradas en este estudio y como evolucionan para favorecer el razonamiento sobre covariación, estas son: a) desde considerar valores aislados hasta la tendencia general de la covariación (evolución de la concepción local); b) desde considerar una sola variable hasta la consideración de la covariación (evolución hacia el reconocimiento de la distribución bidimensional); c) desde los juicios basados en las teorías previas a los juicios basados en los datos presentados (concepción causal). Finaliza con unas implicaciones para la enseñanza y futuras investigaciones.

## Objetivos

De todas las investigaciones anteriores se deduce la complejidad de la enseñanza aprendizaje de la asociación estadística, debido a que se requieren conocimientos y destrezas, gráficos y numéricos y representaciones gráficas y descriptivas. Debido a esta complejidad y dificultades hemos considerado el interés de realizar, en este primer trabajo, un estudio exploratorio que nos proporcione una caracterización del significado que los estudiantes de Educación Secundaria tienen de las nociones iniciales de asociación estadística (correlación y regresión lineal) que se estudian en este nivel de enseñanza, con el fin de adquirir conocimiento útil para planificar la enseñanza del tema realizar otras investigaciones. El presente trabajo tiene como fin alcanzar este *objetivo*.

## Marco Teórico

Adoptamos como marco teórico el enfoque ontológico-semiótico de la cognición matemática (EOS), desarrollado por J. D. Godino y 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).

El significado de un objeto matemático se define en el EOS 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 aceptado. 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 los que utilizan el saber matemático (matemática aplicada, instituciones científicas, profesionales o comerciales que precisan de las Matemáticas).

El significado de un objeto matemático tiene un carácter global, en los estudios didácticos se descompone en entidades más elementales denominadas elementos de significado. El EOS distingue los siguientes elementos de significado: situaciones problemas, lenguaje, las acciones del sujeto, conceptos, propiedades y argumentos.

Comparando el significado institucional con el significado personal alcanzado por los estudiantes, en un cierto nivel de enseñanza, podremos caracterizar los aprendizajes de los estudiantes sobre un objeto matemático determinado. En investigación didáctica se debe delimitar en primer lugar el significado institucional. El significado institucional de la correlación, lo tenemos caracterizado en trabajos anteriores sobre asociación estadística entre 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. Algunas veces, en el estudio del significado institucional, es de interés clarificar el desarrollo histórico de dicho tema; el desarrollo histórico de la ida de asociación lo hemos realizado en un trabajo anterior Estepa y Sánchez (1995).

Muestra

La muestra ha estado formada por 75 estudiantes del primer curso de Bachillerato (16 años de edad) de las modalidades de Ciencias de la Naturaleza y Salud y de Humanidades y Ciencias Sociales a los que se les había impartido el tema de dependencia no funcional, en el que se desarrollaron los contenidos estadísticos de 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.

### Cuestionario

El cuestionario es parte de otro más amplio sobre correlación y regresión. En este trabajo hemos utilizado 12 ítems, 9 de elección múltiple y 3 de emparejamiento (unir con flechas). Estos últimos economizan en tiempo y espacio el cuestionario, pues de cada ítem se extrae 3 ó 4 veces más información que en los anteriores. Los ítems 1 y 2 se han tomado, previa adaptación, de Cruise (1984), otros se han elaborado partiendo de algunos similares de Cruise (1984), y otros son de elaboración propia, siguiendo las indicaciones metodológicas de la bibliografía existente como por ejemplo Muñiz y otros (2005). Los contenidos incluidos en el cuestionario conforman una muestra amplia de elementos de significado de la correlación y regresión en la Educación Secundaria con los que queremos caracterizar el significado que estos estudiantes tienen sobre la correlación y regresión después de haber recibido enseñanza. Nos hemos ceñido más a contenidos conceptuales que a computacionales.

### Distribución Bidimensional

En el estudio de la relación existente entre dos variables es de sumo interés distinguir si las dos variables constituyen una distribución bidimensional o no. Si las variables constituyen una distribución bidimensional se pueden realizar estudios de correlación y regresión en caso contrario la correlación y regresión no tendrían sentido. Por esta razón hemos incluido el ítem 1 en el cuestionario. Dicho ítem, tomado de Cruise (1984), de elección múltiple, tiene cuatro posibles respuestas, una correcta, la b) y tres distractores. Las respuestas de los estudiantes de la muestra se pueden ver en la tabla A.1 (véase Anexo). Este ítem lo ha respondido 71 de los 75

estudiantes, lo que nos indica que la tarea, de entrada, no ha parecido difícil para estos estudiantes.

El 52% de los estudiantes responden adecuadamente, sin embargo el número de respuestas erróneas es considerable: uno de cada tres estudiantes responde la opción c), en la que se admite como distribución bidimensional dos conjuntos de datos relacionados, pero distintos, en este caso la correlación y regresión no tendría sentido y el estudio de la relación entre las variables se puede realizar, por ejemplo, mediante la comparación de muestras independientes. 6 estudiantes eligen alguna de las otras respuestas incorrectas y un estudiante elige la respuesta correcta y una incorrecta. Podemos concluir diciendo que un porcentaje elevado de estudiantes de esta muestra no discrimina correctamente entre una distribución bidimensional y dos variables estadísticas cualesquiera.

## Correlación

Generalmente en el proceso de estudio del tema de correlación y regresión en la enseñanza preuniversitaria se suele estudiar primero la correlación y después la regresión, ya que si se encuentra una correlación lineal adecuada se puede calcular, con sentido, una recta de regresión que puede aproximar los valores de una variable a partir de la otra. Por otra parte, los cálculos se suelen realizar con calculadora u ordenador, en consecuencia, hemos centrado nuestro estudio de la correlación, covarianza y regresión, en elementos de significado con más énfasis conceptual que computacional, en concreto los elementos de significado estudiados son el concepto de coeficiente de correlación, intensidad y signo que nos dan los tipos de correlación, directa inversa e independencia y las propiedades del coeficiente de correlación, haciendo especial hincapié en la correlación y causalidad, ya que como hemos visto, en la sección de antecedentes, la concepción causalista de la asociación es difícil de superar.

## Concepto de coeficiente de correlación

Para indagar la noción de coeficiente de correlación que muestran los estudiantes de esta muestra propusimos el ítem 2, que lo han respondido el 92% de los estudiantes de esta muestra. En la tabla A.2 (véase Anexo) se muestran las frecuencias y porcentajes de

las repuestas a este ítem. El 68% de los estudiantes de la muestra han dado la respuesta correcta. Entre las respuestas incorrectas cabe destacar la respuesta d) en la que los estudiantes admiten que el coeficiente de correlación se puede calcular entre dos variables cualesquiera, aunque en este caso el número de estudiantes es la mitad de los que dieron respuesta similar en el ítem 1 sobre la distribución bidimensional. Las otras dos respuestas incorrectas a) y b) tienen baja frecuencia.

Como resumen de los dos ítems hasta ahora analizados podemos decir que un tercio de estudiantes no distinguen entre distribución bidimensional y dos variables cualesquiera y que un sexto de estudiantes admiten calcular el coeficiente de correlación entre dos variables cualesquiera.

