NÚMEROS
Revista de Didáctica de las Matemáticas
http://www.sinewton.org/numeros
ISSN: 1887-1984
Volumen 81, noviembre de 2012, páginas 5-14

Algunas notas históricas sobre la correlación y regresión y su uso en el aula

Antonio Estepa Castro (Universidad de Jaén. España)
María Magdalena Gea Serrano (Universidad de Granada. España)
Gustavo R. Cañadas de la Fuente (Universidad de Granada. España)
José Miguel Contreras García (Universidad de Granada. España)

Fecha de recepción: 22 de abril de 2012
Fecha de aceptación: 17 de mayo de 2012

Resumen Los currículos actuales nos aconsejan introducir los conocimientos matemáticos a partir de situaciones reales donde intervengan con sentido los objetos matemáticos a estudiar. Una fuente importante de casos reales la proporciona la historia, donde podemos encontrar las situaciones que dieron origen al descubrimiento de los objetos matemáticos e identificar algunas dificultades en su desarrollo que podrían reproducirse en los estudiantes. En este trabajo se analizan brevemente algunos hechos que dieron lugar a la creación de las nociones de correlación y regresión y se hace una reflexión general sobre los posibles usos de la historia de la matemática en la enseñanza.

Palabras clave Correlación, regresión, Historia de la Matemática, uso en el aula, obstáculos.

Abstract Current curricula suggest introduce the mathematical concepts through real situations where these concepts are used in a meaningful way. The history of mathematics provides an important source of real examples, since we can find the situations that originate the discovering of mathematical objects and identify some difficulties in their development that may be reproduced in the students' learning. In this paper we summarize some facts that lead to the creation of correlation and regression and reflect on the possible uses of the history of mathematics in the classroom.

Keywords Correlation, regression, history of mathematics, its use in the classroom, obstacles.

1. Introducción

Entre las nociones estadísticas fundamentales, cuya enseñanza debe optimizarse, se encuentran las de correlación y regresión. Desde la prehistoria hasta nuestros días, el discernimiento sobre la posible relación que puede existir entre dos sucesos ha sido un aspecto importante del conocimiento humano. "*Conocer si los sucesos se relacionan y, con qué intensidad lo hacen, facilita a las personas explicar el pasado, controlar el presente y predecir el futuro*" (Crocker, 1981, p.272). De estas palabras se desprende el valor que tiene por parte de los ciudadanos el dominio de las nociones de correlación y regresión.

En esta dirección podemos destacar los importantes avances que desde diversas disciplinas y ocupaciones del mundo actual, como por ejemplo la Economía, Dirección de Empresas, Estadística ó Sociología, se han llevado a cabo en cuanto al estudio de la toma de decisiones. Dentro de este ámbito, una destreza importante es la realización de juicios sobre la existencia o inexistencia de asociación entre variables (Alloy y Tabachnik, 1984), lo que nos lleva al núcleo del estudio de la correlación y regresión. Una comprensión correcta de la misma es un prerrequisito básico para garantizar la comprensión de muchos otros conceptos y procedimientos estadísticos. Podemos añadir que la mayoría de los trabajos didácticos sobre correlación y regresión (por ejemplo: Estepa y Batanero, 1996; Castro et al., 2009) señalan la dificultad que tienen los estudiantes en el estudio de estos temas, en consecuencia, debemos realizar esfuerzos con el fin de ayudar a disminuir dichas dificultades.

Un punto de arranque imprescindible en toda investigación es el estudio de resultados en áreas de conocimiento afines y aunque las investigaciones didácticas sobre la correlación y regresión son escasas, el tema ha sido objeto de gran interés en Psicología, debido en gran medida a su implicación en la toma de decisiones, dado el comportamiento de un individuo en situaciones de incertidumbre como la puesta en práctica de su racionalidad. Los resultados de estas investigaciones muestran que una concepción correcta de la noción de asociación estadística (que incluye como caso particular la correlación y regresión) no siempre se adquiere espontáneamente, incluso alcanzada la edad adulta. Asimismo, las investigaciones psicológicas muestran los numerosos sesgos en las estrategias empleadas para estimar la correlación o detectar la existencia de la misma y el peso de las teorías previas sobre la interrelación entre variables sobre los juicios emitidos, esto es, el peso de la "correlación ilusoria" (Chapman y Chapman, 1969).

En lo que sigue presentamos algunas notas sobre el origen de los problemas que dieron sentido a las nociones de correlación y regresión, debida, como veremos más adelante, principalmente a Galton, pero también a otros autores que hicieron posible el estudio de la interrelación entre variables aleatorias, y por tanto extienden la idea de dependencia funcional, introduciendo modelos matemáticos en gran número de ciencias como la Biometría o la Psicometría que, hasta entonces no habían contado con un método científico especialmente adaptado a las mismas. A continuación, ofreceremos observaciones de interés didáctico, en cuanto a la enseñanza de estas nociones, atendiendo al eje transversal de la Historia de las Matemáticas.

