

# LA ESTIMACIÓN DE LA CORRELACIÓN: VARIABLES DE TAREA Y SESGOS DE RAZONAMIENTO

MARIA M. GEA

CARMEN BATANERO

GUSTAVO R. CAÑADAS

PEDRO ARTEAGA

## RESUMEN

*La correlación es un concepto estadístico central, pues extiende la idea de dependencia funcional a variables estadísticas. Sin embargo, la investigación previa ha descrito sesgos de razonamiento y dificultades asociadas a su comprensión. En este trabajo se analiza la importancia de la correlación como idea estocástica fundamental y su lugar en el currículo español. Seguidamente nos centramos en las tareas de estimación de la correlación, describiendo las principales variables que determinan su dificultad. Se resumen los resultados de las investigaciones sobre concepciones erróneas y sesgos de razonamiento identificados en las mismas. Finalmente se presentan algunas implicaciones para la investigación y la enseñanza.*

**Palabras clave:** Correlación; tareas de estimación; variables de tarea; sesgos y concepciones erróneas.

## 1. INTRODUCCIÓN

Explicar, controlar y predecir los sucesos que se presentan a lo largo de la vida depende de habilidades y destrezas para detectar covariaciones entre los acontecimientos (Alloy y Tabachnik, 1984). En consecuencia, el razonamiento covariacional es una actividad cognitiva fundamental para el desarrollo del ser humano (Moritz, 2004; Zieffler, 2006; McKenzie y Mikkelsen, 2007) aunque no por ello está exenta de dificultades.

Disciplinas como la psicología, sociología y la didáctica de la estadística han analizado las dificultades en la realización de un juicio de covariación y las variables que influyen en dicha tarea. Dichas investigaciones indican que el razonamiento humano se comporta, en situaciones de incertidumbre, de un modo inadecuado, debido a los valores y creencias del propio individuo, que interaccionan con la situación que se le presenta. Como consecuencia, el procesamiento humano de la información difiere de un proceso algorítmico en el que, paso a paso y siguiendo criterios dados, se produce una solución dentro de una clase de problemas (Batanero, 2001; Moritz, 2004).

La finalidad de este trabajo es realizar una síntesis de estas investigaciones para informar al profesor de estadística sobre estas dificultades, de forma que sean tenidas en cuenta en el diseño de experiencias de enseñanza que contribuyan al desarrollo del razonamiento covariacional de sus estudiantes. Nos centraremos únicamente en la percepción de la covariación entre variables numéricas y la estimación de su intensidad (que llamaremos tareas correlacionales, en adelante).

Dichas nociones aparecen en el estudio de la estadística, en especial en la educación secundaria, bachillerato y universidad. Se suele comenzar el trabajo con las mismas en el análisis de la presencia de covariación entre dos variables a partir de un diagrama de dispersión; para diferenciar el sentido y valorar intuitivamente su intensidad. Esta aproximación intuitiva da paso posteriormente al desarrollo formal de las ideas de covarianza, correlación y coeficiente de correlación y posteriormente para realizar un estudio de regresión. Por su importancia, lleva al profesorado *“a interesarse por el razonamiento de los estudiantes en cuanto a la lectura diagramas de dispersión, a la interpretación de la correlación y otras destrezas que son utilizadas en el estudio e interpretación de los datos bivariados”* (Zieffler, 2006, p. 7).

Para llevar a cabo esta síntesis, hemos revisado los trabajos resumidos en Pérez Echeverría (1990), Estepa (1994, 2004), Shaughnessy (1992, 2007), Shaughnessy, Garfield y Greer (1996) y Engel y Sedlmeier (2011). Se analizaron también las investigaciones relacionadas publicadas en los congresos ICOTS (International Conference on Teaching Statistics) y las principales revistas y congresos de educación estadística.

En lo que sigue presentamos esta síntesis, comenzando por analizar la importancia de la correlación, como idea estocástica fundamental. Seguidamente, analizamos los pasos en la estimación de una tarea correlacional, presentando las variables que pueden influir en su dificultad y las estrategias de los estudiantes en la resolución de estas tareas. A continuación, se resumen los principales resultados de la investigación sobre concepciones erróneas, sesgos y dificultades en la estimación de la correlación y se finaliza con algunas implicaciones para la enseñanza del tema.

## 2. LA CORRELACIÓN: UNA IDEA ESTOCÁSTICA FUNDAMENTAL

Cuando se planifica la enseñanza de una materia, como el caso de la estadística, se debe prestar especial atención al conjunto de conceptos sobre los cuáles se apoya y fundamenta. El primero que habló en educación estadística de ideas fundamentales fue Heitele (1975), quien identificó las mismas en el campo de la probabilidad. Dicho autor toma de Bruner (1960) los siguientes principios:

1. La finalidad de la enseñanza es ayudar a los estudiantes a construir las ideas fundamentales.

2. Estas ideas se caracterizan porque pueden ser enseñadas en forma comprensible a estudiantes de diferente edad, siempre que se elija un lenguaje, grado de formalización y actividades adecuadas a su nivel de desarrollo cognitivo.
3. Las ideas fundamentales deben ser la guía que estructura el currículo, desde la escuela primaria a la universidad, para garantizar una cierta continuidad. Por ello se han de enseñar en diferentes niveles cognitivos y lingüísticos, formando una “espiral curricular”.
4. La transición de un estudiante a un nivel cognitivo superior se facilita si el tema subyacente ha sido estudiado y comprendido en un nivel de abstracción adecuado en las etapas educativas anteriores. Es decir, es preferible que el niño comience a estudiar el tema poco a poco, aunque en un principio sólo lo comprenda de forma limitada, en lugar de esperar a que madure y se le pueda enseñar directamente en forma más abstracta.

En consecuencia, las ideas fundamentales estocásticas pueden presentarse al estudiante en cada etapa de su desarrollo, adaptadas gradualmente, para proporcionarles modelos explicativos a fenómenos estadísticos que observa. De este modo, se propicia que entienda su entorno por sus propios medios, mucho antes de que sea capaz de comprender la complejidad del modelo matemático subyacente, y desde una edad temprana.

Además, los modelos intuitivos que los estudiantes generan sirven de base para el desarrollo del conocimiento analítico posterior, puesto que ayudan a construir intuiciones correctas. Como señala Feller (1950), la intuición estadística puede ser entrenada incluso en los adultos, aunque si un niño adquiere “intuiciones erróneas” cuando es muy pequeño, esto le puede impedir más tarde la adquisición de un conocimiento adecuado. Por ello parece necesario ofrecer a los niños formación estocástica en edad temprana. Sin embargo, las actividades a desarrollar no deben escogerse al azar. Es preciso un principio de organización, en la forma de la espiral curricular, escalonada por las ideas fundamentales.

## **2.1. IMPORTANCIA DE LA CORRELACIÓN EN LA ESTADÍSTICA**

Las ideas de Heitele son retomadas por Burrill y Biehler (2011), y puestas en consonancia con la revisión del currículo de estadística en varios países; el marco de Wild y Pfannkuch (1999) sobre las características que definen el pensamiento estadístico; las características definidas para la cultura estadística, por ejemplo, por Gal (2002); y las recomendaciones del proyecto GAISE (Franklin et al., 2007). A partir de estos fundamentos, los autores definen una lista de ideas estadísticas fundamentales, que complementan las identificadas por Heitele, cuya naturaleza es fundamentalmente probabilística.