### Valor del coeficiente de correlación. Tipos de correlación

Para estudiar el significado que esta muestra de estudiantes asigna al valor del coeficiente de correlación y su relación con el tipo de correlación directa, inversa o nula, se propuso el ítem 10. Como para este estudio se necesitaban varias respuestas, decidimos proponer un ítem de emparejamiento (unir con flechas), con lo que se obtienen cuatro respuestas en un solo ítem. En la tabla A.10 (véase Anexo) se dan los resultados de las respuestas de los estudiantes. En primer lugar se puede observar que entre 14 y 15 estudiantes no respondieron a esta pregunta lo que supone un 18.7%-20% de la muestra, lo que dobla el porcentaje de no respuestas de los 5 primeros ítems. A continuación analizaremos cada una de las alternativas del ítem.

*El coeficiente de correlación es positivo, varía entre 0 y 1.* En este caso, 54 estudiantes el 72% de los estudiantes de la muestra, es decir, el 88.5% de los estudiantes que contestan dan una respuesta adecuada y afirman que si el coeficiente de correlación es positivo la relación entre las variables es directa. Entre las respuestas erróneas podemos destacar los 3 estudiantes que afirma que la relación es inversa y los otros 3 estudiantes que afirma que la relación es de los dos tipos, directa e inversa.

*El coeficiente de correlación es negativo, varía entre -1 y 0.* En este caso, 50 estudiantes, el 66.7% de los estudiantes de la muestra y el 82% de los estudiantes que han respondido, dan una respuesta

correcta, estos porcentajes son ligeramente inferiores al caso anterior, aunque los dos casos son similares. En las investigaciones desde el punto de vista psicológico y didáctico sobre correlación se observa que la correlación negativa es percibida en menor grado por los sujetos, aunque, este hecho no nos debe de extrañar, pues el mismo Galton, padre de la correlación, en sus primeros trabajos sobre la correlación solamente tuvo en cuenta la correlación positiva (Pearson, 1920). 8 estudiantes el 10.7% de los estudiantes de la muestra y el 13.1% de los que han respondido, dan las respuesta incorrecta 2-c, es decir, afirman que en este caso la relación es directa (concepción unidireccional), además 4 de ellos afirman que es de ambos tipos directa e inversa lo que indica una confusión importante.

*El coeficiente de correlación es igual a cero.* En este caso, 50 estudiantes, el 66.7% de la muestra y el 82% de los que han contestado al ítem dan una respuesta correcta. Entre las respuestas incorrectas, cabe destacar que 6 estudiantes afirman que si el coeficiente de correlación toma el valor cero, entonces debe haber un error en los cálculos.

*Obtener en los cálculos un coeficiente de correlación fuera del intervalo [-1, 1].* 51 estudiantes de los 61 que han respondido a la pregunta dan una respuesta correcta. Entre las respuestas erróneas se puede destacar que 7 estudiantes dan como respuesta la independencia. La creencia de que si el valor del coeficiente de correlación está fuera del intervalo [-1,1] significa independencia entre las variables ha sido encontrada por Morris (1997) y Sánchez (1999), aunque también con baja frecuencia.

#### Intensidad del coeficiente de correlación

Para estudiar la incidencia del signo en la evaluación de la intensidad del coeficiente de correlación que realizan los estudiantes de la muestra, propusimos el ítem 3. En las investigaciones de Morris (1997) y Estepa y Sánchez (2001) hay ítems similares a este con la diferencia que en estas dos investigaciones se pide a los estudiantes que ordenen los coeficientes de correlación (0.5, -0.8, 0.2, -0.4, 0), según su intensidad, sin embargo, en este trabajo se ha pedido a los estudiantes que elijan una ordenación determinada. En la tabla A.3 (véase Anexo), se puede observar, que 70 de los 75 estudiantes de la muestra han respondido a la pregunta. Aproximadamente la mitad de

los estudiantes de la muestra dan la respuesta correcta, 60 % en la investigación de Morris (1997), 46.1% en la investigación de Estepa y Sánchez (2001). La opción b) en la que se ordena la intensidad de la correlación, de manera incorrecta, siguiendo la ordenación de los números reales, es elegida por el 37.3% de esta muestra, en la investigación de Morris (1997) dieron esta respuesta el 15% de los estudiantes y en la investigación de Estepa y Sánchez (2001) el 17.1%; hay que tener en cuenta que los estudiantes de estas dos últimas investigaciones eran universitarios y los del presente trabajo son de Educación Secundaria. La persistencia de este error en las tres investigaciones nos sugiere que el conocimiento del orden natural de los números reales se constituye en un obstáculo, en el sentido de Brousseau (1983), para adquirir un conocimiento apropiado de la intensidad del coeficiente de correlación independientemente del signo. Las demás respuestas tienen una frecuencia baja en nuestro estudio.

### Propiedades del coeficiente de correlación

Una gran mayoría de estudiantes ha respondido correctamente a los ítems 4 y 5 (véase tablas A.4 y A.5 del Anexo) relativos a las propiedades del coeficiente de correlación: el coeficiente de correlación varía entre -1 y +1 y el coeficiente de correlación no tiene dimensión, con lo que podemos concluir que los estudiantes de esta muestra conocen adecuadamente estas propiedades. Una observación que podemos hacer es que en el ítem 10 el porcentaje de estudiantes que acepta un coeficiente de correlación no perteneciente al intervalo  $[-1,+1]$  es ligeramente superior.

### Correlación y causalidad

Barbancho (1973) distingue cinco casos en el estudio de la covariación entre variables. Cada caso se extrae del conocimiento que tengamos del contexto en que se encuentran las variables.

- a) Dependencia causal unilateral. Este tipo o clase de covariación se da cuando una variable, X, influye en otra, Y, pero no al contrario. Ejemplo: La cantidad de lluvia caída, X, influye en el rendimiento de la cosecha agrícola, Y.
- b) Interdependencia. Cuando la influencia ente X e Y es recíproca. Ejemplo: La edad del hombre (X) y la edad de la mujer (Y) en la que contraen matrimonio.

c) Dependencia indirecta. Dos variables pueden mostrar una covariación debida a una tercera variable que influye en ellas. Ejemplo: Se puede comprobar estudiando los países del mundo que el consumo diario de proteínas animales correlaciona negativamente con la tasa de natalidad, a mayor consumo de proteínas animales, menor tasa de natalidad y a menor consumo de proteínas animales, mayor tasa de natalidad. La tercera variable es el nivel de vida que se puede medir por la renta “per capita”.

d) Concordancia. A veces sabemos que dos variables son independientes, pero queremos estudiar si en sus variaciones existe una cierta concordancia. Ejemplo: la correlación que exista entre las puntuaciones dadas por dos jueces a una serie de señoritas en un concurso de belleza. Obviamente las puntuaciones dadas por cada juez son independientes, sin embargo, habrá cierta correlación positiva entre las mismas.

e) Covariación casual. Hay casos en que se observa, entre dos variables, una variación sincronizada de la que pudiera deducirse una asociación o dependencia entre dichas variables. No obstante, tal covariación puede ser totalmente casual o accidental.