2. Algunas notas históricas sobre la formación de las nociones de correlación y regresión

La formación de las nociones de correlación y regresión proviene, en gran parte, de estudios realizados en Biología, Biometría y Eugenesia. El primer autor que se interesa en el tema fue Lambert-Adolphe-Jacques Quetelet (1796-1874), conocido como Adolphe Quetelet, nacido en Gante, Bélgica. Obtuvo su doctorado en Matemáticas con una tesis sobre secciones cónicas, llegando a ser director del observatorio astronómico de Bruselas. Fue un hombre de gran energía, entusiasmo y talento organizativo que utilizó para crear varias instituciones internacionales.

Sus aportaciones sobre la correlación y regresión se originan desde sus estudios sobre el hombre medio, estimando empíricamente las medias y desviaciones típicas de medidas antropométricas que, suponía, dependen de varias variables independientes tales como el sexo, edad, profesión o nivel de educación. En sus estudios relaciona dos o más variables, por ejemplo llega a obtener una ecuación de una hipérbola que relaciona la edad y la altura de las personas entre cero y 30 años (Hald, 1998). Su originalidad no consistió en haber calculado las medias de las magnitudes antropométricas, sino haber considerado su dispersión y descubierto que la ley normal (bien conocida en Astronomía) ofrecía una descripción aceptable de tal variabilidad, por lo que utilizó esta distribución como ajuste a sus medidas antropométricas, introduciéndola en Biometría (Seal, 1967).

Augusto Bravais (1811-1863) contribuye al desarrollo de esta teoría desde otro campo: la astronomía, al estudiar los errores en las medidas de las coordenadas de cuerpos espaciales. Fue él quien utilizó por primera vez el término *correlación* en un estudio presentado en 1846 en la Academia de Ciencias en Francia (Seal, 1967). Sin embargo Pearson (1965) indicará que Bravais, al estudiar la teoría de errores, no consideraba variables aleatorias correlacionadas, sino consideraba errores independientes unos de otros; por tanto, no llegó a una verdadera idea de la correlación, tal como hoy la conocemos.

Un avance en el desarrollo de estos conceptos se produce mediante el estudio conjunto de la variación de dos medias realizado por Francis Galton (1822-1911). Según Hald (1998), Galton era hijo de un banquero de Birmingham y se casó con una prima de Charles Darwin, por lo que se interesó por sus estudios sobre la herencia. Estudió Medicina y Matemáticas en Londres y Cambridge, dedicando una primera parte de su vida a estudios geográficos y meteorológicos, llegando a tener un papel destacado en los mapas del tiempo diarios del periódico "The Times". Tenía una destacada facilidad para construir artificios mecánicos, habilidad que utilizó para construir sofisticados aparatos de medida. Tuvo obsesivo interés por medir, hacer recuentos y gráficos de fenómenos de antropología, biología, genética, sociología y psicología. A lo largo del periodo 1865-1890, su principal interés fueron los estudios empíricos de las leyes de la herencia por medio de métodos estadísticos.

Desde el punto de vista de la Estadística matemática se puede considerar a Galton como un ingenioso amateur, ya que, sin conocer los refinados métodos estadísticos de la época (usados, por ejemplo, por Laplace y Gauss) y por medio de investigaciones empíricas, estudia la variabilidad de características humanas. Desarrolla sus propios y rudimentarios métodos para describir observaciones univariadas y bivariadas normalmente distribuidas, explicando la utilidad y el significado de la regresión y correlación, no solamente en el contexto de la herencia, sino en general. Galton no conocía la literatura estadística germana y para ajustar una distribución normal a sus datos utiliza el método de Quetelet, muy simple desde el punto de vista numérico, ya que, requiere solamente el cálculo de frecuencias relativas y la interpolación en la tabla de la binomial acumulativa. Como no domina con soltura la matemática de su tiempo utiliza artificios mecánicos para "probar" las propiedades de la distribución binomial como el que llamó "quincux", hoy conocido como aparato de Galton.