En esta lista incluyen la asociación y correlación, justificando su relevancia con los siguientes motivos (Batanero, 2001):

- Permiten modelizar las relaciones entre variables estadísticas; si las variables son cualitativas llegaremos al concepto de asociación y si son numéricas al de correlación. Complementaria a ella es la idea de regresión que permite predecir una de las variables a partir de la otra, por medio de las ecuaciones de las funciones de ajuste.
- Estas ideas extienden la dependencia funcional, donde a cada valor de una variable independiente  $X$  corresponde un solo valor de la variable dependiente  $Y$ . En la correlación y regresión para cada valor de la variable independiente tenemos una distribución de valores de la variable dependiente. Además se puede dar una medida de la intensidad de la relación por medio de diversos coeficientes (que dependen del tipo de variable y el tipo de relación).
- La regresión y correlación se pueden estudiar desde las diferentes ramas en que tradicionalmente se ha dividido la estadística. Aparecen como parte del cálculo de probabilidades, en el estudio de las variables aleatorias bidimensionales, y de la distribución normal bivariante o multivariante. Asimismo en el estudio de las variables estadísticas bidimensionales dentro de la estadística descriptiva. Por último, en la estadística inferencial se ponen en relación estos dos aspectos, utilizándose los datos de las muestras para estimar los coeficientes de correlación y regresión.
- Son base de muchos otros métodos estadísticos: en análisis multivariante la correlación es una noción fundamental, pues representa el coseno del ángulo que forman dos variables, en la interpretación geométrica de los datos multivariantes. Métodos como el análisis discriminante o el análisis factorial derivan directamente de la correlación y regresión, por lo cual, la comprensión de las anteriores será necesaria para la de éstos. Más aún, tanto el coeficiente de correlación y la proporción de varianza explicada (cuadrado de dicho coeficiente) sirven de base a los modelos de análisis de varianza univariantes y multivariantes.

## 2.2. LA CORRELACIÓN EN EL CURRÍCULO ESPAÑOL

En España sólo hay una breve mención al estudio de la asociación entre variables cualitativas en la Educación Primaria, (6 a 12 años) al referirse a la “*Lectura e interpretación de tablas de doble entrada de uso habitual en la vida cotidiana*” (MEC, 2006, p. 43099) y en la Educación Secundaria Obligatoria (12 a 16 años), también con el estudio de tablas de contingencia, dentro del tema de probabilidad (MEC, 2007a). Aunque en esta etapa se introducen las relaciones funcionales entre variables, la correlación se retrasa a la enseñanza postobligatoria (Bachillerato, 16 a 18 años). El tema se incluye en el primer curso de las modalidades de *Ciencias y Tecnología* y *Humanidades y Ciencias Sociales* (MEC, 2007b).

La modalidad de Humanidades y Ciencias de la Salud incluye el tratamiento de la correlación en el bloque 3 “*Probabilidad y estadística: [...] Distribuciones bidimensionales. Interpretación de fenómenos sociales y económicos en los que intervienen dos variables a partir de la representación gráfica de una nube de puntos. Grado de relación entre dos variables estadísticas. Regresión lineal. Extrapolación de resultados.*” (MEC, 2007b, p. 45475). Igualmente especifica los siguientes criterios de evaluación (MEC, 2007b, p. 45475-45476):

6. Distinguir si la relación entre los elementos de un conjunto de datos de una distribución bidimensional es de carácter funcional o aleatorio e interpretar la posible relación entre variables utilizando el coeficiente de correlación y la recta de regresión.

Se pretende comprobar la capacidad de apreciar el grado y tipo de relación existente entre dos variables, a partir de la información gráfica aportada por una nube de puntos; así como la competencia para extraer conclusiones apropiadas, asociando los parámetros relacionados con la correlación y la regresión con las situaciones y relaciones que miden. En este sentido, más importante que su mero cálculo es la interpretación del coeficiente de correlación y la recta de regresión en un contexto determinado.

El contenido y criterios de evaluación de la modalidad de Ciencia y Tecnología es similar, aunque descrito más someramente. En resumen, el currículo español contempla el estudio de la asociación y correlación gradualmente, llegando en el Bachillerato a bastante profundidad, lo mismo que ocurre en otros países. A pesar de ello, veremos que los resultados de las investigaciones realizadas, en su mayoría con estudiantes universitarios, indican que las dificultades de comprensión persisten finalizado el Bachillerato.

En lo que sigue nos centramos en la capacidad de estimación de la correlación a partir de diferentes representaciones.

### **3. LA ESTIMACIÓN DE LA CORRELACIÓN: TAREAS CORRELACIONALES**

Alloy y Tabachnik (1984) reducen a dos las fuentes de información relevantes para estimar el grado de correlación entre dos variables numéricas:

- Las expectativas previas o creencias. El individuo posee ideas o juicios sobre la asociación de los fenómenos tratados, debidas bien a la experiencia previa directa con dichos fenómenos en situaciones similares o bien de otras fuentes por ejemplo, la transmisión cultural o biológica.
- La información objetiva de la situación, es decir, los datos que el individuo dispone sobre la ocurrencia conjunta o las relaciones de los fenómenos de su entorno.

Teniendo en cuenta estas dos fuentes de información, McKenzie y Mikkelsen (2007) proponen un modelo bayesiano para describir el razonamiento covariacional. Indican que nuestro sistema cognitivo se aproxima a la tarea de covariación desde un marco inferencial amplio, que depende del entorno. Por

ello el ser humano utiliza el procedimiento bayesiano para evaluar la covariación, al ser sensible a las creencias previas y a la información dada por los datos. Las condiciones en que se realiza la estimación igualmente influyen en el razonamiento de los sujetos y pese a los intentos de los experimentadores de descontextualizar las tareas o de eliminar las influencias del mundo real, los sujetos recurren a hipótesis previas sobre parámetros de la tarea.

Por otro lado, al analizar el modo de proceder de un individuo cuando se enfrenta a una tarea covariacional, encontramos pasos o subtareas secuenciales en que se puede descomponer dicha resolución, que van desde la recogida de datos para estudiar la relación, hasta el uso de los resultados obtenidos para producir juicios y predicciones de interés para el investigador. Crocker (1981), distingue los siguientes pasos en dicha tarea:

- a. Decidir los tipos de datos a recolectar, es decir, la población en estudio y las variables que serán observadas;
- b. Elegir la muestra de la población de estudio; decidiendo su tamaño y método de muestreo;
- c. Diferenciar los valores de las variables en cada caso;
- d. Codificar y estimar las frecuencias de los pares de valores que se observan;
- e. Integrar resultados, por ejemplo, en un diagrama de dispersión, una tabla o cualquier otro modo de presentar la distribución bivariada;
- f. Utilizar estos resultados para estimar el grado de asociación o correlación y emitir juicios (sobre el tipo e intensidad de correlación) o bien hacer una predicción.

