

EXPLORANDO EL RAZONAMIENTO COVARIACIONAL MEDIANTE UN AMBIENTE COMPUTACIONAL EN UN CURSO INTRODUCTORIO DE ESTADÍSTICA

Exploring the Covariational Reasoning through a Computer Environment in an Introductory Statistics Course

Inzunsa, S.^a y Ward, S. E. ^b

^aUniversidad Autónoma de Sinaloa, ^bUniversidad Pedagógica de Sinaloa

Resumen

Se reportan resultados de un estudio con 34 estudiantes universitarios que tomaban un curso introductorio de estadística en el cual se utilizó un ambiente computacional para la enseñanza de la covariación. Los resultados muestran que los estudiantes lograron identificar con éxito la dirección de la relación entre variables y desarrollaron un buen sentido de la intensidad al interpretar diagramas de dispersión y el coeficiente de correlación, pero concedieron menos importancia a la forma, agrupamientos y puntos extremos. La interpretación de los coeficientes de la ecuación de regresión y el coeficiente de determinación les resultó complicada a muchos estudiantes y algunos tuvieron dificultades para identificar la variable explicativa y de respuesta.

Palabras clave: *correlación, regresión, razonamiento covariacional, tecnología computacional*

Abstract

Results of a study with 34 university students taking an introductory statistics course in which a computer environment for the teaching of covariation was used are reported. The results show that students were able to successfully identify the direction of the relationship between variables and developed a good sense of intensity to interpret scatterplots and correlation coefficient, but gave less importance at the clusters and outliers. The interpretation of the coefficients of the regression equation and the coefficient of determination were difficult to many students and some had difficulty to identify the explanatory and response variable.

Keywords: *correlation, regression, covariational reasoning, computer technology.*

INTRODUCCIÓN

El razonamiento que se involucra en la interpretación y aplicación de la correlación y la regresión se denomina razonamiento covariacional o razonamiento sobre datos bivariados. El razonamiento sobre datos bivariados involucra saber cómo juzgar e interpretar una relación entre dos variables. En el caso de variables cuantitativas —situación a la que se hace referencia en este estudio—, la relación es representada comúnmente mediante un diagrama de dispersión, el cálculo del coeficiente de correlación y la recta de regresión cuando el ajuste de los puntos es factible.

Una comprensión profunda de la covariación requiere del dominio de diversas ideas y conceptos, tales como la estructura de una relación bivariada, su rol en la construcción de modelos y en la predicción de eventos; por lo que no debe sorprender que los estudiantes tengan dificultades para comprender los conceptos involucrados en la covariación (Garfield, Joan y Ben-Zvi, 2008). En este sentido, las nociones estadísticas de correlación y regresión resultan de gran ayuda en cuanto a discernir si dos variables se relacionan, con qué intensidad o en qué sentido lo hacen; y cuando sea

posible, construir un modelo para predecir una variable con base en la información de otras variables.

La tecnología computacional ofrece un gran potencial de cálculo para realizar los procesos de análisis de datos que tradicionalmente han caracterizado a la estadística. En el caso de la covariación, el cálculo del coeficiente de correlación y la determinación de la línea de regresión —dos de los procedimientos más laboriosos de un análisis de datos— no sólo pueden ser realizados de forma automática y precisa con la tecnología, sino que también pueden ser visualizados en forma dinámica mediante diversas representaciones y explorar su comportamiento cuando algún dato o parámetro de interés es modificado. En este contexto, en el presente trabajo nos hemos planteado las siguientes preguntas: ¿qué estrategias y recursos del ambiente computacional ponen en juego los estudiantes para dar sentido a la información de datos bivariados?, ¿qué factores influyen en el desarrollo del razonamiento covariacional en un ambiente computacional con representaciones dinámicas de los datos?