Darwin propuso a Galton estudiar algún método que pudiese dar soporte a su teoría de la evolución, tratando de comparar las características físicas de los hijos con las de sus padres, pues si estos caracteres se heredan se confirmaría su hipótesis de que las características de los sujetos mejor adaptados pasarían de una a otra generación; a la vez sería necesario una variabilidad dentro de los hijos de los mismos padres (para poder pensar en sujetos mejor y peor adaptados) (Benzecri, 1982). Galton acepta el reto, pero, al no tener suficientes datos humanos diseñó un experimento con semillas de guisantes para estudiar la distribución de los pesos de las semillas en dos generaciones (Hald, 1998). Observó que la distribución de los pesos era normal, seleccionó siete grupos (clasificando los guisantes por su peso, que era igual al peso medio global, más o menos 0, 1, 2, 3, desviaciones típicas), conteniendo cada grupo 70 semillas del mismo peso. Pidió a siete amigos de diferentes partes del país que cultivaran un grupo de semillas y que le enviaran las semillas cosechadas. Sus conclusiones fueron las siguientes:

- El peso medio de las semillas filiales era función lineal del peso de las semillas padres con una pendiente menor que la unidad, es decir, el peso medio de la progenie se desvía menos de la población media que de los padres. Galton llamó a esta propiedad reversión en un discurso pronunciado en 1877 (después se llamará regresión). Por tanto, los padres de peso $M + x$ producen hijos adultos de peso medio $M + r \cdot x$, para $0 < r < 1$.
- Para cada grupo de semillas padres, el peso de las semillas filiales estaba normalmente distribuido. El peso de los hijos, debido a la variación aleatoria de entre hijos del mismo grupo de padres, llega a ser $M + r \cdot x + y$.

- La desviación probable del peso de las semillas filiales es la misma para todos los grupos y más pequeña que la desviación probable del peso de las semillas padres. Esta es la propiedad que hoy día se conoce como *homocedasticidad* u homogeneidad de las varianzas.

Para explicar estas propiedades se basa en gráficos y la identidad estadística de las dos generaciones le sirve para dar la relación entre la varianza condicional (variación entre grupos), la varianza marginal (la variación entre el total de la población) y el coeficiente de reversión. Numerosos estudios, como el citado llevan a Galton a la noción de correlación. En 1869 publica el libro "*Hereditary Genius*", donde estudia la influencia de las características de los padres y otros antepasados en las de los hijos. El método para expresar estas múltiples relaciones se le ocurre a Galton una mañana, repasando unas notas de su libreta mientras esperaba el tren. Lo cuenta del siguiente modo (citado por Newman, 1956, p. 239):

> "Parecía evidente por observación, y había sido completamente confirmado por esta teoría, que existía un "índice de correlación"; o sea, una fracción, que ahora llamamos simplemente *r* que relaciona con la mayor aproximación cada valor de desviación (de la media) por parte del sujeto con el promedio de todas las desviaciones asociadas, del pariente, tal como ha sido descrito. Por lo tanto, la aproximación de cualquier parentesco específico puede ser hallada y expresada con un término único. Si un individuo particular se desvía mucho, el promedio de las desviaciones de todos sus hermanos será una fracción definida de esa cantidad; del mismo modo que los hijos, los padres, primos hermanos, etc. Cuando no hay relación alguna, *r* se vuelve igual a 0; cuando es tan cercana que sujeto y pariente poseen idéntico valor, entonces $r = 1$. Por lo tanto, el valor de *r* reside siempre entre los límites extremos de 0 y 1. Mucho más podría añadirse, pero no sin usar lenguaje técnico, lo cual sería aquí inadecuado". (Newman, 1956, p. 239).

Galton no había considerado más que la distribución conjunta de una medida *x* tomada sobre el padre y la misma medida tomada sobre los hijos, pero con posterioridad se da cuenta de la posibilidad de estudiar variaciones conjuntas de medidas biológicas diferentes sobre los mismos individuos. En su obra, "*Natural Inheritance*", publicada en 1889, Galton propone un formidable programa de investigación biométrica: estudiar estadísticamente la variabilidad de las medidas físicas de distintas especies, a fin de confirmar matemáticamente el mecanismo de la evolución descrito por Darwin (Benzecri, 1982). En consecuencia, Galton fue consciente de que sus descubrimientos parecían dar lugar a un amplio campo de aplicación de problemas que caerían bajo las leyes de la correlación.

La sugerencia fue recogida por prestigiosos autores como Edgeworth, Pearson, Yule, Seppard (Hald, 1998) que en tiempos posteriores desarrollaron estas ideas en profundidad. Por ejemplo Weldon calculó empíricamente coeficientes de correlación entre varias medidas físicas de órganos de camarones, es decir correlaciones entre medidas del mismo sujeto. También contribuyó a crear la teoría sobre la distribución muestral del coeficiente de correlación.