Zieffler (2006) resume los diferentes estudios relativos al razonamiento covariacional desarrollados en psicología, educación, didáctica de la matemática y didáctica de la estadística e indica que en la mayoría de estos estudios, los sujetos se centran en el paso (f) anterior, es decir en la realización de juicios sobre la existencia y tipo de asociación. Por otro lado, Sánchez Cobo (1999) clasifica el conjunto de tareas covariacionales en tres categorías fundamentales:

- *Juicios de asociación en tablas de contingencia.* Cuando se analiza la asociación entre dos variables cualitativas y el sujeto debe utilizar las frecuencias de dicha tabla, comparando porcentajes por filas o columnas, o bien comparando frecuencias observadas y esperadas.
- *Juicios de correlación.* Donde se trata de analizar la relación entre dos variables numéricas, a partir de diferentes representaciones de la misma; por ejemplo, a partir de un diagrama de dispersión, donde la forma o dispersión de la nube de puntos puede ser de gran ayuda para evaluar la relación entre las variables.

- *Juicios de asociación en la comparación de muestras.* Cuando se estudia la relación entre una variable cuantitativa y otra variable cualitativa; por ejemplo comparando medidas de posición central y dispersión o empleando un contraste de hipótesis.

En este trabajo nos centramos en los juicios de correlación, es decir, en realizar un juicio o estimar la correlación, una vez los datos se han integrado. En lo que sigue analizamos estas tareas y las variables que las determinan.

#### **4. VARIABLES EN LAS TAREAS CORRELACIONALES**

De entre las variables que pueden influir en la estimación de la correlación destacamos, entre otras, la forma en que se representan los datos, la explicación de la posible correlación entre las variables, el signo e intensidad de la correlación, y la concordancia entre las teorías previas y la correlación que presentan los datos. A continuación analizamos estas variables.

##### **4.1. REPRESENTACIÓN DE LA CORRELACIÓN**

La variable más influyente en el diseño de una tarea correlacional es la forma en que se proporcionan los datos, es decir, la representación que se utiliza para presentar la distribución bivariada. Las investigaciones desarrolladas por Sánchez Cobo, Estepa y Batanero (2000), Estepa (2007) y Zieffler y Garfield (2009), nos acercan un poco más a comprender el modo en que los estudiantes resuelven este tipo de situaciones problemáticas.

En concreto, la investigación abordada por Sánchez Cobo, Estepa y Batanero (2000) analiza el proceso de traducción de la correlación entre diferentes representaciones y su efecto en la estimación de la correlación así como las estrategias empleadas por los estudiantes en el desempeño de la tarea. Las representaciones que consideran cubren el lenguaje (verbal, tabular, gráfico y numérico), pero no el simbólico, ya que el objeto de este tipo de tareas es la estimación e interpretación y no la formalización mediante fórmulas de la correlación. Se describen a continuación:

- *Descripción verbal*, cuando describimos una distribución bivariada mediante el lenguaje natural. No se proporcionan datos a los estudiantes, sino que se les pide estimar la correlación de dos variables descritas verbalmente. Presentamos un ejemplo de estas tareas tomado de un libro de texto de Bachillerato donde se pide el signo y diferenciar entre dependencia funcional y aleatoria. Observamos que el estudiante tendría que conocer el fenómeno para poder pensar pares posibles de valores y poder responder la pregunta.

Para cada uno de los siguientes casos indica: Cuáles son las variables que se relacionan; Si se trata de una relación funcional o de una relación estadística y, en estos casos, el signo de la correlación:

- Renta mensual de una familia- gasto en electricidad.
- Radio de una esfera- volumen de esta.
- Litros de lluvia recogidos en una ciudad- tiempo dedicado a ver la televisión por sus habitantes
- Longitud del trayecto recorrido en una línea de cercanías- precio del billete
- Peso de los estudiantes de 1º de Bachillerato-número de calzado que usan
- Toneladas de tomate recogidas en una cosecha- precio del kilo de tomate en el mercado.

*Figura 1. Juicio de correlación a partir de descripción verbal* (Colera et al. 2008, p.238)

- *Tabla de valores*, cuando los datos se dan al estudiante en forma de tabla de frecuencia o bien como un listado de pares de valores numéricos de una distribución bivariada. Estas actividades son frecuentes en los libros de texto (ver ejemplo en la Figura 2). Generalmente, el estudiante necesitará transformar este tipo de representación en un gráfico, si quiere estimar la correlación. También podría primero calcular el coeficiente de correlación, pues es difícil visualizar la posible relación a partir de un conjunto de valores numéricos.

En la siguiente tabla se recoge la evolución del IPC (índice de precios al consumo) y el precio del barril de petróleo (*brent*) durante el segundo semestre de 2007.

IPC (%)	2,4	2,2	2,2	2,7	3,6	4,1
Precio del barril (\$)	71,54	77,01	70,73	76,87	82,50	90,16

¿Se puede asegurar que la evolución del IPC está directamente relacionada con el precio del petróleo?

*Figura 2. Estimación de la correlación a partir de listado de valores* (Vizmanos et al. 2008, p.252)

- *Diagrama de dispersión*, cuando el conjunto de pares de valores de una distribución bivariada se presentan mediante un diagrama cartesiano. En este caso, se puede observar la mayor o menor intensidad de la relación (observando la dispersión de los puntos del diagrama), así como el signo en caso de relación lineal. También la forma aproximada de la función que expresaría la relación entre las variables (por ejemplo, exponencial o parabólica). Un ejemplo muy frecuente de este tipo de tareas se muestra en la Figura 3, y otro menos común en la Figura 4, donde el estudiante deberá interpretar la correlación de cada descripción para asociarla a cada uno de los diagramas que construya.

Razona qué valor tomará el coeficiente de correlación.

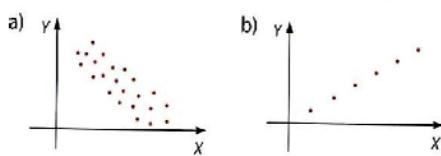


Figura 3. Estimación de la correlación desde el diagrama de dispersión (Antonio et al. 2008, p. 249)

Dibuja un ejemplo de nube de puntos que corresponda a cada uno de los siguientes casos de correlación:

- |                            |                               |                                |
|----------------------------|-------------------------------|--------------------------------|
| a) Lineal negativa débil.  | c) Lineal positiva débil.     | e) Curvilinea positiva débil.  |
| b) Lineal negativa fuerte. | d) Curvilinea negativa débil. | f) Curvilinea positiva fuerte. |

Figura 4. Elaboración de diagrama de dispersión a partir de descripción verbal (Anguera et al. 2008,

p. 223)

- *Coeficiente de correlación*, cuando se da el coeficiente de correlación existente entre las variables y se pide a los estudiantes, en forma aproximada, la forma del diagrama de dispersión o dar un juicio sobre la intensidad y el signo (ver Figura 5). En algunos textos, para facilitar la tarea, se dan varios diagramas y se pide elegir el que corresponde a una correlación dada (ver ejemplo en Figura 6)

Representa nubes de puntos que se ajusten, aproximadamente, a las rectas de regresión de Y sobre X y a los coeficientes de correlación que se indican:

- $y = -x + 4, r = -0,8$
- $y = 2x + 0,5, r = 0,6$
- $y = 0,2x + 1, r = 0,95$
- $y = x, r = 0,1$

Figura 5. Traducción de coeficiente de correlación a diagrama de dispersión (Martínez et al., 2008,

p.262)

14. Indica cuáles de las correlaciones dadas se corresponden con las nubes de puntos de las figuras y explica por qué.

- |                 |                |
|-----------------|----------------|
| a) $r = 0,7$    | b) $r = -0,98$ |
| c) $r = -0,001$ | d) $r = 1$     |