En la investigación didáctica se ha encontrado que algunos estudiantes identifican asociación con causalidad (Estepa y Batanero, 1996), sin tener en cuenta los tres últimos casos de Barbancho (1973) y, además, asignan causalidad a los dos primeros. Para estudiar este extremo introdujimos el ítem 8, distinto a los utilizados por Estepa y Batanero (1996), donde incluimos un contexto comprensible para el estudiante, pero totalmente desconocido, en consecuencia, no se sabe en que caso de Barbancho (1973) estamos y, por consiguiente se debe ser prudente con la asignación de algún tipo de causalidad entre las variables y se debe concluir que Y puede depender de X, pero no afirmar que X depende con seguridad de Y. Las respuestas al ítem se muestran en la tabla A.8 (véase Anexo), donde podemos observar que 13 estudiantes, el 17.3% no responde.

53 estudiantes, el 70.7% de los estudiantes de la muestra y el 85.5% de los que han respondido admiten que las variables muestran una covariación positiva, lo que muestra una correcta interpretación del coeficiente de correlación. Sin embargo el 10.7 de los estudiantes de la muestra, o el 12.9% de los que han respondido no interpretan correctamente el coeficiente de correlación ofrecido, manifestando que existe independencia entre las variables.

Por otra parte, el 45.3 de los estudiantes de esta muestra, es decir, el 54.8% de los estudiantes que han respondido, manifiestan una concepción causalista de la asociación, afirmando que la variable dependiente depende con seguridad de la variable independiente. Ya hemos visto, en la sección 2, que en nuestras investigaciones

anteriores, Batanero, Estepa y Godino, (1997), hemos detectado que la concepción causalista es resistente al cambio y perdura después de la instrucción.

## Covarianza

La covarianza es de sumo interés en el estudio de la correlación y la regresión. Hemos dedicado dos ítems a la covarianza, los números 6 y 7, los dos tienen un índice de no respuestas elevado en comparación con los anteriores del cuestionario. En los ítems del 1 al 5 el promedio de no respuestas ha sido de 5, mientras que en los ítems 5 y 6 las no respuestas han sido 14 y 13 respectivamente, lo que nos indica que ha habido más de un 10% de estudiantes que han omitido su respuesta en relación a los ítems sobre correlación (1-6).

## Covarianza y signo de la relación

La relación de la covarianza con el signo de la correlación es una de las propiedades más importantes de esta, para explorar el significado que los estudiantes de la muestra dan a esta importante relación propusimos el ítem 6. En la tabla A.6 (véase Anexo) se dan la frecuencia y los porcentajes de las respuestas de los estudiantes, que a continuación comentaremos.

En primer lugar, 14 estudiantes, el 18.7% no responde a este ítem. 29 estudiantes, el 38.7% de los estudiantes de la muestra y el 47.5% de los estudiantes que han respondido dan una respuesta correcta. Los tres distractores incluidos tienen respuestas de importancia que pasamos a comentar. En primer lugar, 16 estudiantes, 21.3% de los estudiantes de la muestra y el 26.2% de los estudiantes que han respondido afirman que la covarianza nos muestra que las varianzas son homogéneas. 9 estudiantes el 12% de la muestra y el 14.8% de los estudiantes que han respondido afirman que la covarianza nos indica que las desviaciones típicas son positivas. 7 estudiantes, el 9.3% de la muestra y el 11.5% de los estudiantes que han respondido afirman que la covarianza nos indica que la desviación típica es menor que la varianza. En conclusión 32 estudiantes, el 42.6% de la muestra y el 52.5% de los que han respondido, dan un significado erróneo de la varianza no relacionado con el signo de la covariación.

## Signos de la covarianza, coeficiente de correlación y pendiente de la recta de regresión

Un elemento de significado importante es la relación existente entre el signo de la covarianza, el del coeficiente de correlación y el de la pendiente de la recta de regresión. Para ver como lo percibían los estudiantes de la muestra incluimos el ítem 7, las respuestas a este ítem se pueden observar en la tabla A.7 (véase Anexo). 13 estudiantes el 17.3% de los estudiantes de la muestra no han respondido al ítem, lo que duplica el número de no respuestas respecto a los ítems de correlación. 47 estudiantes, el 62.7 de los estudiantes de la muestra y el 75.8% de los que han respondido dan una respuesta correcta. 15 estudiantes, el 20% de la muestra y el 24.2 % de los que han respondido, dan una respuesta que no relaciona adecuadamente los tres signos.

## Regresión

En el estudio de la asociación estadística, en primer lugar, se suele calcular el coeficiente de correlación, si este es de intensidad suficiente se puede calcular una de las rectas de regresión para predecir, de manera aproximada, los valores de la variable explicada a partir de los valores de la variable explicativa. Para caracterizar el significado que los estudiantes de la muestra dan a la regresión hemos incluido los ítems 9, 11 y 12, haciendo hincapié e los elementos de significados conceptuales más que en los computacionales.

## Coeficiente de correlación y predicción mediante la regresión lineal

Para estudiar este punto utilizamos el ítem 11, se ha elegido de “unir con flechas” porque, de este modo podíamos abarcar este punto y otro de suma importancia y es que salvo que el coeficiente de correlación valga +1 ó -1, en los demás casos siempre es posible que pueda existir una relación no lineal entre las variables. Los resultados de este ítem se dan en la tabla A.11. El porcentaje de no respuestas a este ítem oscila entre 18.7% y el 22.7%, lo que muestra un porcentaje estimable.

*Coefficiente de correlación de intensidad alta*, son los casos 1 y 4 del ítem 11 que se deben de unir con el caso a). En primer lugar, en la tabla A.11 (véase Anexo) podemos observar que el 78.75 de los estudiantes de la muestra, o el 96.7 de los que han respondido dan la respuesta correcta 1-a). En este caso solo un estudiante ha dado una respuesta incorrecta 1-b), la respuesta 1-c) es correcta. En el caso 4, 47 estudiantes el 62.7 de la muestra y el 77.05 de los que han contestado han dado una respuesta correcta. En esta ocasión probablemente el porcentaje más bajo de respuestas se deberá al signo negativo, como ocurre en la concepción unidireccional de la asociación estadística (Batanero, Estepa, Godino y Green, 1996), donde los estudiantes perciben la dependencia entre las variables solamente cuando el signo es positivo. Este hecho no nos debe de extrañar, pues el mismo Galton, padre de la correlación, en sus primeros trabajos sobre la correlación solamente tuvo en cuenta la correlación positiva (Pearson, 1920).

*Coefficiente de correlación de intensidad muy baja*, son los casos 2 y 3 del ítem 11 que se deben unir con el caso b). En el caso 2, 50 estudiantes el 66.7% de los estudiantes de la muestra y el 84.9% de los estudiantes que han respondido a la pregunta dan una respuesta correcta, esto es, la respuesta 2-b) que reconoce que si el coeficiente de correlación toma el valor de 0.09, entonces la relación lineal entre las variables es débil y la recta de regresión no es útil para realizar predicciones. En el caso 3, 45 estudiantes el 60% de la muestra y el 77.6% de los que han respondido han dado la respuesta correcta 3-b). En este caso del coeficiente de correlación negativo, las respuestas correctas han descendido en 5 estudiantes, lo que nos vuelve a corroborar la mayor dificultad de interpretación cuando el coeficiente de correlación es negativo.