A pesar del enorme ingenio y como señala Pearson: "*Galton no había aún alcanzado la idea de correlación negativa*" (Pearson, 1920, p. 199). Se basa en la definición que propone en el artículo "*Co-relación y su medida*", publicado en los Proceedings de la Royal Society, 45, pp. 135-145 en 1888:

> "Co-relación o correlación de estructura es una frase de gran uso en biología y no menos en la rama que estudia la herencia, y la idea es incluso más frecuente que la frase; pero no soy consciente de ningún intento anterior de definirla claramente, de trazar su forma de acción en detalle o mostrar cómo medirla. Dos órganos variables se dicen co-relacionados cuando la variación

> en uno se acompaña en promedio por más o menos variación del otro y en la misma dirección. Así, la longitud del brazo se dice correlacionada con la de la pierna porque una persona de brazo largo usualmente tiene pierna larga y al contrario. Si la correlación es cercana, entonces una persona de brazos muy largos tendrá piernas muy largas; si es moderadamente próxima, entonces la longitud de la pierna sólo será larga, no muy larga; y si no hubiese ninguna correlación, entonces la longitud de la pierna sería, en promedio, mediocre". (citado por Pearson, 1920, p.199).

Pearson continúa el trabajo de Galton, interesado en la Biometría debido a los trabajos de éste y de Weldon, quien en 1892 fue el primero que publicó un artículo en el que señala el significado de un coeficiente de correlación negativo, indicando que era posible determinar una razón (coeficiente), cuyo valor se convierte en $\pm$ 1 cuando un cambio en cualesquiera de los órganos implica un cambio igual en el otro, y 0 cuando los dos órganos son bastante independientes (Pearson, 1920).

Las ideas modernas sobre regresión se originan en los trabajo de Legendre y Gauss, sobre el método de mínimos cuadrados, para ajustar los datos sobre las órbitas de cuerpos celestes. El primer estudio documentado sobre el método de mínimos cuadrados, de donde deriva la idea de regresión, es debido a Legendre (1752-1833) en 1805. Esta técnica de optimización intenta encontrar la función (dentro de una familia) que mejor se ajusta a los datos bivariantes, de acuerdo con el criterio de mínimo error cuadrático, y siempre que los datos cumplan algunas condiciones (como independencia). Se conocían las ecuaciones funcionales de estas órbitas, pero los errores de medida hacía que los cálculos fuesen aproximados y se ajustaban ciertas familias de funciones, usando la teoría de errores (y la distribución normal para describirlos). En 1829 Gauss fue capaz de establecer la razón por la cual este procedimiento es muy adecuado desde el punto de vista estadístico, mediante lo que hoy se conoce como teorema de Gauss-Markov, que muestra que los estimadores obtenidos con este método son insesgados y no se requiere una distribución específica para los datos que se ajustan. El impulso posterior lo dan los trabajos citados de Galton y Pearson sobre la herencia. Yule la utiliza para el estudio de fenómenos sociales, como las causas de pobreza alrededor de 1899.

A lo largo del siglo XIX el trabajo de los estadísticos era mayormente descriptivo; la inferencia estadística se va a desarrollar como consecuencia de la creación de la Escuela Biométrica del University College de Londres bajo la dirección del matemático Karl Pearson (1857-1936), quien trata de aportar bases matemáticas a los descubrimientos de Galton (Botasso, 2009). Pearson defiende que el método científico es esencialmente estadístico, pues sus inferencias se basan en la asociación entre antecedentes y consecuentes. Alrededor de 1895, Pearson había resuelto las propiedades matemáticas del coeficiente de correlación y la regresión simple utilizada para la predicción lineal entre dos variables continuas. También generaliza la idea de regresión de Galton a varias dimensiones, comprobando que muchas variables biométricas siguen la distribución normal multivariante y por tanto se podría aplicar la regresión múltiple para relacionar varias de estas medidas. Yule introduce la notación e ideas para la correlación parcial de dos variables fijando una tercera en un artículo "*Sobre la teoría de la correlación para cualquier número de variables tratado mediante una nueva notación*" publicado también en los Proceedings de la Royal Society en 1907.

3. Uso de la Historia en la enseñanza de la matemática

Las breves notas históricas que acabamos de comentar podrían tener una utilidad didáctica, ya que en los currículos actuales se enfatiza el uso de las Matemáticas en situaciones reales con las cuales los estudiantes comenzarán la construcción de sus propios conocimientos matemáticos. Como se indica en el Real Decreto de enseñanzas mínimas del Bachillerato: "*Detrás de los contenidos que se estudian hay un largo camino conceptual, un constructo intelectual de enorme magnitud, que ha ido* *evolucionando a través de la historia hasta llegar a las formulaciones que ahora manejamos*." (MEC, 2007, p. 45449). Un recurso importante al respecto es el uso de la Historia de la Matemática en el diseño y construcción de secuencias de enseñanza-aprendizaje. Si observamos los sucesivos libros de texto en las últimas décadas, cada vez se incluyen más referencias históricas en el desarrollo de los temas. Creemos que el objetivo de los autores es situar el contenido de estudio y los problemas que dieron origen a los conceptos que se van a estudiar, a la vez que motivar a los estudiantes.