ANTECEDENTES DE INVESTIGACIÓN

En la literatura se reportan estudios sobre los conceptos que se involucran en la covariación desde el nivel elemental hasta el universitario. En este contexto destacan los trabajos de los investigadores españoles, quienes han estudiado la covariación de forma consistente desde hace varios años, analizado el significado que le atribuye la institución de los autores de libros de texto de secundaria, bachillerato y universitarios (e. g., Sánchez Cobo, 1999; Gea, Batanero, Cañadas y Contreras, 2013). Los resultados de estas investigaciones han servido de punto de partida para el diseño de secuencias de aprendizaje por parte de investigadores que se proponen mejorar la enseñanza del tema y profundizar en las dificultades de comprensión que se presentan en el proceso de enseñanza y aprendizaje.

El efecto de la tecnología computacional en las estrategias utilizadas por estudiantes y profesores, así como la comprensión de los diversos conceptos que se involucran en la covariación ha sido objeto de estudio en algunos de estos trabajos (e.g., Estepa, 1993; Batanero, Estepa y Godino, 1997; Batanero, Gea, Díaz y Cañadas, 2014). Entre los resultados obtenidos es importante destacar el uso de estrategias incorrectas para estimar la correlación y concepciones erróneas sobre algunos conceptos que se involucran en la covariación, como es el caso de la concepción unidireccional de la correlación, la cual consiste en percibir la relación entre variables cuando sólo cuando es positiva. Otra concepción recurrente fue considerar que la correlación entre dos variables implica causalidad. Destaca también la dificultad de muchos sujetos de estudio para distinguir entre variable de respuesta y variable explicativa y para razonar con relaciones que son negativas.

Zieffler (2006) realizó un estudio longitudinal durante un semestre para analizar el efecto de dos secuencias de enseñanza con apoyo de tecnología y con diferente orden en el tema de datos bivariados (antes y después de la inferencia estadística) en el desarrollo del razonamiento sobre covariación en un curso introductorio de estadística con estudiantes universitarios en los Estados Unidos. Los resultados muestran que la secuenciación de datos bivariados dentro del curso parece no estar asociada a los cambios en el desarrollo del razonamiento. El mayor cambio en el desarrollo del razonamiento ocurrió entre la primera y segunda evaluación de las cuatro realizadas a lo largo del semestre, justo antes de que los datos bivariados fueran enseñados por métodos formales, lo que podría significar que el razonamiento sobre datos bivariados está cercanamente relacionado al razonamiento sobre datos univariados, y que no se requiere una enseñanza formal del tema para su comprensión.

Fitzallen (2012) en su investigación con estudiantes de primaria y secundaria australianos, explora el pensamiento y el razonamiento de los estudiantes acerca de la covariación y la influencia del software Tinkerplots para determinar relaciones entre variables e identificar tendencias mediante

recursos gráficos como el diagrama de dispersión. Los resultados muestran que Tinkerplots proporciona un poderoso ambiente de aprendizaje para apoyar a los estudiantes en la comprensión de la covariación desde niveles tempranos.

MARCO TEÓRICO

El tipo de representaciones utilizadas en el análisis de datos incorporan elementos que resultan complejos de interpretar por muchos estudiantes. En este sentido, Ben-Zvi y Arcavi (2001) señalan que el análisis requiere poner en juego visiones locales y globales de los datos; una comprensión local de los datos significa enfocarse en algún valor individual (o en algunos) dentro de un conjunto de datos (por ejemplo, la entrada de una tabla o un punto de una gráfica), mientras que la comprensión global requiere de habilidades para buscar, reconocer, describir y explicar patrones generales en un conjunto de datos (por ejemplo, tendencias, variabilidad, forma) ya sea a través de la visualización o por medio del cálculo de medidas estadísticas. El razonamiento covariacional como un caso específico del razonamiento estadístico tiene una liga muy importante con la interpretación de tablas, gráficas y expresiones que expresan la relación entre datos bivariados.