*Existencia de relación no lineal entre las variables*. La posible existencia de relación no lineal entre las variables (opción c) se debería haber elegido en todas las respuestas, sin embargo, eso no ha sido así dependiendo fuertemente del signo y valor del coeficiente de correlación. De la tabla A.11, hemos obtenido la tabla 10.1, que se muestra a continuación, con las respuestas de los estudiantes.

Tabla 10.1. Frecuencia y porcentaje de la elección de la opción c) del ítem 11

Valor del coeficiente de correlación											
r = 0.97			r = -0.98			r = 0.09			r = -0.12		
F	%m	%r	F	%m	%r	F	%m	%r	F	%m	%r
3	4.0	4.9	13	17.3	21.3	28	37.3	47.4	33	44.0	56.9
F = frecuencia absoluta de elección de la opción c) del ítem 11											
%m = porcentaje sobre el total de la muestra (75 estudiantes)											
%r = porcentaje sobre el número de estudiantes que han respondido al ítem 11											

En la tabla 10.1 podemos observar, que la elección de la opción c) aumenta considerablemente desde la mayor intensidad positiva del coeficiente de correlación a la intensidad más baja y negativa del coeficiente de correlación. Dicho de otro modo, cuando el coeficiente de correlación es de alta intensidad y positivo, tan solo el 4.9% de los estudiantes que han respondido, advierten la posibilidad de una relación no lineal. Cuando el coeficiente de correlación es de alta intensidad, pero negativo, dicho porcentaje aumenta, el 21.3% de los estudiantes que han respondido, manifiestan la posibilidad de relación no lineal. Si la intensidad de la correlación es muy baja, pero positiva, casi la mitad de los estudiantes que han respondido, 47.3%, admiten la posibilidad de que exista una relación no lineal entre las variables. Por último, si la intensidad de la correlación es muy baja y negativa, entonces más de la mitad de los estudiantes que han respondido, 56.9%, manifiestan la posibilidad de la existencia de relación no lineal entre las variables. Parece ser que aproximadamente para la mitad de los estudiantes que han respondido al ítem si no existe relación lineal, obligatoriamente debe existir otro tipo de relación y, por el contrario, si existe relación lineal, no debe existir otro tipo de relación. En resumen, la percepción de la posibilidad de la existencia de una relación no lineal entre las variables depende fuertemente de la intensidad del coeficiente de correlación y del signo de la correlación.

### Coeficiente de correlación y regresión

El dominio de la relación existente entre el valor del coeficiente de correlación y la posición relativa de las rectas de regresión implica un conocimiento bastante completo de la correlación y la regresión, en consecuencia, se propuso el ítem 12 para estudiar el significado que los estudiantes de esta muestra dan a este elemento de significado. Las frecuencias y porcentajes de las respuestas de los estudiantes al ítem 12 se dan en la tabla A.12 (véase Anexo), donde podemos observar que el porcentaje de no respuestas oscila entre el 22.7% y el 26.7%, porcentaje no desdeñable.

*Dependencia funcional y rectas de regresión.* En este caso cada una de las opciones 1 y 3 se deben relacionar con las opciones b) y d). En teoría las frecuencias y porcentajes de las respuestas a la opción 1 deberían ser idénticas a la opción 3, no es así, existen algunas diferencias que comentaremos a continuación.

En la opción 1, 19 estudiantes, 25.3% de la muestra y 33.9% de los que han respondido, han dado las dos respuestas correctas, 1-b) y 1-d), y en la opción 3, 16 estudiantes, 21.3% de la muestra, 28.1% de los que han respondido han dado las respuesta correcta, en consecuencia, aproximadamente 3 de cada 4 estudiantes de la muestra no responden totalmente correcto a esta tarea. La opción 1-b) en solitario ha sido respondida por 11 estudiantes. La opción 1-d) ha sido respondida por 12 estudiantes en solitario. En resumen, la opción 1-b) ha sido respondida por 30 estudiantes, el 40.0% de los estudiantes de la muestra, o bien, el 53.6% de los que han respondido a este ítem; la opción 1-d) ha sido respondida por 31 estudiantes, el 41.30% de los estudiantes de la muestra, o bien, el 55.4% de los estudiantes que han respondido a este ítem, aproximadamente un poco más del 50% de los estudiantes de la muestra responden bien o parcialmente bien a este ítem. Las expresiones “*los puntos del diagrama de dispersión están sobre la recta de regresión*” y “*el ángulo que forma las dos rectas de regresión es nulo y las dos rectas coinciden*” son equivalentes, y en consecuencia el porcentaje de respuesta es equivalente. En cuanto a las respuestas erróneas, cabe destacar que 8 estudiantes responden que cuando el coeficiente de correlación vale  $-1$ , entonces las dos rectas de regresión forman un ángulo agudo; y 8 estudiantes responden que cuando el coeficiente de correlación vale  $+1$ , entonces las dos rectas de regresión son perpendiculares. Las demás respuestas incorrectas son de baja frecuencia.

*Dependencia aleatoria y rectas de regresión.* En este caso la respuesta correcta es la 4-c), que ha sido respondida por 36 estudiantes que representan el 48% de la muestra y el 62.1% de los que responden a esta pregunta; 3 estudiantes han dado las respuestas 4-c) y 4-d) conjuntamente que no pueden ser ciertas simultáneamente. Entre las respuestas erróneas destaca la 4-d), 13 estudiantes la responden en solitario y otros 3 ligada a la 4-c), lo que da un total de 16 estudiantes, el 21.3% de la muestra y el 27.6% de los que han respondido que afirman que cuando el coeficiente de correlación es distinto de -1, 0, ó 1 los puntos del diagrama de dispersión están sobre la recta de regresión. También hay 4 estudiantes que afirman que en ese caso las rectas de regresión son perpendiculares. Todo esto nos muestra que uno de cada tres estudiantes que ha respondido no relaciona correctamente la dependencia aleatoria con las rectas de regresión.

*Independencia y rectas de regresión.* La respuesta correcta en este caso es la 2-a), que la han dado 36 estudiantes que representan el 48% de los estudiantes de la muestra y el 65.5% de los estudiantes que han respondido a la pregunta; 3 estudiantes dan la respuesta correcta ligada con otra que las convierte en incoherentes. Entre las respuestas incorrectas podemos destacar la 2-b) que la responden 11 estudiantes en solitario y 2 estudiantes más ligada a la correcta 2-a), lo que la hace incoherente; por consiguiente, 11 estudiantes, el 14.7% de la muestra y 20% de los que han respondido afirman que cuando el coeficiente de correlación toma el valor cero, entonces el ángulo que forman las dos rectas de regresión es nulo y las dos rectas coinciden, seguramente identifican el cero del coeficiente de correlación con el “ángulo nulo”.