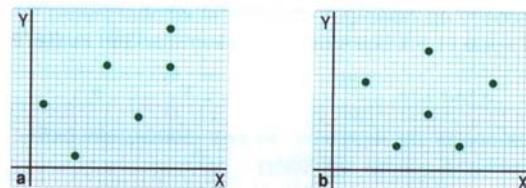


Figura 6. Elección del coeficiente de correlación a partir del diagrama de dispersión (Anguera et al.,

2008, p.229)

Sánchez Cobo (1999) hizo notar que la estimación de la correlación resulta más precisa cuando se realiza a partir de un diagrama de dispersión, así como la tarea inversa, es decir encontrar un diagrama de dispersión que se ajuste a una correlación dada numéricamente. Los estudiantes comenten un

número mayor de errores al construir una nube de puntos a partir de una descripción verbal o al estimar el coeficiente de correlación desde una tabla de valores numéricos. En cuanto a la capacidad de proponer pares de variables (en forma verbal) factibles a un coeficiente de correlación dado, aunque los estudiantes proponen variables cuya correlación tiene el mismo signo que el dado en el problema, son menos precisos con la intensidad.

El autor sugiere que parece más sencillo de comprender la correspondencia entre la intensidad de la dependencia y la dispersión de la nube de puntos o entre el signo de la correlación y la tendencia de la nube, que deducir estos datos a partir de un listado, una tabla o una descripción verbal. Por otro lado, este autor sugiere que los diagramas de dispersión son una de las representaciones que los estudiantes han usado con frecuencia en la enseñanza. Sin embargo, para adquirir un razonamiento correlacional adecuado, además de la capacidad para resolver cada una de las tareas descritas anteriormente, es de vital importancia dominar los procesos de traducción entre cada una de las representaciones que se les asocian (verbal, tabular, gráfica y numérica). La práctica de estos procesos lleva implicadas destrezas y habilidades tales como el cálculo estadístico, el uso de modelos matemáticos para el ajuste de datos o la traducción desde y hacia las expresiones algebraicas y gráficas utilizadas.

Precisamente Duval (1993) resaltó la importancia para el aprendizaje de las matemáticas del dominio de diferentes representaciones del mismo objeto matemático, aunque también advierte que dichos objetos nunca deben ser confundidos con su representación. Esto último podría ocurrir si un estudiante vincula la correlación al diagrama de dispersión, ya que podría encontrar dificultad en el desarrollo de tareas como las relativas a la descripción verbal. Aún así, una conceptualización adecuada implica la coordinación de distintas representaciones complementarias que permitan y faciliten el trabajo matemático.

## 4.2. EXPLICACIÓN DE LA CORRELACIÓN

Una segunda variable es la posible explicación de la correlación (denominada tipo de covariación por Barbancho, 1973). Una vez que encontramos una correlación entre dos variables y nos preguntamos por qué ocurre, la explicación más natural para muchas personas es que existe una dependencia causal unilateral, es decir, que una de las variables actúa como causa de la otra, como por ejemplo, si estudiamos la correlación entre el índice de paro en un país y el producto nacional bruto (que será menor a mayor paro).

La detección de relaciones causales es un componente importante de la conducta adaptativa humana, que algunos investigadores extienden a todo ser vivo, como muestra la siguiente cuestión: “*¿Bajo qué condiciones son los organismos precisos en la detección de la covariación entre sucesos?*”

(Alloy y Tabachnik, 1984, p. 113). El razonamiento causal nos permite saber cuándo dos sucesos están conectados, de modo que la presencia o ausencia de una causa siga la de un efecto (Catena, Perales y Maldonado, 2004).

Díaz y de la Fuente (2005) analizan las relaciones entre causalidad y condicionamiento, que también aparecen en un problema de correlación. Desde el punto de vista de la probabilidad, si un suceso *A* es la causa estricta de un suceso *B*, siempre que suceda *A*, sucederá *B*, por lo que  $P(B/A)=1$  y habrá una correlación entre *A* y *B*. Aunque, igualmente se puede establecer una condicionalidad en situaciones no causales, por ejemplo, en situaciones diagnósticas. En consecuencia, las nociones de causalidad, correlación y condicionamiento están relacionadas pero no son equivalentes.

Para comprobar que la asociación es causal se deben considerar los siguientes aspectos (Álvarez y Pérez, 2004): fuerza de asociación, congruencia (si la asociación ha sido reproducida por investigadores diferentes), relación temporal, gradiente dosis-respuesta (a más magnitud más intensidad del efecto), credibilidad (concordancia de la asociación con el conocimiento que se tiene en la actualidad), especificidad (genera un solo efecto). Sin embargo, en el tratamiento de tareas correlacionales, algunos estudiantes sólo consideran la existencia de correlación cuando se percibe una relación causal en las variables asociadas, confundiendo correlación y causación (Estepa, 1994), sin plantearse algunas de las siguientes razones:

- *Las variables pueden ser interdependientes* (cada variable afecta a la otra), como en el caso de la longitud de piernas y altura de una persona;
- *La existencia de una tercera variable* que determine la correlación, esto es las variables muestran dependencia pero es indirecta. Un ejemplo sería relacionar el índice de natalidad y la esperanza de vida, que presentan una correlación negativa. La proporción de mujeres que trabaja en un país afecta al producto nacional bruto, y al índice de natalidad (que disminuye) y con ello a la esperanza de vida. Así es que estas dos variables están por ello correlacionadas, de modo indirecto, ya que su relación viene motivada por la influencia de otras variables como el número de mujeres empleadas en un país.
- La *concordancia* o coincidencia en preferencia u ordenación de una misma serie de datos, como por ejemplo si dos profesores de modo independiente califican un mismo examen ya que, las calificaciones están correlacionadas pero la nota de uno no influye en la del otro.
- *Covariación casual o espúrea*: 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.

#### **4.3. OTRAS VARIABLES**

Otras variables que también influyen en la dificultad de una tarea correlacional son las siguientes (Sánchez Cobo, Estepa y Batanero, 2000):

- *El signo de la correlación*, que puede ser positivo en caso de dependencia directa, negativo en caso de dependencia inversa o nula en caso de independencia. Aunque matemáticamente la dependencia directa o inversa sean semejantes, los estudiantes no las perciben de igual modo, sino que tienen más facilidad en estimar correctamente las correlaciones positivas. En el trabajo de Erlick y Mills (1967) se expone que la asociación negativa se estima como muy próxima a cero.
- *Intensidad de la dependencia*, siendo más fácil por los estudiantes percibir una correlación fuerte. De hecho, los estudiantes presentan dificultades al comparar diferentes valores del coeficiente de correlación distintos de: -1, 0 y 1 (Sánchez- Cobo, 1999).
- *Concordancia entre los datos y las teorías previas* sugeridas por el contexto. Como veremos al hablar de los sesgos en el razonamiento correlacional, muchos sujetos se guían preferentemente por sus teorías (en vez de usar los datos) cuando estiman una correlación. Si piensan que dos variables deben estar correlacionadas tratan de encontrar una correlación entre ellas llegando a veces a sobrevalorarla, aunque, los datos en la tarea presenten poca correlación o hasta se trate de variables independientes. Por ello, es más fiable la estimación de la correlación cuando hay concordancia entre los datos y estas teorías previas que en caso contrario.
- *El número de datos*. Respecto a la tarea de estimar un valor del coeficiente de correlación, Sánchez Cobo et al., (2000) indican que mejora cuando hay más datos.