En el análisis del razonamiento estadístico sobre un tema en particular resulta útil establecer categorías jerárquicas que tengan en cuenta la cantidad de conceptos y la forma como los estudiantes establecen relaciones entre ellos y el contexto, al hacer argumentaciones basadas en datos. Un modelo que se utiliza con frecuencia en educación estadística para establecer jerarquías de razonamiento o niveles de comprensión es el modelo SOLO (Structure of Observed Learning Outcomes) desarrollado por Biggs y Collis (1982). El modelo consta de los niveles preestructural, univariado, multivariado, relacional y abstracto extendido, los cuales van en orden ascendente de complejidad en el razonamiento.

La tecnología computacional en el análisis de datos no solo representa un recurso amplificador de la capacidad de cálculo y construcción de representaciones gráficas, sino que constituye a su vez un recurso con potencial reorganizador de la mente de los usuarios, con capacidad para provocar cambios estructurales en el sistema cognitivo a través de la organización y transformación de las diversas representaciones de los datos y la visualización de patrones que emergen al realizar cambios en algunas de las componentes de dichas representaciones o en los datos mismos. Lo anterior le otorga a la tecnología computacional un estatus de herramienta cognitiva en el sentido establecido por Pea (1987, p.91), quien la define como “cualquier medio que ayuda a trascender las limitaciones de la mente, en el pensamiento, el aprendizaje y las actividades de resolución de problemas”.

Se han propuesto modelos para el diseño de ambientes de aprendizaje que promueven el desarrollo del razonamiento estadístico, en los cuales la tecnología computacional es un importante elemento de articulación. Ejemplo de ello es el modelo propuesto por Cobb y McClain (2004) cuyos principios instruccionales se resumen a continuación:

- Enfocarse en el desarrollo de ideas estadísticas centrales en lugar del énfasis en procedimientos de cálculo.
- Utilizar datos reales y motivadores para comprometer a que los estudiantes realicen y prueben conjeturas.
- Utilizar actividades en el salón de clases para apoyar el desarrollo del razonamiento de los estudiantes.
- Integrar el uso apropiado de herramientas tecnológicas que permitan a los estudiantes probar sus conjeturas, explorar y analizar datos y desarrollar su razonamiento estadístico.

- Promover un discurso en el salón de clases que incluya argumentos estadísticos e intercambios que se enfoquen en ideas estadísticas significativas.

METODOLOGÍA

Los sujetos de estudio fueron 34 estudiantes (18-19 años) que tomaban un curso introductorio de estadística en el primer grado de la carrera de Informática. Sus antecedentes sobre el tema eran prácticamente nulos, lo cual quedó de manifiesto en un cuestionario diagnóstico aplicado antes de las sesiones de enseñanza. La investigación tuvo lugar en un aula de cómputo con 35 computadoras durante 6 sesiones de una hora en la parte final del bloque de análisis de datos, por lo que los estudiantes recién habían visto el análisis de datos univariados donde utilizaron el software Fathom para el cálculo de medidas descriptivas y construcción de gráficas.

En el diseño de las actividades de enseñanza se utilizó un paquete de diapositivas que el profesor investigador preparó con los conceptos que se contemplan en la covariación y siguiendo los principios instruccionales sugeridos por Cobb y McClain (2004). Las actividades contemplaban conjuntos de datos reales multivariados que se tomaron de algunos sitios de internet, y se pedía a los estudiantes que exploraran pares de variables para analizar las características de la relación (forma, dirección, intensidad, datos extremos, agrupamientos). Como punto de partida del análisis se construyeron diagramas de dispersión y se enfocó la atención de los estudiantes a visualizar sus características, posteriormente se calcularon coeficientes de correlación buscando conectar su valor con las características del diagrama. Finalmente, se introdujo la recta de regresión para ajustar la tendencia de los puntos y su ecuación, con el valor del coeficiente de determinación que por omisión proporciona el software.