## Rectas de regresión

Una propiedad importante en el estudio de la regresión es que las dos rectas de regresión se cortan en el punto definido por sus respectivas medias, propiedad que utiliza la ecuación punto-pendiente de la recta de regresión. Para estudiar el significado que los estudiantes de esta muestra dan a esta propiedad propusimos el ítem 9. En la tabla A.9 (véase Anexo), se dan los resultados de las respuestas de los estudiantes a este ítem, donde podemos observar que, en relación con los anteriores este ítem ha tenido más número de respuestas, concretamente lo ha respondido 66 estudiantes, es

decir el 88% de la muestra. 55 estudiantes, el 73.3% de la muestra y el 83.3% de los que han respondido han dado una respuesta adecuada afirmando que las dos rectas de regresión pasan por el punto definido por las dos medias. En cuanto a las respuestas erróneas podemos destacar que 8 estudiantes, el 10.7% de la muestra y el 12.1% de los que han respondido afirman que ambas rectas de regresión siempre coinciden, probablemente estos estudiantes han calculado y representado pocas veces las dos rectas de regresión de una misma distribución bidimensional.

## Resumen y Conclusiones

En la literatura consultada hemos encontrado resultados de investigación que pueden ser útiles para el diseño de la enseñanza de este tema, también dichos resultados se han relacionado con las distintas secciones de este trabajo. Los resultados de nuestra investigación los resumimos a continuación.

Hemos comenzado nuestro estudio, explorando el significado que los estudiantes de esta muestra otorgan a la distribución bidimensional, encontrando que poco más de la mitad, le dan un significado adecuado a la distribución bidimensional y un tercio de los estudiantes admiten como distribución bidimensional dos variables que no forman dicha distribución, aunque este porcentaje baja a la mitad cuando se trata de calcular el coeficiente de correlación entre variables cualesquiera, en consecuencia este extremo se deberá tener en cuenta en la planificación de la enseñanza, sin embargo, cuando se pregunta a los estudiantes que es el coeficiente de correlación, dos tercios de los estudiantes responden adecuadamente.

Aproximadamente tres cuartos de los estudiantes afirman que si el coeficiente de correlación es positivo la dependencia es directa, esta fracción baja a los dos tercios cuando el coeficiente de correlación es negativo y afirman que la dependencia es inversa; se obtiene igual porcentaje cuando el coeficiente de correlación es nulo y existe independencia. Aproximadamente, un diez por ciento de los estudiantes muestran la concepción unidireccional.

La mitad de los estudiantes ordenan adecuadamente los coeficientes de correlación según su intensidad, pero un poco más de la tercera parte, ante esta tarea eligen el orden natural de los

números reales, como este hecho lo hemos encontrado en otras investigaciones podemos deducir que el orden natural de los números reales se constituye en un obstáculo.

Las propiedades del coeficiente de correlación: el coeficiente de correlación varía entre +1 y -1 y el coeficiente de correlación es adimensional ha obtenido una alta tasa de respuestas correctas.

La concepción causalista de la asociación, encontrada en investigaciones anteriores, así como su persistencia al cambio después de la instrucción, aparece aquí en aproximadamente la mitad de los estudiantes.

Poco menos de la mitad de los estudiantes no relaciona adecuadamente el signo de la covarianza y el tipo de la correlación (directa, inversa, independencia), mientras que aproximadamente la cuarta parte no relaciona adecuadamente los signos de la covarianza el coeficiente de correlación y el coeficiente de regresión.

La posibilidad de predicción de los valores de una variable a partir de los valores de la otra variable mediante la ecuación de regresión lineal es distinguida por casi la totalidad de los estudiantes que responde al ítem correspondiente si el valor del coeficiente de correlación es alto y positivo, este porcentaje desciende si es alto y negativo. Sin embargo cuando el coeficiente de correlación es de intensidad muy baja, aproximadamente el 85% de los estudiantes que responden dan una respuesta correcta en el caso positivo y en el caso negativo el porcentaje baja al 77.6%.

El reconocimiento a la existencia de relación no lineal entre las variables, por parte de los estudiantes de la muestra, depende de la intensidad y signo del coeficiente de correlación, variando el porcentaje de estudiantes que han respondido y que no dan la posibilidad de relación no lineal, desde 4.9% para un coeficiente de correlación de intensidad alta y positiva a un 56.9% para un coeficiente de correlación de intensidad muy baja y negativa. Para los estudiantes de la muestra, la posibilidad de la existencia de relación no lineal depende fuertemente del signo y de la intensidad del coeficiente de correlación.

La posición de las rectas de regresión cuando existe dependencia funcional es discriminada correctamente por el 25% de los estudiantes de la muestra y parcialmente correcta por aproximadamente un poco más del 50% de los estudiantes que han

respondido, sin embargo cuando la dependencia es aleatoria, 62.1 de los estudiantes que han respondido dan una respuesta correcta, lo mismo ocurre cuando existe independencia.

Una propiedad importante de la ecuación de regresión, que es utilizada para el cálculo de la ecuación punto-pendiente, es que ambas rectas de regresión pasan por el punto definido por las dos medias de las variables, esta propiedad es admitida por el 83.3% de los estudiantes que responden al ítem.

Nuestros resultados pueden ser de interés para la planificación de la enseñanza del tema, pueden ser utilizados cuando se deciden intervenciones de enseñanza y cuando se diseñan actividades para desarrollar dicha enseñanza, como por ejemplo, para un error o dificultad, de no muy elevado porcentaje de aparición se puede hacer más énfasis en la enseñanza y en el diseño de actividades; si el error o dificultad es más persistente, como la concepción causal, que es difícil de superar, ya que está influenciada, entre otras cosas, por el énfasis determinista de los currículos escolares, se debe trabajar en la enseñanza, haciéndola aparecer para que los estudiantes encuentren contradicciones que le lleven a ir superando paulatinamente dicha concepción.

En cuanto a la investigación, del análisis realizado se pueden diseñar investigaciones que profundicen en las dificultades y errores que hemos detectado, o bien experimentos de enseñanza, donde la enseñanza sea controlada y se estudien la evolución con la instrucción de las dificultades y errores que hemos encontrado.

#### Nota

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

#### Referencias Bibliográficas

ALLOY, L. B. Y TABACHNIK, N. Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*, v. 91, pp. 112-149, 1984.

ARKES, H. R. Y HARKNESS, A. R. Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General*, v. 112, n. 1, pp. 117-135, 1983.

BARBANCHO, A. G. *Estadística elemental moderna*. Barcelona: Ariel, 1973. 4ª edición, reimpresión de 1975.

BATANERO, M. C.; ESTEPA, A.; GODINO, J. D. Y GREEN, D. Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, v. 27 n. 2, pp. 151-169, 1996.

BATANERO, M. C.; ESTEPA, A. Y GODINO, J. D. Students' understanding of Statistical association in a computer environments. In J. Garfield & G. Burrill (Eds .) *Research on the Role of Technology in Teaching and Learning Statistics*. Voorburg, The Netherlands: International Statistical Institute, pp. 191-205, 1997.