### **5. ESTRATEGIAS EN LA ESTIMACIÓN DE LA CORRELACIÓN**

En cuanto al análisis de estrategias en los juicios de asociación, diversas investigaciones (Estepa, 1993; Sánchez Cobo, 1999) llevan a cabo un estudio pormenorizado del modo en que los estudiantes resuelven una tarea correlacional, donde generalmente los datos se dan en forma de diagramas de dispersión. Clasifican las estrategias encontradas en correctas parcialmente correctas e incorrectas:

- *Estrategias correctas*: serían las estrategias que llevan siempre a una estimación correcta de la correlación entre los datos, independientemente de los datos concretos del problema. Entre ellas se encuentran las siguientes, que deben combinarse para deducir simultáneamente el signo y la intensidad de la correlación:
  - *Comparación en base a la mayor o menor dispersión del diagrama de dispersión*. Cuando el sujeto basa el juicio o la estimación de la correlación entre las dos variables observando la dispersión de los puntos en torno a una línea. Esta estrategia es correcta y permite estimar la

intensidad de la correlación, pues a menor dispersión, aumenta la correlación. Los casos extremos serían cuando no existe dispersión (dependencia funcional) y el caso de independencia (máxima dispersión en la gráfica).

- *Crecimiento o decrecimiento.* Esta estrategia permite ver el signo de la correlación en caso de que la nube de puntos se ajuste a una función lineal. Consiste en estudiar la tendencia de la nube de puntos para justificar el tipo de dependencia.
- *Estrategias parcialmente correctas:* Son estrategias que funcionan a veces, debido a los datos particulares de la tarea, pero no siempre son adecuadas. Entre ellas encontramos las siguientes:
  - *Comparación con un modelo matemático.* El estudiante compara la forma de la nube de puntos con una función conocida, por ejemplo, lineal o cuadrática. Si la forma de la nube se ajusta al modelo, deduce que hay correlación, y estima la intensidad de acuerdo al mayor o menor ajuste de los datos a éste. La estrategia funciona a veces, sobre todo si la dispersión es pequeña y dependiendo del modelo elegido para comparar. Por ejemplo, si la relación es cuadrática. Este tipo de razonamiento puede venir motivado por una confusión entre las definiciones de coeficiente de correlación y coeficiente de determinación.
- *Estrategias incorrectas:* Son estrategias que ofrecen pocos resultados aceptables, y en general vienen motivadas por no tener en cuenta todos los datos del problema. Encontramos las siguientes:
  - *Interpretación incorrecta de puntos aislados.* El estudiante basa la estimación o el juicio de correlación en pares de valores aislados. Por ejemplo, usa sólo los puntos extremos de la gráfica y con ello argumenta su aproximación.
  - *Teorías previas.* Se usan las teorías previas sobre el contexto para argumentar la correlación, en lugar de considerar los datos disponibles. Por ejemplo, si se piensa que no hay relación entre la tasa de natalidad y la esperanza de vida y se argumenta que la correlación es nula, a pesar de que el diagrama de dispersión muestre una correlación negativa fuerte.
  - *Otras variables.* Cuando la existencia de otras variables que puedan influir en la dependiente es considerada como motivo para la no existencia de asociación. Por ejemplo, estimar la correlación entre el número de partidos empatados y el puesto que ocupa en la liga un cierto equipo, el estudiante argumenta que la correlación es cero, pues el puesto depende, sobre todo, del número de partidos ganados.
  - *Rechazo de la dependencia aleatoria.* Cuando se argumenta que no existe correlación, al ser la dependencia aleatoria (haber dispersión en la nube de puntos).

Por su parte, Moritz (2004) estudió tres tipos de tareas: (a) generar un listado de datos bivariados a partir de descripciones verbales de relaciones (b) interpretar diagramas de dispersión y (c) interpretar

tablas de frecuencias de datos covariados. Encuentra una nueva estrategia errónea consistente en considerar las variables implicadas en forma asilada, en lugar de ponerlas en relación.

## 6. SESGOS Y CONCEPCIONES ERRÓNEAS SOBRE LA CORRELACIÓN

Como se ha visto al describir las estrategias erróneas, en ocasiones se observa la presencia de sesgos de razonamiento o concepciones incorrectas en los juicios de correlación o en la estimación de su signo e intensidad (Crocker, 1981; Alloy y Tabachnik, 1984). En lo que sigue describimos los estudios que se centran en la identificación de los mismos.

El sesgo que más interés ha suscitado es el denominado *correlación ilusoria* (Chapman y Chapman, 1967; 1969) que hace referencia a la presencia de errores sistemáticos producidos por variables inherentes a los estímulos que el sujeto percibe y que conllevan a juicios erróneos de covariación. Como sus autores indican:

”Correlación ilusoria” es un error sistemático en el informe de tales relaciones. Es definida según el informe dado por un observador de una correlación entre dos clases de eventos que en realidad (a) no están correlacionados, o (b) están correlacionados en menor medida que lo manifestado, o (c) están correlacionados en sentido contrario del que se manifiesta en el informe” (Chapman y Chapman, 1967, p. 194).

Chapman y Chapman (1969) describen la presencia en muchas personas de este sesgo, que consiste en formarse teorías sobre la relación entre variables, que impide evaluar correctamente las contingencias empíricas. En algunas ocasiones, el cerebro tiende a interpretar la realidad de manera sesgada, creyendo detectar relaciones que en realidad no existen. Por ejemplo, si dos eventos van frecuentemente juntos, es probable que se concluya que existe algún tipo de relación entre ellos, incluso cuando esto sea falso, e incluso se sobreestima la frecuencia de aparición conjunta de estos sucesos. Estudiaron este efecto en pruebas psicodiagnósticas como la titulada “*Dibujar una persona*”, (Chapman y Chapman, 1967), motivados por la idea de que muchas interpretaciones clínicas tienen su génesis en la correlación ilusoria. Los sujetos del estudio asociaron las personas de ojos grandes con el diagnóstico de paranoia en sus dibujos, aunque tal correlación no existe, indicando la poca utilidad de las pruebas basadas en dibujos realizados por el enfermo en la diagnosis de los trastornos mentales.

Una explicación de este sesgo es la vulnerabilidad del ser humano a creer lo que intuye o siente frente a lo que cualquier informe le pueda proporcionar (Chapman, 1967). Es por ello, que muchos autores proponen advertir a nuestros estudiantes de la dificultad que conlleva el razonamiento covariacional correcto, con el fin de sensibilizarles de las consecuencias que acarrean sus sesgos o errores (Chapman y Chapman, 1969; Moritz, 2004).

