

INTERPRETANDO LA CORRELACIÓN

Carmen Batanero, María M. Gea, Rafael Roa, Pedro Arteaga y Gustavo R. Cañadas

Universidad de Granada (España)

batanero@ugr.es, mmgea@ugr.es, rroa@ugr.es, parteaga@ugr.es, grcanadas@ugr.es

Palabras clave: Correlación, enseñanza, interpretación, bachillerato, universidad

Key words: Correlation, teaching, interpretation, high school, university

RESUMEN: El escrito trata de proporcionar al profesor información que le ayude a la mejor planificación de la enseñanza y a la evaluación de sus estudiantes. Se consideró la representación de datos bidimensionales, variables de las tareas de correlación, dificultades de los estudiantes y recursos tecnológicos. Dicha información se proporcionó de forma práctica, involucrando a los asistentes en la resolución de tareas que pueden proponer a sus estudiantes y en el posterior análisis didáctico de las mismas.

ABSTRACT: In this workshop, held during the conference, we tried to provide the teacher with information useful for the planning of teaching and the assessment of students. We considered the representations of bidimensional data, variables in correlation tasks, students' difficulties and technological resources. This information was provided in practical way, involving participants in the solution of tasks they can propose to their students and in the didactical analysis of the same.

■ INTRODUCCIÓN

El estudio de los datos bivariantes extiende la dependencia funcional a fenómenos caracterizados por la incertidumbre y es muy útil en el estudio de las relaciones entre variables, que es frecuente en la ciencia, tecnología y gestión. Explicar, controlar y predecir los sucesos que se presentan en nuestro día a día depende de habilidades y destrezas para detectar covariaciones entre variables, es decir, en la percepción de la correlación (Alloy y Tabachnik, 1984). En consecuencia, el razonamiento correlacional se encuentra presente en la vida cotidiana del ser humano, como actividad cognitiva fundamental para su vida diaria (Moritz, 2004; Zieffler, 2006; McKenzie y Mikkelsen, 2007), aunque no por ello está exenta de dificultades.

En este taller nos hemos centrado en datos estadísticos cuantitativos, cuya relación se modeliza mediante la correlación y regresión. Este tema se incluye en España en el primer curso de Bachillerato (estudiantes de 16 y 17 años), tanto en la especialidad de Ciencias y Tecnología, como en la de Humanidades y Ciencias Sociales. Aparece en las orientaciones curriculares de España, promulgadas por el Ministerio de Educación y Ciencia (MEC, 2007), con contenido similar en ambas modalidades, que se describe en la forma siguiente: “Distribuciones bidimensionales. Relaciones entre dos variables estadísticas. Regresión lineal” (MEC, 2007: 45449). También se especifica; “Se pretende comprobar la capacidad para estimar y asociar los parámetros relacionados con la correlación y la regresión con las situaciones y relaciones que miden” (MEC, 2007: 45450). En forma muy similar se contempla en el nuevo decreto de Educación Secundaria y Bachillerato (MECD, 2015).

La enseñanza del tema no es simple, pues se basa en la comprensión de múltiples conceptos como los de variable estadística y distribución bidimensional; diferentes tipos de frecuencias; dependencia estadística, funcional e independencia; covarianza y correlación; regresión, variable dependiente e independiente, modelo y modelo lineal, bondad de ajuste y coeficiente de determinación. La investigación sobre el tema, por otro lado, sugiere que se pueden presentar dificultades, tanto por el efecto de las creencias previas sobre la estimación de la correlación, como por los errores en la interpretación de los conceptos implicados en su estudio (Engel y Sedlmeier, 2011).

El objetivo principal del taller fue sensibilizar y motivar a los profesores sobre la problemática de una buena enseñanza, que permita el aprendizaje correcto de la correlación por parte de sus alumnos. Para ello se plantean dos objetivos específicos:

1. Proporcionar ejemplos de tareas que potencien el tratamiento de la correlación en el aula con una triple finalidad: a) favorecer el dinamismo de la enseñanza y el interés de los estudiantes; b) proporcionar a los profesores una actividad de evaluación para detectar y diagnosticar algunos de los sesgos descritos en la investigación desarrollada sobre correlación y c) ayudar a concienciar a los estudiantes de estos sesgos permitiendo así desarrollar en ellos un adecuado razonamiento correlacional.
2. Presentar a los participantes los principales resultados de las investigaciones sobre el desarrollo del razonamiento de los estudiantes en correlación en forma resumida y sencilla.