BATANERO, C.; GODINO, J. D. Y ESTEPA, A. Building the meaning of statistical association through data analysis activities (Research Forum). In A. Olivier & K. Newstead (Eds.). *Proceedings of the 22<sup>nd</sup> Conference of the International Group for the Psychology of Mathematics Education*. Stellenbosh (South Africa): University of Stellenbosh, 1998, v. 1, pp. 221-236.

BATANERO, C. & GODINO, J. D. Understanding Graphical and Numerical Representations of Statistical Association in a computer environment. En: Pereira-Mendoza, L., Seu, L., Wee, T. y Wong, W.-K. (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics* Voorburg (The Netherlands): ISI Permanent Office, 1998, v. 2, pp. 1017-1023.

BATANERO, C.; GODINO, J. Y ESTEPA, A. 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. En J. R. Pascual (Eds.). *II Simposio de la Sociedad Española de Investigación en Educación Matemática (SEIEM)*. Pamplona: Universidad Pública de Navarra. 2001, pp. 151-169.

BEYTH-MAROM, R. (1982). Perception of correlation reexamined. *Memory and Cognition*, v. 10, n. 6, 1982, pp. 511-519.

BROUSSEAU, G. (1983). Les obstacles epistemologiques et les problemes en Mathematiques. *Recherches en Didactique des Mathématiques*, v. 4, n. 2, 1983, pp. 165-198.

CROCKER, J. (1981) Judgment of covariation by social perceivers. *Psychological Bulletin*. 90, 2, 272-292.

CRUISE, R. J.; DUDLEY, R. L. Y THAYER, J. D. *A Resource Guide for Introductory Statistics*. Dubuque (Iowa): Kendall/Hunt Publishing Company, 1984.

ERLICK, D. E. Y MILLS, R. G. Perceptual quantification of conditional dependency. *Journal of Experimental Psychology*, v. 73, n. 1, pp. 9-14, 1967.

ESTEPA, A. 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. 1994.

ESTEPA, A. y BATANERO, M. C. Concepciones iniciales sobre la asociación estadística. *Enseñanza de las Ciencias*, Barcelona, España, v. 13, n. 2, pp. 155 – 170, 1995

ESTEPA, A. Y BATANERO, M. C. Judgments of correlation in scatter plots: An empirical study of students' intuitive strategies and preconceptions. *Hiroshima Journal of Mathematics Education*, Hiroshima (Japan), v. 4, pp. 25-41, 1996.

ESTEPA, A., BATANERO, C. y SÁNCHEZ, F. T. Students' understanding of regression lines. In O. Zaslavsky (Ed.). *Proceedings of the 23<sup>rd</sup> Conference of the International Group for the Psychology of Mathematics Education – PME 23*, Vol. 2,. Haifa (Israel): Technion-Israel Institute of Technology, 1999, pp. 2-313, 2-320

ESTEPA, A. y SÁNCHEZ, F. T. Desarrollo histórico de la idea de asociación estadística. *Epsilon*, N. 30, v, 10 f. 3, pp. 61 – 73, 1995

ESTEPA, A. Y SÁNCHEZ, F.T. 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* (vol. 2), Voorburg (The Netherlands): ISI Permanent Office, 1998, pp. 671-676).

ESTEPA, A. Y SÁNCHEZ, F. T. Empirical Research On The Understanding Of Association And Implication For The Training Of Researchers. En C. Batanero (Ed.) *Training Researchers in the Use of Statistics*. Granada (Spain): International Association for Statistical Education. International Statistical Institute, 2001, pp. 37-51.

- 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, 2003, recuperable en (9-julio-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*, v. 14, n. 3, 1994, pp. 325-355.
- HILDEBRAND, D. K., LANG, J. D. y ROSENTHAL, H. *Analysis of nominal data*. London:: Sage University Paper, 1977.
- INHELDER, B. Y PIAGET, J. *De la logique de l'enfant à la logique de l'adolescent*. Paris: Presses Universitaires de France, 1955. (Traducción castellana, Primera reimpresión, 1985. Barcelona: Paidós).
- JANVIER, C. *Problems of representation in the teaching and learning of mathematics*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers, 1987a, n. p. 247.
- JANVIER, C. Translation processes in Mathematics Education. En C. Janvier. *Problems of representation in the teaching and learning of mathematics*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers, 1987b, pp. 27-32.
- JENNINGS, D. L., AMABILE, T. M. Y ROSS, L. (1982). Informal covariation assessment: Data-based versus theory-based judgments. En D. Kahneman, P. Slovic y A. Tversky (Ed.), *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press, 1982. pp. 211-230.
- MORITZ, J. Reasoning about covariation. In D. Ben-Zvi & J. Garfield (Eds.). *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*, n. p. 423. Dordrecht (The Netherlands): Kluwer Academic Publishers, 2004, pp. 227-255.
- MORRIS, E. J. An investigation of students' conceptions and procedural skills in the statistical topic correlation. *Centre for Information Technology in Education*, Report n. 230, London: The Open University, 1997, n, p. 64.
- MUÑIZ, J.; FIDALGO, A. M.; CUETO, E. G<sup>a</sup>.; MARTÍNEZ, R. Y MORENO, R. *Análisis de ítems*. Madrid: Editorial "La Muralla", 2005, n. p. 181.

- MURPHY, G. L. Y MEDIN, D. L. The role of theories in conceptual coherence. *Psychological Review*, v. 92, n. 3, pp. 289-316, 1985.
- NISBETT, R. Y ROSS, L. *Human inference: strategies and shortcomings of social judgment*. Nueva Jersey: Prentice Hall, 1980.
- PEARSON, K. Notes on the history of correlation. *Biometrika*, 13, 25-45, 1920. En Pearson, E. S. y Kendall, M. (1970), (Eds.), *Studies in the history of Statistics and Probability*, 1970, v. 1. London: Charles Griffin, pp. 185-205.
- PETERSON, C. R. Y BEACH, L. R. Man as an intuitive statistician. *Psychological Bulletin*, v. 68, n. 1, 1967, 29-46.
- PÉREZ ECHEVERRÍA, M. P. *Psicología del razonamiento probabilístico*. Madrid: Ediciones de la Universidad Autónoma de Madrid, 1990.
- POZO, J. I. v. *Aprendizaje de la ciencia y pensamiento causal*. Madrid: Visor, 1987.
- SMEDLUND, J. The concept of correlation in adults. *Scandinavian Journal of Psychology*, 4, 165-174, 1963.
- SÁNCHEZ, F. T. Significado de la correlación y regresión para los estudiantes universitarios. Tesis doctoral. Departamento de Didáctica de la Matemática. Universidad de Granada, 1999.
- SÁNCHEZ, F.T.; ESTEPA, A. Y BATANERO, C. Un estudio experimental de la estimación de la correlación a partir de diferentes representaciones. *Enseñanza de las Ciencias*, v. 18, n. 2, 297-310, 2000.
- SIERSPINSKA, A. Some remarks on understanding in mathematics. For the learning of mathematics, v. 10, n. 3, pp. 24-36, 1991.
- TANUR, J. M. (Ed.) *Statistics: A guide to the Unknown*. San Francisco: Holden Day, 1972.
- TVERSKY, A. Y KAHNEMAN, D. (1982b). Judgment under uncertainty: Heuristics and biases. En D. Kahneman, P. Slovic y A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press, 1982, pp. 3-20. (Traducido al español en M. Carretero y J. A. García Madruga (1984). *Lecturas de psicología del pensamiento. Razonamiento, solución de problemas y desarrollo cognitivo*. Madrid: Alianza Editorial, 1984, pp. 169-184.