Por su parte, Engel y Sdelmeier (2011) describen los siguientes sesgos complementarios en la estimación de la correlación:

- *No tener en cuenta la existencia de variables extrañas* que pueden explicar la correlación entre dos variables dadas. Es bien conocida la paradoja de Simpson (Contreras et al., 2012), que muestra que en determinados casos se produce un cambio en la relación entre un par de variables, cuando se controla el efecto de una tercera variable. Como indican Malinas y Bigelow (2009), dependiendo de los datos con los que trabajemos, la correlación entre dos variables se puede llegar a invertir cuando la población de estudio se divide en subpoblaciones. Esto ocurre, según Blyth (1972) cuando se analiza una variable dependiente respecto a otras variables independientes en algún estudio o experimento. Incluso algunas veces una de las variables (la que hace cambiar el tipo o intensidad de correlación entre las otras) es una variable extraña o no controlada, por lo que el investigador puede no ser consciente de este efecto y llegar a una conclusión incorrecta en su estudio.
- *El efecto de regresión.* Es habitual en una situación de test- retest que los sujetos con puntuaciones atípicas en la primera prueba vuelvan hacia el valor medio en la segunda. Esto se considera por algunas personas indicativo de un cambio debido al tratamiento, cuando es una propiedad de la regresión.

En su estudio, Estepa (1994) observa una aparente inconsistencia entre estrategia y estimación de la correlación, pues algunos estudiantes dan una estimación incorrecta de la correlación con una estrategia correcta o al contrario (dan una estimación correcta con una estrategia incorrecta). Analizando estos casos con detalle, muestra que ello se debe a que el estudiante tiene concepciones erróneas sobre la correlación. A lo largo de sus trabajos describe las siguientes (Estepa, 1993; Estepa y Batanero, 1996; Batanero, Estepa y Godino, 1997; Batanero, Godino y Estepa, 1998):

- *La concepción determinista de la correlación.* Los estudiantes tienden a asignar un único valor de la variable independiente a cada uno de los valores considerados de la variable dependiente. Esto es, sólo se asume la correlación de las variables si existe una dependencia funcional, pues se confunden los dos conceptos. En casos extremos, los estudiantes exigen la existencia de una fórmula que ligue las dos variables para indicar que las variables están relacionadas; Estepa (1994) lo llama *concepción algebraica* pues el estudiante indica que no hay correlación, ya que no puede encontrar una expresión que ligue las dos variables.
- *La concepción local de la correlación.* Cuando los estudiantes utilizan parte de los datos del estudio para estimar la correlación o juzgar si existe correlación y de qué tipo. Se limitan a confirmar la relación en un subconjunto de los datos que de algún modo justifique algún tipo de

patrón, obviando la tendencia global de éstos. Por ejemplo, si al unir dos puntos del diagrama de dispersión se observa una tendencia creciente, pueden indicar que la relación de las variables es directa, sin tener en cuenta si esta tendencia se mantiene o no en otros puntos.

- *Concepción unidireccional:* En este caso el estudiante no admite la correlación inversa, considerándose la intensidad de la asociación, pero no su signo. Se presenta en los casos en los que los estudiantes consideran la dependencia inversa como independencia.
- *Concepción causal:* Cuando el sujeto sólo considera la correlación entre variables si puede adjudicarse a la presencia de una relación causal entre las mismas. Por ejemplo, si hay tercera variables que afecten la relación, piensan que no hay relación entre las dos primeras.

Otras concepciones incorrectas identificadas posteriormente son las siguientes:

- *Identificar correlación positiva y dependencia lineal.* En concreto, en el estudio de Sánchez- Cobo (1999) casi la mitad de los sujetos del estudio consideran que si existe correlación positiva, ésta se deduce en una dependencia lineal.
- *Concepción transitiva de la correlación.* Es descrita por Castro-Sotos, Van Hoof, Van den Noortgate y Onghena, (2009). El coeficiente de correlación no tiene propiedad transitiva; es decir, dos variables  $A$  y  $B$  pueden estar correlacionadas, así como  $B$  y  $C$ , sin estarlo  $A$  y  $C$ . Esta propiedad no se comprende por parte de los estudiantes, quienes generalizan indebidamente la transitividad al coeficiente de correlación.

## 7. EFECTO DE LA ENSEÑANZA

Además de los estudios de evaluación resumidos, encontramos diversas experiencias de aula diseñadas para facilitar el aprendizaje de la correlación a los estudiantes (Barret, 2000; Groth y Powell, 2004; Holmes, 2001; Shaughnessy y Pfannkuch, 2002; Horton et al., 2004; Todd et al., 2004; Obremsky, 2008). Todas se desarrollan en cursos de introducción al tema, ya sea a nivel universitario o no universitario, pero no informan sobre los resultados del aprendizaje.

El estudio más completo que encontramos sobre este tema es el de Batanero, Estepa y Godino (1997), quienes, para analizar la evolución de las concepciones, desarrollan un curso de estadística basada en el ordenador y con especial énfasis en el estudio de la asociación y correlación. En la evaluación final del aprendizaje evidencian que gran parte de los estudiantes de la muestra superaron la concepción determinista (asimilando la noción de dependencia aleatoria en los datos) y la concepción local (cuando comprenden la importancia de considerar el conjunto total de los datos), llegando a

utilizar diferentes distribuciones condicionales en lugar de procedimientos aditivos para justificar sus juicios o estimaciones.

En cuanto a la concepción unidireccional, sólo algunos estudiantes lograron superarla, no encontrándose, en general, una mejora respecto a la concepción causal. Estepa (1994) también propuso a sus estudiantes una prueba final para ser resuelta con ayuda del ordenador. Esta evaluación era abierta (en el sentido que el estudiante podía usar cualquier programa del software proporcionado) e incluyó una pregunta sobre la correlación numérica. Aunque algunos estudiantes lo resolvieron correctamente mediante el estudio del coeficiente de correlación o a partir del diagrama de dispersión, muchos otros lo trataron incorrectamente como un problema de comparación de las dos variables (y no de correlación) aplicando procedimientos inapropiados, por ejemplo, comparar promedios de las dos variables. En consecuencia, como en otras investigaciones, se deduce que el ordenador por sí sólo no es un recurso suficiente para que los estudiantes adquieran competencia en el análisis de datos bivariados. Por consiguiente, los autores plantean la necesidad de diseñar nuevas actividades que ayuden a los estudiantes a reflexionar sobre este aspecto.

En su estudio Sánchez-Cobo realiza una evaluación de la comprensión después de la enseñanza en un curso introductorio de estadística. Además de los resultados ya resumidos sobre las traducciones entre diferentes representaciones de la correlación, observa dificultades para ordenar correctamente diferentes coeficientes de correlación según su intensidad (pues los estudiantes los ordenan como si fuesen números enteros); otros estudiantes no aprecian la adimensionalidad de la correlación o confunden el coeficiente de correlación y el coeficiente de determinación (el cuadrado del coeficiente de correlación).

Por su parte, Zieffler (2006) pretende caracterizar el desarrollo del razonamiento covariacional de los estudiantes y estudiar factores explicativos de dicho razonamiento. Lo describe con la metáfora de “modelo cuadrático” porque, aunque los estudiantes muestran inicialmente un gran avance en su razonamiento covariacional, con el tiempo decrece su intensidad, incluso retrocediendo:

Estos resultados podrían sugerir la presencia de una saturación en nuestros estudiantes en cuanto al razonamiento sobre datos bivariados que provoca una decadencia que impide su avance durante el curso (Zieffler, 2006, p. 86).