A continuación se describen los diferentes apartados del taller y las actividades llevadas a cabo.

■ REPRESENTACIONES DE DATOS BIDIMENSIONALES

Las investigaciones desarrolladas por Sánchez Cobo, Estepa y Batanero (2000) y Estepa (2007) junto con la de Zieffler (2006), nos acercan un poco más a comprender el modo en que los alumnos adquieren la noción de correlación. Estas investigaciones consideran cuatro formas de representar la correlación:

1. Descripción verbal, cuando describimos una distribución bivariada mediante el lenguaje natural; por ejemplo, indicamos que entre la talla y el peso de las personas hay una correlación alta y positiva.
2. Tabla de valores, o presentación de un conjunto de pares de valores numéricos de una distribución bivariada; en el ejemplo anterior, proporcionaríamos una tabla con datos de talla y peso de un grupo de personas.
3. Diagrama de dispersión, cuando el conjunto de pares de valores de una distribución bivariada se presentan mediante un diagrama cartesiano.
4. Coeficiente de correlación, cuando se da un valor numérico como medida de asociación existente entre las variables que conforman la distribución.

Para adquirir un razonamiento correlacional adecuado es de vital importancia dominar los procesos de traducción entre estas representaciones. Es decir, el alumno ha de saber traducir entre la tabla de datos, sus representaciones gráficas, las descripciones verbales sobre la correlación y el valor del coeficiente.

Por ello se comenzó el taller analizando las representaciones más usuales de datos bivariantes que se utilizan en los libros de texto para este nivel educativo (Gea, Batanero, Arteaga, Cañadas y Contreras, 2014). A continuación, se hizo observar a los participantes la ventaja de cada tipo de representación y se plantearon algunas tareas de traducción entre representaciones, analizando la actividad matemática requerida del estudiante en cada una de ellas y posibles dificultades que podrían encontrar.

■ VARIABLES EN LAS TAREAS DE ESTIMACIÓN DE LA CORRELACIÓN

Un modo de comenzar la enseñanza de la correlación y regresión en los cursos de Bachillerato, es pedir a los estudiantes que estimen la intensidad de la asociación a partir de los diagramas de dispersión. Un estudio de este tipo se realiza en Estepa (2008), cuyos resultados resumimos a continuación:

- El 90% de los estudiantes leen correctamente las coordenadas de los puntos;
- El 73% de los estudiantes discriminan la dependencia funcional de la aleatoria, siendo más fácil de detectar la dependencia funcional lineal;
- La detección de la correlación positiva depende de la intensidad de la correlación, no ocurre igual con la dependencia negativa;
- El ajuste correcto de una línea recta a un diagrama de dispersión depende del gráfico y de la intensidad de la correlación;

Es por ello importante tener en cuenta las siguientes variables, que afectan a la dificultad de las tareas de correlación (Sánchez Cobo et al., 2000):