La Historia de la Matemática nos revela el origen de los problemas donde surgieron los objetos matemáticos, cómo se resolvieron dichos problemas (con las sucesivas aproximaciones a las soluciones, hasta su resolución actual), construyendo de esta forma el conocimiento matemático. Dichos problemas, la mayor parte de las veces, fueron extra-matemáticos y otras veces surgieron del propio saber matemático y sus relaciones.

Como apunta Brousseau (1986), generalmente los conocimientos matemáticos se presentan en los libros de una manera acabada, articulada, secuenciada y sin ningún indicio de su gestación y evolución, ocultando los problemas que dieron origen a su germinación y desarrollo y el proceso de elaboración que siguieron hasta constituirse en conocimientos socialmente aceptados y reconocidos como objetos del saber matemático. Cuando un matemático, después de su labor de investigación, encuentra un nuevo conocimiento matemático lo comunica de una manera "limpia" y, generalmente, ocultando los problemas y situaciones que dieron lugar a la investigación, las conjeturas iniciales, los intentos fallidos, las dificultades encontradas, las generalizaciones de interés; en pocas palabras, ocultando la historia de la formación de ese saber. Sin embargo, pocas veces el científico encuentra un nuevo saber directamente, sino que realiza conjeturas, inicia las primeras aproximaciones, muchas de ellas erráticas, da las primeras soluciones, muchas de ellas incompletas, otras triviales, estas primeras soluciones las va mejorando y delimitando hasta que encuentra el objeto matemático, que intenta generalizar al máximo nivel.

El proceso de generación de un nuevo saber matemático está personalizado por el investigador, contextualizado por las situaciones y problemas que se han estudiado. Sin embargo, el productor del saber, cuando lo comunica, lo despersonaliza y lo descontextualiza, quedando oculto todo el proceso de generación. Dicho proceso de generación puede tener un interés didáctico para que los que intentan aprender se convenzan de su validez, aunque no tengan que seguir el mismo camino de descubrimiento. Una vez constituido un saber, otras personas, lo transforman, reformulan, aplican, generalizan, según sus necesidades. De acuerdo a Brousseau, el trabajo del profesor es, en cierta medida, inverso al del productor del saber, debe producir una recontextualización y una repersonalización de los conocimientos. Estos conocimientos van a ser los conocimientos del alumno, es decir, una respuesta natural a las condiciones particulares y personales que son indispensables para que los nuevos conocimientos tengan sentido para el estudiante.

La tarea del profesor de recontextualización del conocimiento para ser comunicado a los estudiantes, puede quedar favorecida si conoce la génesis y evolución de los conocimientos matemáticos, es decir, la historia de su formación y evolución. En el proceso de adquisición, el estudiante lo personaliza contextualizado en las situaciones que ha trabajado y que han dado sentido al objeto matemático en cuestión. Cuando el estudiante se ha apropiado de los nuevos conocimientos con sentido y ha adquirido un cierto dominio de ellos, debe redescontextualizar y redespersonalizar sus nuevos conocimientos e identificarlos con los existentes en la comunidad matemática científica. Por ello, el conocimiento de la historia de la formación de los objetos matemáticos ayuda al estudiante a comprenderlos mejor.

En general hay dos estrategias básicas para introducir la Historia de las Matemáticas en la enseñanza (Radford, 1997; Fried, 2001): la *estrategia de la adición* y la *de acomodación*. La estrategia de la adición consiste en introducir en los programas anécdotas, extractos de biografías, problemas, métodos de resolución, etc. Esta estrategia no altera el currículo, pero lo extiende; esto es un problema, pues hay poco tiempo para desarrollar todos los contenidos. La estrategia de acomodación consiste en utilizar un desarrollo histórico, la explicación de una técnica o una idea, para organizar una lección de acuerdo con dicho tema histórico. Esta organización de los temas siguiendo su desarrollo histórico nos proporcionaría una oportunidad para discutir la motivación que llevó al estudio del tema tanto dentro de las Matemáticas como en otros campos científicos, aunque el profesor debe filtrar de la Historia lo que es relevante de lo que no lo es (Fried, 2008).