Así es, que la mayor parte del cambio en el desarrollo de este razonamiento se produce al comienzo de la enseñanza formal del tópico. Por ello, Zieffler y Garfield (2009) sugieren prestar más atención al desarrollo de razonamiento sobre datos univariados, con el fin de proporcionar una base sólida al desarrollo posterior del razonamiento sobre datos bivariados. Aun así, no se observó una influencia significativa de la secuenciación del tema sobre el desarrollo del razonamiento covariacional, aunque sí se plantea su influencia en otros tópicos como la inferencia, que presenta como posibles líneas de

investigación futura. Al respecto, Sánchez Cobo (1999) defiende la secuenciación de trabajar en primer lugar la correlación y luego abordar la regresión. Además, este autor aboga por:

“que no se haga una aproximación al concepto de variable estadística bidimensional a partir de dos variables estadísticas unidimensionales, como se efectuaba en algunos manuales de bachillerato, ya que podría favorecer el que los alumnos sean poco conscientes de que dichas variables unidimensionales tienen que estar referidas a la misma unidad estadística, como ha puesto de manifiesto los resultados de esta investigación” (Sánchez Cobo, 1999, p. 291).

Finalmente resaltamos la investigación de Estepa (2007) donde sólo algo más de la mitad de los participantes son capaces de dar una definición adecuada del coeficiente de correlación. Otro porcentaje importante indican que la correlación se aplica sólo para relacionar dos conjuntos de datos tomados de diferentes muestras, sin comprender que se puede aplicar para correlacionar dos variables tomadas sobre los sujetos de la misma muestra.

## 8. CONCLUSIONES

Aunque el estudio de la competencia en la estimación de la correlación ha tenido poca relevancia en didáctica de la matemática, encontramos una amplia literatura en psicología que describe el razonamiento covariacional como una competencia fundamental para la toma de decisiones. El análisis de las estrategias intuitivas de los estudiantes en la estimación de la correlación a partir de diversas representaciones muestra, por un lado, la existencia de numerosas estrategias incorrectas; por otro, la de concepciones incorrectas y sesgos en la detección de la correlación, que a veces se mantienen tras la enseñanza (Batanero, Estepa y Godino, 1997). Todo ello nos aporta información sobre las posibles dificultades de los estudiantes que los futuros profesores debieran conocer.

Algunos estudiantes muestran una concepción local, realizando su estimación a partir de sólo una parte de los datos. Moritz (2004) sugiere enseñar a los estudiantes a leer progresivamente punto a punto hasta una posterior generalización con los datos disponibles. Otra ayuda sería la utilización de un enfoque de variación temporal que permita a los estudiantes centrarse en el cambio de una variable a lo largo del tiempo para una posterior correspondencia entre variables no temporales.

Es también interesante alentar la expresión de las creencias y teorías previas que serían poco a poco equilibradas mediante la información proporcionada por el estudio. Para ello conviene utilizar tareas que impliquen un razonamiento covariacional contraintuitivo donde, se cuestione de modo natural, la fiabilidad del conjunto de datos de que se dispone.

Shaughnessy y Pfannkuch (2002), resaltan la importancia de situar la estadística en contexto y propiciar el proceso de transnumeración (Wild y Pfannkuch, 1999), donde el estudiante debe interpretar los datos y trabajar con ellos en diferentes representaciones para obtener conclusiones y hacer predicciones. En este sentido, se puede mostrar a los estudiantes algunos ejemplos reales cómo la

paradoja de Simpson, que ha llevado a conclusiones equivocadas en la investigación. Un ejemplo, reciente (Saari, 2001) se produjo cuando en 1999 el Estado de California estableció un sistema de primas para los profesores de los colegios públicos en los que se lograra mejorar el rendimiento de los estudiantes. La evaluación de los centros mostró (con satisfacción de las autoridades educativas) que el rendimiento de los estudiantes había mejorado en la mayoría de los colegios, tanto para el grupo racial dominante, como para la minoría (hispanos). Sin embargo, al mezclar los datos de los dos grupos, el rendimiento global bajó en casi todos los centros, debido a una variación de la proporción de hispanos en los centros a lo largo del periodo.

Finalmente estamos de acuerdo con Lavalle et al. (2006), quienes sugieren adelantar la enseñanza, que podría comenzar entre el segundo y el tercer año de secundaria (14 a 15 años), cuando ya los estudiantes conocen los conceptos de función, y función lineal. En principio, se podría hacer un tratamiento intuitivo, ampliando lo aprendido sobre distribuciones univariantes con el estudio descriptivo y la representación gráfica de datos bivariados. Luego se llevaría a cabo una aproximación a la correlación lineal a través del análisis de gráficos y, por último, se incorporarían los cálculos referidos al coeficiente de correlación y determinación.

## **AGRADECIMIENTOS**

Proyecto EDU2010-14947, FPI-BES-2011-044684 (MICINN-FEDER) y grupo FQM126 (Junta de Andalucía).

## **REFERENCIAS**

- Alloy, L. B. y Tabachnik, N. (1984). Assessment of covariation by humans and animals: the joint influence of prior expectations and current situational information, *Psychological Review*, 91(1), 112-149.
- Álvarez, H. y Pérez, E. (2004). Causalidad en medicina. Historia y filosofía de la medicina. *Gaceta Médica de México*, 140(4), 467-472.
- Anguera, J., Biosca, A., Espinet, M. J., Fandos, M.J., Gimeno, M. y Rey, J. (2008). *Matemáticas I aplicadas a las Ciencias Sociales*. Barcelona: Guadriel.
- Antonio, M., González, L., Lorenzo, J. Molano, A., del Río, J., Santon, D. y de Vicente, M. (2008). *Matemáticas aplicadas a las Ciencias Sociales I*. Madrid: Santillana.
- Barbancho, A. G. (1973). *Estadística elemental moderna*. Barcelona: Ariel.
- Barret, G. B. (2000). The Coefficient of Determination: Understanding  $r^2$  y  $R^2$ . *Mathematics Teacher*, 93 (3), 230-234.
- Batanero, C. (2001). *Didáctica de la Estadística*. Granada: Departamento de Didáctica de la Matemática.
- Batanero, C., Estepa, A. y Godino, J. D. (1997). Evolution of students' understanding of statistical association in a computer based teaching environment. En J. B. Garfield y G. Burrill, (Eds.),

*Research on the Role of Technology in Teaching and Learning Statistics. IASE Round Table Conference Papers* (pp. 191-205). Voorburg, The Netherlands: International Statistical Institute.