1. Signo de la correlación entre las variables, pues podemos considerar tres casos: Dependencia directa, dependencia inversa e independencia. Aunque matemáticamente la dependencia directa e inversa se analizan de igual modo, con el coeficiente de correlación o la covarianza, psicológicamente no son percibidas de la misma forma por los estudiantes. Para resaltar ese hecho, Estepa (1994) definió la concepción unilateral o unidireccional de la correlación como un sesgo de algunos estudiantes que consideran la dependencia inversa como independencia.
2. Intensidad de la dependencia, pues la dependencia se percibe más fácilmente si es intensa. Algunos estudiantes equiparan una correlación pequeña con la independencia.
3. Concordancia entre los datos y las teorías previas sugeridas por el contexto. Muchos sujetos se guían preferentemente por sus teorías (en vez de usar los datos) cuando analizan la posible correlación entre dos variables. Chapman y Chapman (1967) describen un razonamiento común, que denomina correlación ilusoria y que fue propuesto para designar la correlación que perciben los observadores entre dos clases de sucesos que o bien (a) no están correlacionados, (b) se correlacionan con un menor grado del que se declara, o (c) se correlacionan en la dirección opuesta de la que se declara.
4. Tipo de covariación. Barbancho (1973) indica que la correlación entre dos variables puede deberse a una dependencia causal unilateral, donde una de las variables es una causa y la otra un efecto. Pero también puede deberse a una interdependencia (cada variable afecta a la otra), una dependencia indirecta (una tercera variable afecta a otras dos), concordancia (coincidencia en preferencia u ordenación de la misma serie de datos por dos jueces) o covariación casual. Pero algunos estudiantes sólo admiten la relación causal, confundiendo correlación y causación (concepción causal, según Estepa, 1994).

Para informar a los asistentes sobre la importancia de estas variables se propusieron algunas tareas de estimación de la correlación, pidiéndoles identificar el signo de la correlación, estimar sus intensidad y decidir el tipo de función de ajuste (en caso de correlación moderada o alta). Seguidamente se realizó un análisis de las principales variables implícitas en las tareas propuestas y de la forma en que estas variables permiten graduar la dificultad de estas tareas cuando queremos utilizarlas en la evaluación de los estudiantes (Gea, Batanero, Cañadas, Arteaga, y Contreras, 2013).

En el taller se pidió, además, interpretar los tipos de covariación que podrían explicar dicha correlación, diferenciando la relación de correlación y de causalidad.

■ ANÁLISIS DE LAS DIFICULTADES DE LOS ESTUDIANTES

La investigación sobre razonamiento correlacional se encuentra ampliamente desarrollada en el ámbito de la Psicología, donde se ha evidenciado la presencia de errores o sesgos en la estimación de la correlación al existir en los sujetos expectativas o esquemas referidos a los estímulos presentes en la situación a que se enfrentan.

Para informar a los profesores de estas investigaciones, en este apartado del taller se trató de identificar las dificultades potenciales de los estudiantes en las tareas propuestas anteriormente.

Además de recoger las propuestas por los profesores y observadas en su experiencia didáctica, se proporcionó información sobre concepciones incorrectas y sesgos en la interpretación de la correlación descritas en la investigación.

En este taller se ha tenido en cuenta la siguiente bibliografía sobre correlación y regresión, presentando ejemplos de los errores más comunes:

1. Estudios sobre razonamiento correlacional, que describen su importancia como actividad cognitiva fundamental del ser humano (Moritz, 2004; Zieffler, 2006; McKenzie y Mikkelsen, 2007). Algunos trabajos indican sesgos en este razonamiento, como la correlación ilusoria, donde los sujetos crean sus propias teorías sobre la correlación entre dos variables, sin tener en cuenta la correlación en los datos (Chapman y Chapman, 1967). Otro sesgo descrito es el efecto de la regresión, donde en investigaciones experimentales pueden confundirse efectos del tratamiento con la tendencia de la variable a acercarse a su media, en dos medidas consecutivas de la misma magnitud (Engel y Sedlmeier, 2011).
2. Destrezas requeridas para traducir datos bivariados de representación numérica o descripción verbal a gráfica o tabular, y viceversa (Moritz, 2004; Sánchez Cobo, et al., 2000). Algunos estudiantes tienen dificultades para traducir entre diferentes representaciones, sobre todo partiendo de la descripción verbal y coeficiente de correlación, cuando se les pide cambiar a otro tipo de representación.
3. Concepciones incorrectas sobre la correlación: Estepa (1994) describe las siguientes concepciones incorrectas sobre la correlación: a) concepción determinista, que consiste en aceptar sólo la dependencia funcional; b) concepción local, propia de los estudiantes que tienden a pensar que se puede medir la correlación sólo con parte de los datos; c) concepción unidireccional, cuando no se acepta como correlación la correlación inversa; y d) concepción causal, que consiste en confundir correlación y causalidad. Estas concepciones permanecen después de la enseñanza si no se tienen en cuenta en la misma (Batanero, Godino y Estepa, 1998).