VERGNAUD, G. La théorie des champs conceptuels. Recherches en Didactique des Mathématiques, v. 10, n. 2-3, pp. 133-170, 1990.

## Anexo: Cuestionario y frecuencia y porcentaje de respuestas

*(la respuesta correcta está marcada con una X)*

Tabla A.1. Frecuencia y porcentaje a las respuestas del ítem 1

Ítem 1	Señale con una "X" la respuesta correcta. Una distribución bidimensional consiste en:	Frecuencia	Porcentaje
	a) Dos conjuntos de datos, valores o puntuaciones correspondientes a características variables de dos grupos distintos de personas, animales o cosas	3	4.0
X	b) Dos conjuntos de datos, valores o puntuaciones correspondientes a dos características variables de un solo grupo de personas, animales o cosas	39	52.0
	c) Dos conjuntos de datos, valores o puntuaciones de dos grupos relacionados (por ejemplo: la edad de los habitantes de un pueblo y la edad de los habitantes del pueblo más cercano).	25	33.3
	d) Un conjunto de datos, valores o puntuaciones de dos grupos no relacionados	3	4.0
	b) y c)	1	1.3
	No responde	4	5.3
	TOTAL	75	100.0

Tabla A.2. Frecuencia y porcentaje a las respuestas del ítem 2

Ítem 2	Señale con una "X" la respuesta correcta. El coeficiente de correlación es,	Frecuencia	Porcentaje
	a) un coeficiente que sirve para estimar los valores de una variable conocidos los valores de otra variable	2	2.7
	b) un coeficiente sirve para calcular la ecuación de una recta que se ajuste adecuadamente a los datos	3	3.0
X	c) un coeficiente que describe el grado de relación existente entre los valores de una distribución bidimensional (el peso y la altura de los alumnos de la clase A)	51	68.0
	d) un coeficiente que describe el grado de relación existente entre dos variables cualesquiera (el peso de los alumnos de la clase A y la altura de los alumnos de la clase B)	13	17.3
	No responde	6	8.0
	TOTAL	75	100.0

\*\*\*

Tabla A.3. Frecuencia y porcentaje a las respuestas del ítem 3

Ítem 3	Señale con una "X" la respuesta correcta. La ordenación de los coeficientes de correlación: 0,23; 0,87; -0,72; 0,0; -0,18; de menor intensidad de relación entre las variables a mayor intensidad de relación entre las variables es la siguiente:	Frecuencia	Porcentaje
X	a) Menor intensidad 0,0; -0,18; 0,23; -0,72; 0,87 mayor intensidad	38	50.7
	b) Menor intensidad -0,72; -0,18; 0,0; 0,23; 0,87 mayor intensidad	28	37.3
	c) Menor intensidad 0,0; -0,18; -0,72; 0,23; 0,87 mayor intensidad	3	4.0
	d) Menor intensidad 0,0; -0,72; -0,18; 0,23; 0,87 mayor intensidad	0	0.0
	e) Otra (escribir) _____	0	0.0
	a) y c)	1	1.3
	No responde	5	6.7
	TOTAL	75	100.0

Tabla A.4. Frecuencia y porcentaje a las respuestas del ítem 4

Ítem 4	De las siguientes afirmaciones señale con una "X" la que es correcta.	Frecuencia	Porcentaje
	a) El coeficiente de correlación puede ser cualquier número positivo, ya que se obtiene haciendo cálculos con números muy diversos	1	1.3
	b) El coeficiente de correlación puede ser cualquier número negativo, ya que se obtiene haciendo cálculos con números muy diversos	0	0.0
X	c) El coeficiente de correlación varía entre -1 y +1	69	92.0
	d) El coeficiente de correlación no puede valer cero	0	0.0
	a), b) y d)	1	1.3
	No responde	4	5.3
	TOTAL	75	100.0

\*\*\*

Tabla A.5. Frecuencia y porcentaje a las respuestas del ítem 5

Ítem 5	Señale con una "X" la respuesta correcta. Se ha calculado el coeficiente de correlación entre la altura (medida en centímetros) y el peso (medido en kilogramos) de un grupo de personas, obteniendo un valor para el coeficiente de correlación de 0,63. ¿En qué unidad o unidades se mide este coeficiente de correlación?	Frecuencia	Porcentaje
	a) En centímetros o en kilogramos, entonces sería 0,63 cm, o bien 0,63 kg	0	0.0
	b) En cm/kg, entonces sería 0,63 cm/kg	2	2.7
	c) En kg/cm, entonces sería 0,63 kg/cm	0	0.0
	d) En cm por kg, entonces sería 0,63 cm·kg	1	1.3
X	e) El coeficiente de correlación es adimensional y no se expresa en unidades de medida	65	86.7
	No responde	7	9.3
	TOTAL	75	100.0

Tabla A.6. Frecuencia y porcentaje a las respuestas del ítem 6

Ítem 6	Señale con una "X" la respuesta correcta. La covarianza nos indica que:	Frecuencia	Porcentaje
	a) las desviaciones típicas son positivas	9	12.0
	b) las varianzas son homogéneas	16	21.3
X	c) la relación entre las variables es directa, inversa o nula	29	38.7
	d) la desviación típica es menor que la varianza	7	9.3
	No responde	14	18.7
	TOTAL	75	100.0

Tabla A.7. Frecuencia y porcentaje a las respuestas del ítem 7

Ítem 7	Señale la respuesta verdadera	Frecuencia	Porcentaje
	a) Si la covarianza y el coeficiente de correlación tienen el signo positivo, entonces la pendiente de la recta de regresión tiene signo negativo	4	5.3
	b) Si la covarianza y el coeficiente de correlación tienen el signo negativo, entonces la pendiente de la recta de regresión tiene signo positivo	7	9.3
	c) Si la covarianza y el coeficiente de correlación tienen el mismo signo, entonces la pendiente de la recta de regresión tiene signo negativo	4	5.3
X	d) La covarianza, el coeficiente de correlación y la pendiente de la recta de regresión siempre tienen el mismo signo	47	62.7
	No responde	13	17.3
	TOTAL	75	100.0