- Batanero, C., Godino, J. D. y Estepa, A. (1998). Building the meaning of statistical association through data analysis activities. En A. Olivier y K. Newstead (Eds.), *Proceedings of the 22nd Conference of the International Group for the Psychology of Mathematics Education* (v.1, pp. 221-236). University of Stellenbosch.
- Bruner, J. S. (1960). *The process of education*. Cambridge, MA: Harvard University Press.
- Burrill, G., y Biehler, R. (2011). Fundamental statistics ideas in the school curriculum and in training teachers. En C. Batanero, G. Burrill y C. Reading (Eds.). *Teaching Statistics in School Mathematics- Challenges for Teaching and Teacher Education. A Joint ICMI/IASE Study* (pp. 57-70). New York: Springer.
- Blyth, C. R. (1972). On Simpson's paradox and the sure-thing principle. *Journal of the American Statistical Association*, 67, 364-366.
- Castro-Sotos, A. E., Vanhoof, S., Van Den Noortgate, W. y Onghena, P. (2009). The transitivity misconception of Pearson's correlation coefficient. *Statistics Education Research Journal* 8 (2), 33-55. Disponible en: [www.stat.auckland.ac.nz/~iase/serj/](http://www.stat.auckland.ac.nz/~iase/serj/).
- Catena, A., Perales, J. C. y Maldonado, J. A. (2004). Judgment frequency effects in generative and preventative causal learning. *Psychological* 25, 67-85.
- Chapman, L. J. (1967) Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior*, 6 (1), 151-155.
- Chapman, L. J. y Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations, *Journal of Abnormal Psychology*, 72 (3), 193-204.
- 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 (3), 271-280.
- Colera, J., Oliveira, M. J., García, R. y Santaella, E. (2008). *Matemáticas aplicadas a las Ciencias Sociales I*. Madrid: Anaya.
- Contreras, J. M., Batanero, C., Cañadas, G. y Gea, M. M. (2012). La paradoja de Simpson. *SUMA*, 71, 27-34.
- Crocker, J. (1981). Judgment of covariation by social perceivers, *Psychological Bulletin* 90 (2), 272-292.
- Díaz, C. y de la Fuente, E. I. (2005). Razonamiento sobre probabilidad condicional e implicaciones para la enseñanza de la estadística. *Épsilon*, 59, 245-260.
- Duval, R. (1993). *Semiosis et Noesis. Lecturas en Didáctica de la Matemática: Escuela Francesa*. México: Sección de Matemática Educativa del CINVESTAV-IPN.
- Engel, J. y Sedlmeier, P. (2011). Correlation and regression in the training of teachers. En C. Batanero, G. Burrill, and C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education: A Joint ICMI/IASE study* (pp. 247-258). New York: Springer.
- Erlick, D. E. y Mills, R. G. (1967). Perceptual quantification of conditional dependency, *Journal of Experimental Psychology*, 73 (1), 9-14.
- Estepa, A. (1993). *Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores*. Tesis doctoral no publicada. Universidad de Granada.

- Estepa, A. (2004). Investigación en Educación Estadística. La asociación estadística. En R. Luengo (Ed.). *Líneas de investigación en Educación Matemática* (pp. 227-255). Badajoz: Servicio de Publicaciones. Universidad de Extremadura.
- Estepa, A. (2007). Caracterización del significado de la correlación y regresión de estudiantes de Educación Secundaria, *Zetetiké*, 15(28), 119-151.
- Estepa, A. y Batanero, C. (1996). Judgments of correlation in scatter plots: Students' intuitive strategies and preconceptions. *Hiroshima Journal of Mathematics Education*, 4, 25-41.
- Feller, W. (1950). *An introduction to probability theory and its applications*. Vol. I. New York: John Wiley.
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., y Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report*. Alexandría: American Statistical Association.
- Gal, I. (2002). Adult's statistical literacy. Meanings, components, responsibilities. *International Statistical Review*, 70 (1), 1-25.
- Groth, R.E y Powell, N.N. (2004). Uso de los proyectos de investigación para ayudar al desarrollo del pensamiento estadístico en los estudiantes. *Mathematics Teacher*, 97 (2), 106-109.
- Heitele, D. (1975). An epistemological view on fundamental stochastic ideas. *Educational Studies in Mathematics*, 6, 187-205.
- Holmes, P. (2001). Correlation: From Picture to formula. *Teaching Statistics*, 23 (3), 67-71.
- Horton, M.R. Phillips, V y Kenelly, J. (2004). Construir tu propio modelo de regresión. *Mathematics Teacher*, 97 (4), 284-289.
- Lavalle, A. L., Micheli, E. B. y Rubio, N. (2006). Análisis didáctico de regresión y correlación para la enseñanza media. *RELIME*, 9 (3), 383-406.
- Malinas, G. y Bigelow, J. (2009). Simpson's paradox. En N. Edward (Ed.), *The Stanford Encyclopedia of Philosophy*. Online: <http://plato.stanford.edu/archives/fall2009/entries/paradox-simpson>.
- Martínez, J. M., Cuadra, R., Heras, A. (2008). *Matemáticas aplicadas a las Ciencias Sociales. 1º Bachillerato*. Madrid: Autor.
- Mckenzie, C.R.M., y Mikkelsen, L.A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology*, 54 (1), 33-61.
- MEC (2006). *REAL DECRETO 1513/2006, de 7 de diciembre, por el que se establecen las enseñanzas mínimas de la Educación primaria*. Madrid: Autor.
- MEC (2007a). *REAL DECRETO 1631/2006, de 29 de diciembre, por el que se establecen las enseñanzas mínimas correspondientes a la Educación Secundaria Obligatoria*. Madrid: Autor.
- MEC (2007b). *REAL DECRETO 1467/2007, de 2 de noviembre, por el que se establece la estructura del bachillerato y se fijan sus enseñanzas mínimas*. Madrid: Autor.
- Moritz, J. (2004). Reasoning about covariation. En D. Ben-Zvi y J. Garfield (Eds.). *The Challenge of developing statistical literacy, reasoning and thinking* (pp.221-255). Dordrecht (The Nederlands): Kluwer.
- Pérez Echeverría, M. P. (1990). *Psicología del razonamiento probabilístico*. Madrid: Ediciones de la Universidad Autónoma de Madrid.
- Saari, D. (2001). *Decisions and elections. Explaining the unexpected*. Cambridge University Press.

- Sánchez Cobo, F. T. (1999). *Significado de la correlación y regresión para los estudiantes universitarios*. Tesis doctoral no publicada. Universidad de Granada.
- Sánchez Cobo, F. T., Estepa, A. y Batanero, C. (2000). Un estudio experimental de la estimación de la correlación a partir de diferentes representaciones. *Enseñanza de las Ciencias*, 18 (2), 297-310.
- Shaughnessy, J. M. (1992). Research in probability and statistic: Reflections and directions. En D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465-494). New York: Macmillan Publishing Company.
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. En F. Lester (Ed.), *Second handbook of research on mathematics Teaching and Learning* (pp. 957-1010). Greenwich, CT: Information Age Publishing y NCTM.
- Shaughnessy, J. M. y Pfannkuch, M. (2002). How faithful is Old Faithful? Statistical thinking: A story of variation and prediction. *Mathematics Teacher*, 95 (4): 252-259.
- Shaughnessy, J. M., Garfield, J., y Greer, B. (1996). Data handling. En A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, y C. Laborde (Eds.), *International handbook of mathematics education* (Vol. 1, pp. 205-237). Dordrecht: Kluwer.
- Todd, J.G. Hagtvedt, R y Jones, K. (2004). A VBA-based Simulation for Teaching Simple Linear Regression. *Teaching Statistics*, (26) 2, 36-41.
- Vizmanos, J. R., Hernández, J. Alcaide, F., Moreno, M. y Serrano, E. (2008). *Matemáticas aplicadas a las Ciencias Sociales I*. Madrid: Grupo SM.
- Wild, C. J. y Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67 (3), 223-265.
- Zieffler, A. S. (2006). *A longitudinal investigation of the development of college students' reasoning about bivariate data during an introductory statistics course*. Tesis doctoral no publicada. Universidad de Minnesota.
- Zieffler, A. S. y Garfield, J. (2009). Modeling the growth of students' covariational reasoning during an introductory statistics course. *Statistics Education Research Journal*, 8 (1), 7-31.