

Recursos virtuales para el estudio de la correlación y regresión

M. Magdalena Gea; Danilo Díaz-Levicoy; M. del Mar López-Martín y Gustavo
R. Cañadas

email: mmgea@ugr.es; dddiaz01@hotmail.com; mariadelmarlopez@ugr.es;
grcanadas@ugr.es

Universidad de Granada

RESUMEN

A pesar de la importancia de la correlación y regresión, la investigación sobre el tema sugiere pobres capacidades de los adultos en la estimación de la correlación, así como existencia de sesgos asociados. Son también numerosas las dificultades que permanecen después de la enseñanza del tema. En este trabajo analizamos los componentes del sentido de la correlación y regresión, y algunos ejemplos de recursos interactivos en Internet que podrían utilizarse para desarrollar este sentido en los estudiantes.

Palabras clave: sentido estadístico, bachillerato, correlación, regresión, recursos virtuales.

1. Introducción

La correlación y regresión poseen gran relevancia en el dominio de la Estadística. En un primer nivel, constituye una herramienta muy útil en el proceso de investigación estadístico, donde el conocimiento sobre la intensidad de la relación (análisis de correlación) que pueda existir entre las variables implicadas ofrece información suficiente como para matematizar tal relación (análisis de regresión) en el ajuste de un modelo con fines predictivos. En un segundo nivel, sirven de base para entender otros muchos conceptos y procedimientos estadísticos más avanzados.

A pesar de la relevancia que el razonamiento covariacional posee dentro del desempeño del proceso estadístico [1], la correlación y regresión son tratadas únicamente, en el sistema educativo español, en la enseñanza postobligatoria de Bachillerato, pues en la Educación Secundaria Obligatoria (ESO), tan sólo se introduce al alumnado el uso de tablas de contingencia en cuarto curso en los bloques de contenido de Estadística y Probabilidad (en sus dos opciones A y B) ([2], p. 758 y p. 759).

La correlación y regresión se fijan como contenidos en las materias de Matemáticas I (“Distribuciones bidimensionales. Relaciones entre dos variables estadísticas. Regresión lineal”, perteneciente al bloque 4 “Estadística y Probabilidad”) ([3], p.45449) y Matemáticas aplicadas a las Ciencias Sociales I (“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”, perteneciente al bloque 3 “Probabilidad y estadística”) ([3], p.45475). Estas materias pertenecen a las modalidades de Ciencias y Tecnología y Humanidades y Ciencias Sociales, respectivamente [3].

La importancia del razonamiento estadístico ha sido resaltada por numerosos investigadores ([4], [5], [6]). Será entonces responsabilidad de los profesores desarrollar en sus estudiantes este razonamiento, que, como indican los autores citados, complementa al razonamiento matemático, al permitir modelizar situaciones inciertas. Pero, a pesar de que la estadística actualmente ha tomado presencia en todos los niveles educativos: desde la educación primaria al Bachillerato, en formación profesional y en la mayoría de carreras universitarias de grado y postgrado, son muchas las investigaciones que nos alertan de la variedad y amplitud de extensión de errores en la interpretación y aplicación de conceptos estadísticos elementales.

En [7] se explica esta situación contradictoria por la falta de sentido estadístico en los estudiantes. Los autores indican que con frecuencia, la enseñanza de la estadística se realiza en forma rutinaria, no se presta atención a la interpretación de los conceptos por medio de ejemplos tomados de la vida cotidiana que los estudiantes puedan comprender. Tampoco se realizan investigaciones completas en que el estudiante recorra todos los pasos del método estadístico. Por ello, es habitual encontramos estudiantes que han finalizado sus estudios sin tener una comprensión profunda de los conceptos estadísticos fundamentales, y que además no son capaces de seleccionar el procedimiento más adecuado que deben aplicar cuando se enfrentan a un problema de estadística en su vida futura..

Los autores proponen un modelo de sentido estadístico formado por la unión de la cultura y razonamiento estadístico. En la primera incluyen la comprensión de las ideas estadísticas fundamentales [8] y unas actitudes y creencias adecuadas. Respecto al modelo de razonamiento estadístico, asumen el propuesto en [9]. En un trabajo propio [10] se han desarrollado estas ideas para el caso específico de la correlación y regresión.

La enseñanza de este tema no es simple. La investigación ha descrito sesgos de razonamiento y dificultades de comprensión, como no apreciar la correlación inversa, tener un sentido determinista o local de la correlación o identificar correlación con causalidad ([11]; [12]; [13]). Dichas creencias, en algunos casos, resisten al cambio incluso después de la enseñanza [14]. También se han observado errores al interpretar los coeficientes de correlación y regresión ([15]; [16]; [17]).