El conocimiento de la Historia puede servir también para estudiar la epistemología de los objetos matemáticos. Desde este punto de vista, la Historia de la Matemática se puede ver como un laboratorio de epistemología en el que se estudia y analiza el desarrollo del conocimiento matemático (Radford, 1997). Un ejemplo de este enfoque es el estudio de los obstáculos epistemológicos en la formación y evolución de los objetos matemáticos. Esta noción, debida a Bachelard (1938), fue introducida en la Didáctica de la Matemática por Guy Brousseau, para explicar errores frecuentes y persistentes que cometen los estudiantes en su proceso de estudio de las nociones matemáticas y que tradicionalmente se han interpretado como falta de conocimiento o de dominio de los conceptos matemáticos. Brousseau cambia esta concepción del error:

> “el error y el fracaso no tienen el papel simplificado que a veces se les atribuye. El error no es solamente el efecto de la ignorancia, la incertidumbre o el azar, sino de un conocimiento anterior, que tenía su interés, su éxito, pero que ahora se revela falso o simplemente inadaptado. Los errores de este tipo no son erráticos e imprevisibles, se constituyen en obstáculos” (Brousseau, 1983, p. 173).

El autor distingue tres tipos de obstáculos: ontogenéticos, didácticos y epistemológicos. Estos últimos se pueden encontrar en la evolución histórica de los propios conceptos y en los modelos espontáneos de los alumnos. Como ejemplo, para el caso de la correlación, podemos citar los encontrados en diferentes investigaciones (Estepa y Batanero, 1996; Truran 1995, 1997; Morris 1997, 1998; Batanero, Estepa y Godino, 1997; Batanero y Godino, 1998; Batanero, Godino y Estepa, 1998) orientadas al estudio de las concepciones y comprensión de la correlación y regresión por estudiantes universitarios. Se ha observado que los alumnos no aprecian, a veces, la correlación inversa, que tienen un sentido determinista o local de la correlación y que identifican, con frecuencia, la correlación con la causalidad. Hemos mostrado cómo Galton, en sus primeros trabajos, no considera la correlación inversa, igual que algunos de nuestros estudiantes cuando comienzan el estudio de estos temas, en consecuencia lo podemos considerar como un obstáculo epistemológico en el sentido descrito de Brousseau, que se llega a superar con el estudio del tema. Por otra parte, como es conocido, asociación y causalidad no siempre son coincidentes, a pesar de que uno de los pasos en la búsqueda de relaciones causales es estudiar la covariación de las variables (Pozo, 1987).

El desarrollo en los estudiantes del pensamiento estadístico es objetivo prioritario en la Educación Estadística porque ayudará a los estudiantes a superar las dificultades y obstáculos que encuentren en su aprendizaje de los temas estadísticos. El uso de la Historia puede ayudar a desarrollar el pensamiento estadístico, ya que, como afirman Pfannkuch & Wild (2004), el desarrollo del pensamiento estadístico en los estudiantes presenta cuatro retos en su enseñanza y, en consecuencia a los profesores. El primer reto para el educador es tomar conciencia de las características del pensamiento estadístico, el segundo reto es el reconocimiento del pensamiento estadístico en una variedad de contextos y situaciones y justificarlo en ellos, el tercer reto es desarrollar estrategias que promuevan y faciliten la construcción del pensamiento estadístico en los estudiantes, el cuarto y último reto es planificar e implementar una enseñanza que evalúe las estrategias que centran el desarrollo del pensamiento estadístico en los estudiantes, para ello los estudiantes deben comprender cómo los estadísticos razonan y actúan dentro de la ciencia estadística, la historia del desarrollo de los conceptos estadísticos ayudan en estos retos.

4. Conclusión

La mayoría de nuestros estudiantes de Educación Secundaria creen que los estudios sobre correlación y regresión se deben a Karl Pearson, seguramente por el nombre que se le da al coeficiente de correlación lineal (coeficiente de Pearson) (Stanton, 2001). Sin embargo, aunque Pearson realizó un riguroso desarrollo matemático del coeficiente de correlación, fue el ingenio y la imaginación de Sir Francis Galton lo que le permitió llegar a los conceptos de correlación y regresión tal y como hoy los entendemos. Ello nos da una importantísima lección didáctica: personas con formación inicial diferente a la matemática, como Galton, partiendo de problemas reales y con una fuerte motivación, llegan a descubrir conceptos matemáticos, que han llegado a tener una importancia transcendental. Algo parecido a lo que hoy nos proponen los nuevos currículos, donde se sugiere que partiendo de problemas reales se debe conseguir que nuestros estudiantes descubran conceptos matemáticos importantes.