Los datos fueron recogidos en la actividad final de la secuencia de enseñanza, pero previamente los estudiantes habían resuelto una actividad de estimación del coeficiente de correlación utilizando un applet que proporcionaba un diagrama de dispersión como punto de partida, en el cual hacían la estimación y luego verificaban el resultado con la opción de cálculo que el mismo applet proporcionaba. La otra actividad fue muy similar a la que se presenta en este artículo, en un contexto de variables medidas a un conjunto de tablets. Decidimos reportar los datos de la última actividad porque fue la que más conceptos contemplaba y era la más completa de las tres.

En el análisis de la información generada al resolver cada actividad se tuvieron en cuenta el uso de representaciones y las características de la relación entre variables identificadas visualmente de un diagrama de dispersión, así como medidas numéricas como los coeficientes de correlación y de determinación. Para la interpretación de diagramas se definieron niveles de razonamiento basados en el modelo SOLO.

RESULTADOS Y DISCUSION

El propósito de esta actividad era explorar diversos aspectos del razonamiento covariacional que habían desarrollado los estudiantes. Los datos analizados provenían de diversas variables que fueron medidas a 80 automóviles compactos que se vendieron en el mercado mexicano en 2014 (ver Figura 1). Las unidades del rendimiento están dadas en kilómetros por litro de gasolina, la emisión de CO₂ en gramos por kilómetro, el gasto en pesos por año, la potencia en caballos de fuerza y el precio está expresado en pesos.

1. Con las variables que se muestran construye dos diagramas con relación negativa y otros dos con relación positiva. Interpreta cada uno de ellos.

La construcción de los diagramas de dispersión fue realizada con éxito por todos los estudiantes, sin embargo no todos utilizaron las mismas representaciones y estrategias para resolver la tarea, algunos se basaron sólo en la tendencia de la nube de puntos, otros agregaron la recta de regresión

para tener mayor certeza del signo y la intensidad de la relación y otros más recurrieron al cálculo del coeficiente de correlación (ver Tabla 1).

	MARCA	MODELO	RENDIMIENTO_CIUADAD	EMISION_CO2	GASTO_ESTIMADO_COMBUSTIBLE	POTENCIA	PRECIO
1	HONDA	CIVIC HÍB...	23.8	119	8700	90	266900
2	MAZDA	MAZDA3...	15.4	159	13400	153	236900
3	MAZDA	MAZDA3...	15.7	157	13200	153	236900
4	MAZDA	MAZDA3...	15	162	13800	153	258900
5	MAZDA	MAZDA3...	15.7	158	13200	153	258900
6	HONDA	CITY LX ...	13.2	182	15700	118	219000

Figura 1. Variables de 80 modelos de vehículos compactos vendidos en México en 2014

Tabla 1. Representaciones utilizadas para resolver la tarea

Tipo de representación	Frecuencia
Diagrama de dispersión	20
Diagrama de dispersión y recta de mínimos cuadrados	5
Diagrama de dispersión y coeficiente de correlación	9
Total	34

Con base las representaciones y las características de la relación entre variables que fueron utilizadas, clasificamos la interpretación de los diagramas de dispersión en tres categorías del modelo SOLO como se describe a continuación (ver Tabla 2).

Tabla 2. Niveles del modelo SOLO en la interpretación de diagrama de dispersión

Niveles de interpretación	Frecuencia
Uniestructural	18
Multiestructural	15
Relacional	1
Total	34

Nivel uniestructural: Se emplea una sola característica del diagrama de dispersión para describir la relación entre las dos variables. Aun cuando agregan o calculan algún elemento adicional al diagrama (por ejemplo la recta de regresión y el coeficiente de correlación), solo dependen de una representación para el análisis.

Nivel multiestructural: Se emplea más de una propiedad del diagrama de dispersión para describir la relación entre las dos variables. Se apoyan en el diagrama y algún elemento o cálculo adicional para la interpretación.