■ RECURSOS TECNOLÓGICOS

Los currículos actuales en España (MEC, 2007; MECD, 2015) recomiendan el uso de la tecnología en la enseñanza de la estadística, por la ventaja que supone en el cálculo y representación gráfica, el trabajo con datos reales y el aprendizaje de conceptos a través de la simulación. Hoy en día existe una gran variedad de recursos tecnológicos como la calculadora, hoja de cálculo, applets y programas de ordenador específicos, que pueden facilitar la realización de cálculos y gráficos. El aprendizaje de y a través de la tecnología es esencial en esta etapa educativa, sobre todo para desarrollar este tema, y se incluye como objetivo general en bachillerato: “Utilizar con solvencia y responsabilidad las tecnologías de la información y la comunicación” (MEC, 2007, p.45382).

El uso de la tecnología en la enseñanza de la estadística ha sido reconocido, entre otros, por Pratt, Davies y Connor (2011), que destacan la reducción del tiempo de cálculo y la ampliación del tipo de gráficos que el alumno puede realizar interactivamente. Igualmente señalan la posibilidad de trabajar con proyectos, en los cuales alumno parte de un problema de investigación y completa

todos los pasos de una investigación (Wild y Pfannkuch, 1999). Por este motivo se finalizó el taller describiendo algunos recursos para el estudio de la correlación.

Entre ellos destacamos los conjuntos de datos reales, que hoy día son accesibles desde muchas instituciones en Internet, y que permiten plantear a los estudiantes, proyectos de investigación como el descrito en Batanero, Díaz y Gea (2011). Una gran ventaja al utilizar estos datos es que se potencia la interdisciplinariedad en clase de estadística, permitiendo aprender contenidos que no se adquieren habitualmente con problemas tomados de los libros de texto; por ejemplo, el efecto de valores atípicos sobre el cálculo de un estadístico (Hall, 2011). Al facilitar el cálculo y la representación gráfica, la tecnología disminuye el problema tradicional en la enseñanza de la estadística en cuanto al desfase entre la comprensión de los conceptos y los medios técnicos de cálculo para poder aplicarlos (Batanero y Díaz, 2011).

También se mostraron las potencialidades de las utilidades de Excel para el estudio de la correlación y regresión, así como algunos applets disponibles en Internet, que se describen con más detalle en Gea, Díaz-Levicoy, López-Martín y Cañadas (2015). Estos recursos permiten visualizar la correlación y sus diferentes tipos, así como observar el efecto de los puntos atípicos sobre el valor del coeficiente de correlación. Asimismo, hacen asequible a los estudiantes el criterio de mínimos cuadrados, al representar las desviaciones de los puntos a la recta de regresión y comprobar su efecto sobre la suma de cuadrados. En algunos de ellos es posible ajustar a ojo la recta de regresión para comparar con la recta de mínimos cuadrados. En otros casos es posible realizar tareas de estimación y traducción de la correlación, que pueden ser también de gran ayuda para la formación de profesores y estudiantes.

■ CONSIDERACIONES FINALES

La preparación del profesor de matemáticas para enseñar estadística es una preocupación actual entre las sociedades de profesores y las sociedades de investigación en educación estadística. Es por ello que los investigadores tratan de acercarse al profesor para informarle de los resultados de la investigación educativa y ayudarles a prepararse para la docencia. Al mismo tiempo, las reflexiones de los profesores y la interacción con los mismos durante estos talleres amplían el conocimiento del investigador sobre el trabajo en el aula.

Esperamos que el tipo de tareas sugeridas en este taller, junto con la bibliografía comentada, pueda ser útil a los profesores responsables de impartir el tema de correlación y regresión. Recomendamos al profesor, asimismo, incluir algunas de estas tareas en el trabajo con sus alumnos, para ayudarles a desarrollar un razonamiento correlacional potente que les ayude en la toma de decisiones.