En el apartado anterior hemos expuesto algunas ideas sobre la conveniencia del uso de la Historia en la enseñanza. Otras razones señaladas por Fried (2001) son: (a) La Historia de las Matemáticas humaniza las Matemáticas, ya que los desarrollos históricos se han llevado a cabo en culturas diferentes, proporciona a los estudiantes el papel de los modelos utilizados, además de conectar el estudio de las Matemáticas con las motivaciones humana; (b) La Historia hace las Matemáticas más interesantes, más comprensibles y más accesibles porque da variedad a la enseñanza, disminuye el miedo de los estudiantes a las Matemáticas y las sitúa en la sociedad; y (c) La Historia da perspicacia y agudeza (insight) al estudiante respecto a los conceptos, problemas y resolución de problemas, ya que proporciona ideas y contextos para los problemas bajo estudio, sugiere diferentes enfoques para su resolución y muestra las relaciones que existen entre las ideas, definiciones y aplicaciones.

Este uso no implica un incremento de los contenidos, sino tener en cuenta las situaciones que dieron lugar a los conceptos matemáticos que, a nuestro juicio, motiva y enriquece el conocimiento de nuestros estudiantes. Además, se puede aprovechar el ejemplo del nacimiento de la noción de correlación en los trabajos de Galton a partir del estudio de la herencia biológica para seguir las recomendaciones de desarrollo de una idea matemática a través de casos reales.

En nuestra descripción hemos visto ejemplos del paralelismo de algunos hechos del proceso de descubrimiento y evolución de los conceptos matemáticos con el proceso de apropiación de las nociones matemáticas por algunos de nuestros estudiantes. Por ejemplo, se constata que nuestros estudiantes, al comienzo del estudio de la correlación, no admiten la correlación negativa, al igual que ocurrió con Galton, en sus primeros estudios, aunque después se supere en estudios posteriores. Es por ello que el conocimiento de la Historia puede también ayudar a los profesores para prevenir dificultades en sus estudiantes.

Agradecimientos: Proyecto EDU2010-14947 (MCIN), beca BES-2011-044684 (MCIN), beca FPU-AP2009-2807 (MCIN) y grupos FQM126 y HUM793 (Junta de Andalucía).

Bibliografía

Alloy, L.G. y Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*. 91, 112-149.

Bachelard, G. (1938). *La formation de l'esprit scientifique*. Paris: Libraire Utin.

Batanero, C. y Godino, J.D. (1998). Understanding graphical and numerical representations of statistical association in a computer environment. En L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee y W. Wong, (eds). *Proceedings of the Fifth Conference on Teaching Statistics*. 1017-1024. Voorburg: International Statistical Institute.

Batanero, C., Godino, J.D., y Estepa, A. (1998). Building the meaning of statistical association through data analysis activities (Research Forum). En A. Olivier y K. Newstead, (eds.), *Proceedings of the 22nd Conference of the Internacional Group for the Psychology of Mathematics Education*. Vol 1, 221-236. Stellembosch (South Africa): Universidad de Stellenbosh.

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*. 191-205. Voorburg: Internacional Statistical Institute.

Benzecri, J.P. (1982). *Histoire et prehistoire de l'analyse des données*. París: Bordás.

Brousseau, G. (1983). Les obstacles épistémologiques et les problems en Mathématiques. *Recherches en Didactique des Mathématiques*. Vol. 4(2), 165-198.

Brousseau, G. (1986). *Théorisation des phénomènes d'enseignement des Mathématiques*. Thèse d'État. Université de Bordeaux I. Francia

Botasso, O. (2009). El coeficiente de correlación, una historia de debates movilizantes. *Revista Médica de Rosario*. 75, 80-82.

Castro, A.E. Vanhoof, S. Noortgate, W.V. D. & Onghena, P. (2009). The transitivity misconceptions of Pearson's correlation coefficient. *Statistics Education Research Journal*. 8(2), 33-54. Recuperado el 21 de marzo de 2012, de http://www.stat.auckland.ac.nz/~iase/serj/SERJ8(2)_Sotos.pdf

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, 271-280.

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

Estepa, A. y Batanero, M.C., (1996). Judgments of correlation in scatter plots: An empirical study of students' intuitive strategies and preconceptions. *Hiroshima Journal of Mathematics Education*. 4, 25-41.

Fried, M.N. (2001). Can mathematics education and history of mathematics coexist. *Science & Education*. 10, 391-408.

Fried, M.N. (2008). History of mathematics in mathematics education: a Saussurean perspective. T*he Montana Mathematics Enthusiast*. 5 (2 y 3), 185-197.

Hald, A. (1998). *A history of mathematical statistics. From 1750 to 1930*. New York: John Wiley.

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.

Pfannkuch, M. & Wild, C. (2004). Understanding of Statistical Thinking. En D. Ben-Zvi y J. Garfield (eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*, pp. (17-46). Dordrecht (The Nederlands): Kluwer Academic Publishers. ISBN 1-4020-2277-8 (HB).