La finalidad de este trabajo es contribuir a la mejora de la enseñanza, analizando algunos recursos virtuales que pueden ayudar al desarrollo del sentido de la correlación y regresión en los estudiantes. A continuación analizamos, en primer lugar, los componentes del sentido de la correlación y regresión, para a continuación llevar a cabo el análisis de algunos recursos de interés relacionados con el tema en Internet.

2. Componentes del sentido de la correlación y regresión

Como se ha indicado, en nuestro modelo de sentido estadístico incluimos, tanto la comprensión de las ideas estadísticas fundamentales, como el desarrollo de un razonamiento adecuado y unas actitudes favorables. A continuación analizamos cada uno de estos puntos.

2.1. Comprensión de las ideas fundamentales

En primer lugar es necesario que los estudiantes adquieran las ideas fundamentales sobre correlación y regresión. Las ideas fundamentales son aquellas que se pueden enseñar a diversos niveles de complejidad, siguiendo un currículo en espiral y que son esenciales dentro de la estadística. Para el caso de la estadística fueron definidas en [8]. Las correspondientes a este tema las describimos en [10] del modo siguiente:

Datos y Distribución bivariada.

La importancia de los datos en estadística se debe a que esta ha sido definida por Moore [18] como la ciencia de los datos. El contexto de los datos es tan importante como las operaciones que con ellos se realizan en estadística, mientras que en otras ramas de matemática el contexto no tiene un papel tan esencial [7]. Por ejemplo, una correlación de 0,7 puede ser pequeña cuando se analizan datos de un fenómeno físico, como la correlación entre la velocidad de un móvil y el espacio recorrido, e indicará un gran error de medida. Pero el mismo valor de correlación es grande si se refiere a medidas biológicas, como la altura de padres e hijos, o más si se refiere a datos sociales.

En el estudio de la correlación y regresión se trabaja con datos bivariados. Es decir, para cada individuo de una muestra analizamos a la vez los valores en dicho individuo de dos variables estadísticas que se pretende relacionar. Ello nos lleva a considerar una variable estadística bidimensional formada por un conjunto de pares de valores posibles, que con las frecuencias conjuntas correspondientes a cada par constituye una distribución bidimensional. Asociada a la misma, será necesario que los estudiantes diferencien las frecuencias absolutas y relativas conjuntas, marginales y condicionales. Será entonces importante presentar al estudiante actividades variadas con datos provenientes de diversas fuentes como encuestas realizadas en clase, prensa o Internet donde puedan diferenciar estos tipos de frecuencia.

Representación tabular y gráfica.

Las representaciones juegan un papel esencial en toda la matemática pues permite concretizar los objetos matemáticos abstractos, pero en estadística tienen un valor añadido. Su importancia se debe a que constituyen un instrumento de análisis en sí mismos pues al pasar los datos desorganizados a un gráfico encontramos nueva información.

En el estudio de la correlación y regresión los datos se organizan en una tabla de doble entrada, cuyas celdas representan la frecuencia conjunta de los valores de la variable, que se fijan en sus correspondientes filas y columnas. Pero estas tablas no son sencillas de interpretar, pues además de las celdas (frecuencias absolutas) hay que tener en cuenta los diversos tipos de frecuencia relativa que se deducen de ellas. Además, se ha mostrado la dificultad de interpretación de estas frecuencias por parte de los estudiantes [19]. Así mismo, la información se representa gráficamente, siendo la representación más utilizada para representar la correlación en los textos de Bachillerato el diagrama de dispersión o nube de puntos, aunque también encontramos histogramas, pictogramas y gráficos de burbuja (Ver Figura 1). Hacemos notar que en los datos bivariados, al pasar de un listado al diagrama de dispersión podemos observar la intensidad de la relación (a través de la mayor o menor dispersión de la nube de puntos), su sentido (si la relación es directa o inversa) y el tipo (lineal o no), observando su tendencia [17]. El diagrama de burbujas permitiría también visualizar, simultáneamente, hasta tres variables (representando la tercera mediante el diámetro) o incluso cuatro, si mediante el color pudiera representarse una cuarta variable. Respecto a estas representaciones conviene plantear actividades de construcción y de lectura, así, como de cambio entre representaciones.

Construye la tabla de doble entrada correspondiente, a partir del diagrama de dispersión, teniendo en cuenta la frecuencia de los datos que figura entre paréntesis.

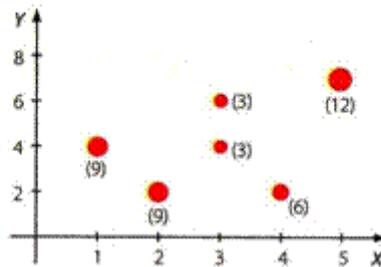


Figura 1. Diagrama de burbujas ([20], p.259)

Dependencia funcional, aleatoria e independencia.