En este sentido es importante reforzar el uso de la tecnología, teniendo en cuenta la alta idoneidad didáctica del uso de recursos tecnológicos en la enseñanza, sobre todo desde la componente afectiva, pues estos recursos son muy motivadores para el estudiante. Además, al facilitar el aprendizaje y potenciar la creatividad se refuerza la componente cognitiva, al poder ampliar las aplicaciones del tema, lo que lleva a conectar las matemáticas con otras materias.

Agradecimiento: Proyecto EDU2013-41141-P (MEC) y grupo FQM126 (Junta de Andalucía).

■ REFERENCIAS BIBLIOGRÁFICAS

- 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.
- Barbancho, A.G. (1973). *Estadística elemental moderna*. Barcelona: Ed. Ariel. (Cuarta edición, 1.975).
- Batanero, C. y Díaz, C. (2011). *Estadística con proyectos*. Universidad de Granada.
- Batanero, C., Díaz, C. y Gea, M. M. (2011). Estadísticas de la pobreza y desigualdad. En C. Batanero y C. Díaz (Eds.), *Estadística con proyectos* (pp. 97-124). Granada: Departamento de Didáctica de la Matemática.
- Batanero, C., Godino, J. D., & Estepa, A. (1998). Building the meaning of statistical association through data analysis activities. In A. Olivier y K. Newstead, (Eds.), *Proceedings of the 22nd Conference of the Internacional Group for the Psychology of Mathematics Education* (pp. 221-236). Stellenbosch, South Africa: Universidad de Stellenbosh.
- Chapman, L.J. & Chapman, J.P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology* 72 (3), 193-204.
- Engel, J. & Sedlmeier, P. (2011). Correlation and regression in the training of teachers. In C. Batanero, G. Burrill, & 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.
- Estepa, A. (1994). *Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores*. Tesis doctoral no publicada. Universidad de Granada.
- Estepa, A. (2007). Caracterización del significado de la correlación y regresión de estudiantes de Educación Secundaria. *Zetetiké*, 15 (28), 119-151. Recuperado de <http://www.fae.unicamp.br/zetetike/viewissue.php?id=5>
- Estepa, A. (2008). Interpretación de los diagramas de dispersión por estudiantes de bachillerato. *Enseñanza de las Ciencias*, 26 (2), 257-271.
- Gea, M. M., Batanero, C., Cañadas, G. R., Arteaga, P. y Contreras, J. M. (2013). La estimación de la correlación: Variables de tarea y sesgos de razonamiento. En A. Salcedo (Ed.) *Educación Estadística en América Latina: Tendencias y Perspectivas*. Caracas (361-384). Universidad Central de Venezuela.
- Gea, M. M., Batanero, C., Arteaga, P., Cañadas, G. R., y Contreras, J. M. (2014). Análisis del lenguaje sobre la correlación y regresión en libros de texto de bachillerato. *SUMA*, 76, 47-45.
- Gea, M., Díaz-Levicoy, D., López-Martín, M. M. y Cañadas, G. R. (2015). Recursos virtuales para el estudio de la correlación y regresión. Presentado en las *17 Jornadas Nacionales de Aprendizaje y la Enseñanza de las Matemáticas* (JAEM). Cartagena. Julio de 2015: Federación Española de Profesores de Matemáticas.

- Hall, J. (2011). Engaging teachers and students with real data: benefits and challenges. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education. A joint ICMI and IASE study* (pp. 335-346). New York: Springer.
- MEC (2007). *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.
- MECD (2015). *Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato*. España: MECD.
- Mckenzie, C.R.M. & Mikkelsen, L.A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology*, 54 (1), 33-61.
- Moritz, J. (2004). Reasoning about covariation. In D. Ben-Zvi y J. Garfield (Eds.), *The Challenge of developing statistical literacy, reasoning and thinking*, (pp. 221-255). Dordrecht (The Netherlands): Kluwer.
- Pratt, D., Davies, N. & Connor, D. (2011). The role of technology in teaching and learning statistics. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education. A joint ICMI and IASE study* (pp. 97-107). New York: Springer.
- 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.
- Wild, C. J., & 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. EUA: Universidad de Minnesota.