Morris, E.J. (1997). *An investigation of students' conceptions and procedural stills in the statistical topic correlation*. Report nº 230. Centre for information Technology in Education: The Open University.

Morris, E.J. (1998). Link: The principled design of a computer assisted learning program for correlation. En L. Pereira-Mendoza, L. Seu Kea, T Wee Kee y W. Wong (eds.), *Proceedings of the*

Fifth Conference on Teaching Statistics (vol. 2, 1033-1040). Voorburg: International Statistical Institute.

Newman, J.R. (1956). Sigma. El mundo de las matemáticas. Barcelona: Grijalbo.

Pearson, K. (1920). Notes on the history of correlation. *Biometrika*. 13, 25-45.

Pearson, K. (1965). Somes incidents in the early history of biometry and statistics 1890-1894, *Biometrika*. 52, 3-18.

Pozo, J.I. (1987). *Aprendizaje de la ciencia y pensamiento causal*. Madrid: Visor.

Radford, L. (1997). On psychology, historical epistemology, and the teaching of mathematics: towards a socio-cultural history of mathematics. *For the learning of Mathematics*. 17 (1), 26-33

Seal, H.L. (1967). The historical development of the Gauss linear model. *Biometrika*. 54, 1-24.

Stanton, J.M. (2001). Galton, Pearson, and the peas: a brief history of linear regression for statistics instructors. Journal of Statistics Education. 9(3). Recuperado el 21 de marzo de 2012, de http://www.amstat.org/publications/jse/v9n3/stanton.html

Truran, J.M. (1995) Some undergraduates' understanding of the meaning of a correlation coefficient. En B. Atweh y S. Flavel, S. (eds.). *Proceedings of the Eigteenth Annual Conference of the Mathematics Education Research Group of Australasia (MERGA)*. 524-529. Northern Territory University, Darwin, Australia.

Truran, J.M. (1997). Understanding of association and regression by first year economics students from two different countries as revealed in responses to the same examination questions. En J. B. Garfield y J. M. Truran (eds.), *Research Papers on Stochastics Educations from 1.997*. 205-212. Department Educational Psychology University of Minnesota.

Antonio Estepa Castro, nacido (13/05/1952) en Valdepeñas de Jaén (Jaén). Doctor en Matemáticas. Catedrático de Didáctica de la Matemática de la Facultad Humanidades y Ciencias de la Educación de la Universidad de Jaén. Su campo de investigación es la Didáctica de la Matemática, especialidad de Didáctica de la Estadística y la Probabilidad. Ha publicado diversos trabajos de investigación en este campo en revistas, libros y congresos. También ha realizado bastantes revisiones de trabajos de investigación para revistas y congresos importantes en dicho campo de investigación.
Universidad de Jaén. Email: aestepa@ujaen.es

María Magdalena Gea Serrano, nacida (18/02/1980) en Cantoria (Almería). Licenciada en Matemáticas (Universidad de Murcia), licenciada en CC. y TT. Estadísticas (Universidad de Granada), posee el Máster en Estadística Aplicada (Universidad de Granada) y el Diploma de estudios avanzados (Universidad de Jaén) bajo la tutela de D. Antonio Estepa Castro. Su investigación se proyecta en torno a la enseñanza y aprendizaje de la asociación estadística. En la actualidad es becada bajo la tutela de Dra. Carmen Batanero Bernabeu por el Plan de Formación de Personal Investigador (2011).
Universidad de Granada. Email: mmgea@ugr.es

Gustavo R. Cañadas de la Fuente, nacido el 2/2/1983 en Linares (Jaén) es licenciado en CC. y TT. Estadísticas (Universidad de Granada), posee el Máster en Metodología de las Ciencias del Comportamiento por la UNED y el Máster en Didáctica de las Matemáticas por la Universidad de Granada. Fue becado en el Plan de Formación del Profesorado Universitario para trabajar en la Universidad de Granada en la convocatoria del 2009 bajo la tutela de Dra. Carmen Batanero. Ha publicado trabajos relacionados con las tablas de contingencia.
Universidad de Granada. Email: grcanadas@ugr.es

José M. Contreras García, nacido en Granada; profesor ayudante doctor de la Universidad de Granada. Licenciado en Ciencias Matemáticas (esp. Estadística e I.O.), licenciado en C.C. y T.T. Estadísticas, Diploma de estudios avanzados en Estadística e I.O., Máster en didáctica de la matemática, Máster en Estadística Aplicada y doctor en Didáctica de la Matemática. Publicaciones en didáctica de la probabilidad.
Universidad de Granada. Email: jmcontreras@ugr.es