Nivel relacional: Se emplean múltiples propiedades del diagrama de dispersión para describir la relación entre las dos variables. Se apoyan en el diagrama y elementos o cálculos adicionales para la interpretación. Además hacen referencia a otros elementos o al contexto para describir o justificar la relación.

La única respuesta ubicada en el nivel relacional fue proporcionada por Jesús Enrique (ver Figura 2): “en este diagrama observamos una asociación negativa con una relación lineal fuerte que se ve a simple vista, se puede observar un dato atípico en la parte más baja de la emisión de CO₂, esta gráfica nos muestra que entre mayor sea el rendimiento del automóvil la emisión de CO₂ será menor”.

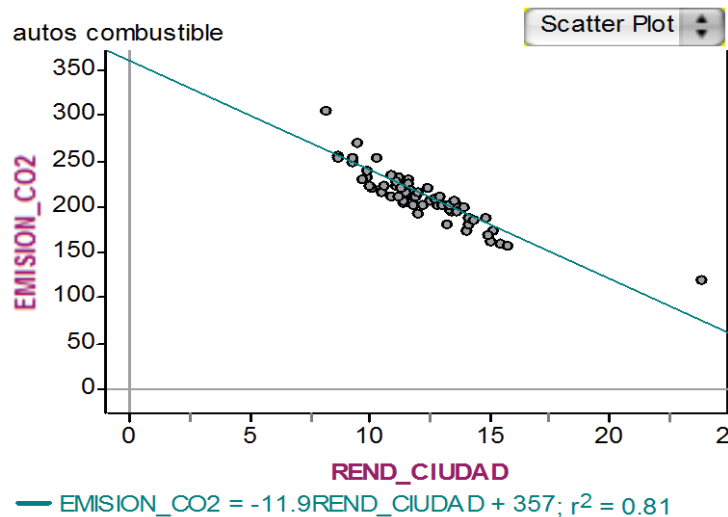


Figura 2. Diagrama de dispersión con recta de regresión construido por Jesús Enrique

2. Considera la variable precio como variable de respuesta y a las demás variables como variables explicativas. Determina la variable que tiene mayor influencia o relación con el precio. Explica las razones de tu elección.

Un total de 32 estudiantes establecieron correctamente que la potencia era la variable que estaba más relacionada con el precio. La principal estrategia consistió en construir los cuatro diagramas de dispersión, pero se observaron varias variantes como se muestra en la Tabla 3.

Tabla 3. Estrategias utilizadas por los estudiantes para resolver la tarea

Tipo de representación	Frecuencia
Diagrama de dispersión	22
Diagrama de dispersión con recta de mínimos cuadrados	6
Diagrama de dispersión y coeficiente de correlación	6
Total	34

Los estudiantes dependieron más de estrategias visuales que de estrategias de cálculo para responder la tarea; tal fue el caso de quienes utilizaron sólo el diagrama de dispersión o combinado con la recta de regresión. Estos estudiantes observaron las cuatro nubes de puntos y seleccionaron la que les parecía que tenía una mayor relación con el precio, no dependieron de ningún cálculo aun cuando la recta de regresión estaba ligada con valor del coeficiente de determinación r^2 . Un caso representativo de estos estudiantes es José Efraín (ver Figura 3) quien argumenta: “los datos de la potencia son los que están más agrupados en relación con el precio ya que la línea de regresión pasa por mayor cantidad de datos que las demás gráficas que se encuentran más dispersos en relación al precio”. El estudiante relaciona el agrupamiento de datos alrededor de la línea de regresión con la intensidad de la relación, pero no hace referencia al coeficiente de determinación que aparece en la parte inferior del diagrama; también hace referencia a que la línea de regresión pasa por la mayor cantidad de puntos, lo cual no es condición obligada de la recta de regresión.