Estos conceptos son los más importantes en el estudio de la correlación y regresión, y es fundamental que el estudiante llegue a diferenciarlos, para lo cual es recomendable que el profesor presente actividades variadas con distintos grados de dependencia entre las variables. Debe hacer observar al estudiante que en una dependencia funcional a cada valor de una variable X (independiente) corresponde un solo valor de otra variable Y (dependiente), pero en la dependencia aleatoria a cada valor de X corresponde una distribución de valores de Y , por lo que este concepto amplía el de dependencia funcional. Además, es posible medir la intensidad y dirección de la relación lineal mediante el coeficiente de correlación de Pearson y la dependencia funcional puede considerarse un caso extremo de la aleatoria.

La definición de independencia es importante, pues la comprensión de esta idea es base de muchos temas estadísticos posteriores; por ejemplo, en inferencia, la mayor parte de contrastes estadísticos e intervalos de confianza están definidos bajo el supuesto de independencia estadística de los datos muestrales.

Covarianza y correlación.

Como se ha indicado, la covarianza y la correlación permiten medir la intensidad de la relación entre las variables y su signo, si la dependencia es lineal. La covarianza es un primer coeficiente utilizado para analizar el signo de la relación, pero su valor es engañoso al no ser una medida acotada. Una estrategia de introducción intuitiva para la explicación de su significado es la división en cuatro cuadrantes de la nube de puntos por las rectas correspondientes a las medias de cada variable, propuesta en [21], y como se muestra en la Figura 2. Así mismo, este análisis propicia que el estudiante comprenda más significativamente el cálculo del coeficiente de correlación lineal de Pearson en base al anterior.

Interpretación de la covarianza

Según sea el signo de la covarianza, se interpreta:

- Covarianza positiva:** al aumentar los valores de la variable X , aumentan los valores de la variable Y . La nube de puntos se orienta a la derecha y hacia arriba.
- Covarianza negativa:** al aumentar los valores de la variable X , disminuyen los valores de la variable Y . La nube de puntos se orienta a la derecha y hacia abajo.



Si se calcula el centro de gravedad $G(\bar{x}, \bar{y})$ y se toman unos ejes con el origen en este centro, se observa:

- Si los puntos están en el 1^{er} y 3^{er} cuadrantes, mayoritariamente los productos $(x_i - \bar{x})(y_i - \bar{y})$ son positivos.
- Si los puntos están en el 2^o y 4^o cuadrantes, mayoritariamente los productos $(x_i - \bar{x})(y_i - \bar{y})$ son negativos.

Figura 2. Interpretación de la covarianza ([22], p. 246)

Es importante que los estudiantes comprendan que la correlación mide la relación lineal y que puede haber una relación fuerte entre las variables (por ejemplo, parabólica) con un bajo valor del coeficiente de correlación de Pearson. Para ello, el profesor debe mostrar ejemplos de relaciones lineales y no lineales con diversos grados de intensidad.

Regresión.

Encontrada una relación moderada o fuerte entre las variables, un siguiente paso es tratar de deducir un modelo matemático que permita predecir una variable dependiente en función de una variable independiente. Para ello ajustamos una función a los datos, generalmente siguiendo el criterio de mínimos cuadrados, aunque también es posible seguir otros criterios.

La diferencia entre variable dependiente e independiente ocupa un lugar central en el análisis de la regresión, ya que una vez aceptada la dependencia entre las variables del estudio, y con objeto de expresar en forma de ecuación o modelo una variable en función de otra, se necesita seleccionar qué variable servirá como dependiente o predictora. Es importante aclarar esta diferencia a los estudiantes para que no deduzcan la existencia de una única recta de regresión.

Modelos de regresión y sus parámetros.

Al ajustar una función a los datos estamos realizando una modelización. Como se señala en [23], la regresión es un modelo general para comprender las relaciones entre variables. El método de mínimos cuadrados permite obtener aquella recta que minimiza la suma de los cuadrados de la diferencia existente entre los datos teóricos (puntos sobre la línea de regresión) y los reales (datos). Será importante diferenciar los parámetros que definen la recta, centrando el interés en la pendiente de la recta, y diferenciando los dos coeficientes de regresión, dependiendo de qué variable se considere como dependiente o independiente. En Bachillerato raramente se alude al tratamiento de datos atípicos, aunque encontramos un texto en la bibliografía analizada por los autores en investigaciones previas (Ver, por ejemplo, [24]), en el que se presenta el procedimiento de cálculo de la recta de Tukey para obtener la recta de regresión respecto a la mediana.

Estimación y bondad de ajuste.