Tabla A.8. Frecuencia y porcentaje a las respuestas del ítem 8

Ítem 8	Hemos calculado el coeficiente de correlación entre dos variables X (número de manos de una muestra de habitantes del planeta P) y la variable Y (número de lenguas que hablan la misma muestra anterior de habitantes del planeta P, no sabemos si pueden hablar con las manos), obteniendo un valor de 0'98, entonces podemos afirmar que: (Elija la mejor respuesta de entre las siguientes).	Frecuen-cia	Porcen-taje
	a) El número de lenguas que habla un individuo de esta muestra de habitantes del planeta P depende con seguridad del número de manos que tenga dicho habitante	34	45.3
X	b) El número de lenguas que habla un individuo de esta muestra de habitantes del planeta P puede depender del número de manos que tenga dicho habitante	19	25.3
	c) El número de lenguas que habla un individuo de esta muestra de habitantes del planeta P no esta relacionado con el número de manos que tenga dicho habitante	8	10.7
	d) Cuantas más manos tenga un individuo de esta muestra de habitantes del planeta P, menos lenguas hablará	1	1.3
	No responde	13	17.3
	TOTAL	75	100.0

Tabla A.9. Frecuencia y porcentaje a las respuestas del ítem 9

Ítem 9	Elija la frase que siempre es verdadera	Frecuen-cia	Porcen-taje
	a) Las dos rectas de regresión pasan solamente por dos puntos de la nube de puntos	0	0.0
	b) Las dos rectas de regresión pasan por los puntos más alejados de la nube de puntos	2	2.7
X	c) Las dos rectas de regresión pasan por el punto $(\bar{x}, \bar{y})$ , definido por cada uno de los valores de las medias aritméticas de las variables	55	73.3
	d) Las dos rectas de regresión siempre coinciden	8	10.7
	c) y d)	1	1.3
	No responde	9	12.0
	TOTAL	75	100.0

Ítem 10. Una con flechas (puede “salir”/”llegar” más de una flecha de/a cada rectángulo)

1.	Si el coeficiente de correlación toma un valor entre 0 y 1		a)	entonces la relación entre las variables es inversa, cuando una variable aumenta, la otra disminuye
2.	Si el coeficiente de correlación toma un valor entre -1 y 0		b)	entonces, debe haber un error en los cálculos
3.	Si el coeficiente de correlación toma un valor menor que -1 ó mayor que 1 (por ejemplo: -3,7, o bien, 7,8)		c)	entonces la relación entre las variables es directa, cuando una variable aumenta, la otra también aumenta
4.	Si el coeficiente de correlación es igual a cero		d)	entonces no existe relación entre las variables, estas son independientes

Tabla A.10 Frecuencia y porcentaje de las respuestas al ítem 10

Res- puesta	Fr.	%	Res- puesta	Fr.	%	Res puest a	Fr.	%	Res puesta	Fr.	%
1-a)	3	4,0	2-a)	50	66,7	3-b)	51	68,0	4-a)	1	1,3
1-c)	54	72,0	2-b)	1	1,3	3-c)	2	2,7	4-b)	4	5,3
1-d)	1	1,3	2-c)	4	5,3	3-d)	1	1,3	4-c)	2	2,7
1-a) y 1-c)	3	4,0	2-a) y 2-c)	4	5,3	3-a) y 3-b)	1	1,3	4-d)	50	66,7
No responde	14	18,7	2-a) y 2-d)	1	1,3	3-b) y 3-d)	6	8,0	4-a) y 4-d)	1	1,3
TOTAL	75	100,0	No responde	15	20,0	No respon de	14	18,7	4-b) y 4-d)	2	2,7
			TOTAL	75	100,0	TOTAL	75	100,0	No responde	15	20,0
									TOTAL	75	100,0

Ítem 11. Una con flechas (puede “salir”/”llegar” más de una flecha de/a cada rectángulo)

1.	Si el coeficiente de correlación es igual a 0,97		a)	entonces la relación lineal entre las variables es fuerte y se pueden predecir, de manera aproximada, los valores de una variable a partir de los valores de la otra mediante la ecuación de la recta de regresión
2.	Si el coeficiente de correlación es igual a 0,09		b)	entonces la relación lineal entre las variables es débil y no se puede utilizar la recta de regresión para predecir los valores de una variable a partir de los valores de la otra
3.	Si el coeficiente de correlación es igual a -0,12		c)	puede existir una relación no lineal entre las variables
4.	Si el coeficiente de correlación es igual a -0,98			

Tabla A.11 Frecuencia y porcentaje de las respuestas al ítem 11

Res- puesta	Fr.	%	Res- puesta	Fr.	%	Res- puesta	Fr.	%	Res- pues- ta	Fr.	%
1-a)	57	76,0	2-a)	3	4,0	3-b)	25	33,3	4-a)	45	60,0
1-b)	1	1,3	2-b)	28	37,3	3-c)	13	17,3	4-b)	3	4,0
1-c)	1	1,3	2-c)	6	8,0	3-b) y 3-c)	20	26,7	4-c)	7	9,3
1-a) y 1-c)	2	2,7	2-b) y 2-c)	22	29,3	No respon- de	17	22,7	4-a) y 4-c)	2	2,7
No respon- de	14	18,7	No respon- de	16	21,3	TOTAL	75	100,0	4-b) y 4-c)	4	5,3
TOTAL	75	100,0	TOTAL	75	100,0				No respo- nde	14	18,7
									TOTA L	75	100,0

Ítem 12. Una con flechas

1.	Si el coeficiente de correlación vale -1,	a)	entonces las dos rectas de regresión son perpendiculares
2.	Si el coeficiente de correlación vale 0,	b)	entonces, el ángulo que forman las dos rectas de regresión es nulo y las dos rectas coinciden
3.	Si el coeficiente de correlación vale +1,	c)	entonces, las dos rectas de regresión se cortan formando un ángulo agudo
4.	Si el valor del coeficiente de correlación es distinto de -1, 0, +1, pero comprendido ente -1 y +1	d)	entonces, los puntos del diagrama de dispersión están sobre la recta de regresión

Tabla A.12 Frecuencia y porcentaje de las respuestas al ítem 12

Res- puesta	Fr.	%	Res- puesta	Fr.	%	Respu esta	Fr.	%	Respu esta	Fr.	%
1-a)	2	2,7	2-a)	36	48,0	3-a)	8	10,7	4-a)	4	5,3
1-b)	11	14,7	2-b)	11	14,7	3-b)	18	24,0	4-b)	1	1,3
1-c)	8	10,7	2-c)	4	5,3	3-d)	12	16,0	4-c)	36	48,0
1-d)	12	16,0	2-d)	1	1,3	3-a) y 3-c)	1	1,3	4-d)	13	17,3
1-a) y 1-c)	1	1,3	2-a) y 2-b)	2	2,7	3-a) y 3-d)	1	1,3	4-c) y 4-d)	3	4,0
1-a) y 1-d)	2	2,7	2-a) y 2-d)	1	1,3	3-b) y 3-c)	1	1,3	4-a), 4-b) y 4-d)	1	1,3
1-b) y 1-c)	1	1,3	No respon de	20	26,7	3-b) y 3-d)	16	21,3	No respon de	17	22,7
1-b) y 1-d)	19	25,3	TOTAL	75	100,0	No respon de	18	24,0	TOTA L	75	100,0
No respon de	19	25,3				TOTA L	75	100,0			
Total	75	100,0									