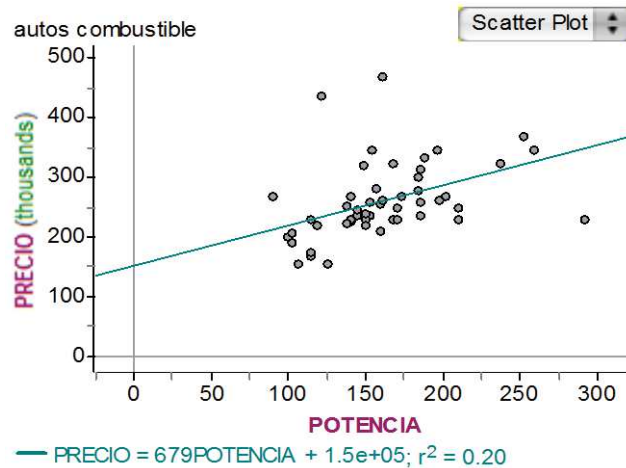


Figura 3. Diagrama de dispersión construido por José Efraín

En sus justificaciones 21 estudiantes se basaron en la dirección de la nube de puntos, señalando sólo que a mayor potencia el precio era mayor, lo que los ubica en un nivel de razonamiento uniestructural. Otros 13 estudiantes además de la dirección, complementaron sus justificaciones con otras propiedades como la intensidad (ver Tabla 4) por lo que fueron ubicados en un nivel multiestructural.

Tabla 4. Niveles del modelo SOLO en la determinación de la variable de mayor influencia

Niveles de interpretación	Frecuencia
Uniestructural	21
Multiestructural	13
Relacional	0
Total	34

3. Considera el rendimiento como variable explicativa y la emisión de CO₂ como variable de respuesta. Construye un diagrama de dispersión y construye la recta de regresión.

a) Explica un poco sobre la precisión de la recta para la predicción de valores de CO₂.

La mitad de los estudiantes dieron explicaciones sobre la relación entre precisión y predicción en las que mostraron una comprensión intuitiva de la relación entre estos conceptos, atendiendo principalmente a criterios visuales que les proporcionaban los diagramas de dispersión. Por ejemplo, Hugo responde: “la precisión de la recta es bastante acertada, la mayoría de los puntos están cerca de la recta, sólo un punto está bastante alejado de los demás, la predicción sobre los valores sería algo acertada porque para un próximo valor de la gráfica se pudiera saber qué es lo que seguiría”. La otra mitad emitió argumentos más bien relacionados con la interpretación del diagrama y no sobre lo que se les preguntaba. Adicionalmente se observó que 7 estudiantes tuvieron dificultades para ubicar correctamente la variable explicativa y la variable de respuesta en los ejes correspondientes.

b) Explica qué representa en el contexto del problema el valor de -11.9 que aparece en la expresión matemática de la recta de regresión de la parte inferior de la gráfica (ver Figura 2).

Este inciso fue el más complicado de todos, pues sólo 16 estudiantes señalan correctamente que representa la pendiente de la recta, pero no dan una explicación de su significado en el contexto de los datos, es decir, que por cada kilómetro por litro que se incrementa el rendimiento, la emisión de CO₂ disminuirá -11.9 gramos. Otros 11 estudiantes dejaron en blanco el inciso, y el resto respondieron que se trataba del coeficiente de correlación o el error.

CONCLUSIONES

Los estudiantes utilizaron estrategias que privilegiaban el uso de representaciones gráficas sobre las representaciones simbólicas, no obstante lo sencillo que resultaba calcular con el software el coeficiente de correlación y de determinación. La aparente sencillez de un diagrama de dispersión queda en entredicho con las respuestas que han dado los estudiantes, pues sólo uno de ellos logró ubicarse en el nivel relacional del modelo SOLO en dos preguntas que involucraban la construcción e interpretación de un diagrama de dispersión. Si bien, la dirección y la intensidad fueron características que los estudiantes identificaron y estimaron bastante bien, dejaron de lado características importantes como los datos extremos y la forma del diagrama. Además hizo falta conectarlas con el contexto, dando cuenta con ello de una visión más local que global de los datos.

En cuanto a la regresión, algunos estudiantes tuvieron dificultades para distinguir la variable de respuesta de la variable explicativa, lo cual concuerda con otros estudios como el realizado por Estepa (1993). Las respuestas de algunos estudiantes muestran que lograron apreciar en forma intuitiva la relación entre la variación y la precisión de una estimación en la variable de respuesta, pero faltó relacionarla con la suma de cuadrados y el coeficiente de determinación. Las mayores dificultades las tuvieron al interpretar los coeficientes de la ecuación de regresión en el contexto de los datos y en valorar la bondad del ajuste.

No obstante el reducido tiempo del estudio y su carácter exploratorio, el presente trabajo aporta algunos elementos para considerar que el amplio recurso representacional dinámico e interactivo de los ambientes computacionales como Fathom, son un factor importante en el diseño de secuencias para el aprendizaje de la covariación. A conclusiones similares han llegado otros estudios como el realizado por Fitzallen (2012), quien ha utilizado un ambiente computacional con características similares en el estudio de la covariación con estudiantes de nivel medio.

Los factores que parecen tener mayor influencia en el desarrollo del razonamiento covariacional en un ambiente computacional, son el dinamismo de las representaciones y su simultaneidad en la pantalla, la interactividad que permite manipular de forma directa las representaciones o alguna parte de ellas, así como el contexto de los datos utilizados que motiva el interés de los estudiantes para explorar sus relaciones.

Referencias

- 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: Proceedings of the 1996 IASE Roundtable Conference*. (pp. 198-212). Voorburg, Los Países Bajos: International Statistical Institute.
- Batanero, C., Gea, M., Díaz, C. y Cañadas, G. (2014). Building high school pre-service teachers' knowledge to teach correlation and regression. En K. Makar, B. de Sousa y R. Gould (Eds.). *Proceedings of the Ninth International Conference on Teaching Statistics* (pp. 1-6). Voorburg, Los Países Bajos: International Statistical Institute.
- Ben-Zvi, D. y Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics*, 45(1), 35-65.
- Biggs, J. y Collis, K. (1982). *Evaluating the Quality of Learning: the SOLO taxonomy*. New York: Academic Press.
- Cobb, P. y McClain, K. (2004). Principles of Instructional Design for Supporting the Development of Students' Statistical Reasoning. En Ben-Zvi, D. y Garfield, J. (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking* (pp. 375-396). Dordrecht, Los Países Bajos: Kluwer.
- 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. Universidad de Granada.

- Fitzallen, N. (2012). *Reasoning about covariation with Tinkerplots*. Tesis Doctoral. Universidad de Tasmania.
- Garfield, J., Joan, B. y Ben-Zvi, D. (2008). *Developing Students' Statistical Reasoning. Connecting Research and Teaching Practice*. Dordrecht, Los Países Bajos: Springer.
- Gea, M. M., Batanero, C., Cañadas, G. R. y Contreras, J. M. (2013). Un estudio empírico de las situaciones-problema de correlación y regresión en libros de texto de bachillerato. En A. Berciano, G. Gutiérrez, A. Estepa y N. Climent (Eds.), *Investigación en Educación Matemática XVII* (pp. 293-300). Bilbao: SEIEM
- Pea, R. (1987). Cognitive Technologies for Mathematics Education. En A. Schoenfeld (Ed.) *Cognitive Science and Mathematics Education* (pp. 89-122). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sánchez Cobo, F.T. (1999). *Significado de la correlación y regresión para los estudiantes universitarios*. Tesis doctoral. Universidad de Granada.
- Zieffler, S. A. (2006). *A longitudinal investigation of the development of college students' reasoning about bivariate data during an introductory statistics course*. Tesis doctoral. University of Minnesota. Consultado en Diciembre de 2014, <http://iase-web.org/documents/dissertations/06.Zieffler.Dissertation.pdf>