Finalmente, es importante resaltar la utilidad de la recta de regresión para realizar estimaciones del valor de la variable dependiente en función de la independiente, e introducir el coeficiente de determinación como medida de la bondad del ajuste. Una vez calculada la función de ajuste a la distribución bidimensional, es habitual encontrar en los libros de textos su utilidad planteando ejercicios de estimación de un valor de la variable dependiente (Y) conocido un valor de la variable independiente (X). Pero además de esta gran utilidad, conviene que los estudiantes aprendan a usar el coeficiente de determinación r^2 como medida de la bondad del ajuste. Además, es necesario que entiendan que la bondad de la estimación también depende de la distancia del punto utilizado para la estimación al centro de gravedad de la distribución.

2.2. Razonamiento correlacional

Además de la comprensión de las ideas que se han descrito en el apartado anterior, será también necesario desarrollar en los estudiantes el razonamiento sobre la correlación y regresión. El razonamiento correlacional sería un componente del razonamiento estadístico y en [18] se señala los siguientes elementos fundamentales:

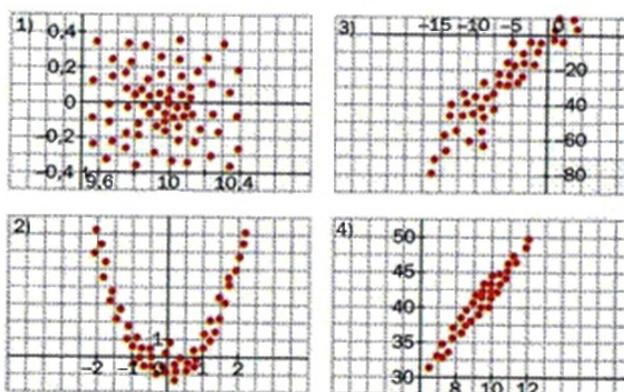
- Aceptar y reconocer la omnipresencia de la variabilidad aleatoria, en contraposición a la visión determinista de la realidad; en el caso de datos bivariados implica aceptar que el conjunto de datos tiene cierta dispersión respecto al modelo de regresión.
- Comprender la necesidad de los datos en los procesos de estudio estadístico, donde la prioridad es buscar en los datos que se necesiten. Por tanto, el estudio de la correlación y regresión debe basarse en una muestra de datos de tamaño suficiente.
- El diseño adecuado de la recogida de datos, teniendo presente la variabilidad que se puede presentar en los datos, la muestra o el método de recogida. Lo ideal es considerar una muestra aleatoria; si ello no es posible, se deben controlar las variables extrañas que puedan afectar a la correlación.
- La cuantificación de la variación. La variación aleatoria se describe, matemáticamente, por

la probabilidad y la estadística mediante la dispersión. En el caso que nos ocupa, la correlación y la covarianza permiten cuantificar la variación.

- Explicación de la variación. El análisis estadístico busca efectos sistemáticos detrás de la variación aleatoria. En nuestro caso el coeficiente de determinación permite explicar la parte de la variación debida al modelo de regresión y la parte aleatoria.

Para mejorar el razonamiento sobre correlación, sería necesario mejorar la competencia para la estimación de la correlación a partir de diversas representaciones de datos, pues la investigación previa muestra sesgos en dicha estimación. Para ello, conviene presentar a los estudiantes ejemplos variados de diagramas de dispersión (Ver Figura 3), teniendo en cuenta las variables que influyen en la dificultad de la estimación de la correlación [17]: signo de la relación (positiva o negativa); intensidad de la correlación (alta, media, baja o nula); número de datos en la distribución; y si la correlación en los datos coincide o no con la esperada por el estudiante (consistencia con las teorías previas del estudiante).

Considera las nubes de puntos de la figura.



a) Indica si hay relación de dependencia entre la variable X y la variable Y. En caso de haberla, ¿puede considerarse esta relación lineal?

b) Asigna a cada gráfico, una de las siguientes rectas:

$$y = x \quad y = 1 - 0,2x \quad y = 2 + 4x$$

Figura 3. Interpretación de la relación de dependencia en el diagrama de dispersión ([25], p.260)

Es también importante desarrollar la comprensión de la diferencia entre correlación y causalidad mediante la discusión con los estudiantes sobre la explicación de la correlación. Además de una relación causa-efecto, otras posibles explicaciones de la correlación son [26]:

- Interdependencia (cada variable afecta a la otra), como en el caso de la longitud de piernas y la altura de una persona.
- Dependencia indirecta, donde una tercera variable es la que determina la correlación; por ejemplo, la relación entre la tasa de fertilidad y el producto interior bruto están inversamente relacionados; pero en realidad las dos variables dependen de la proporción de mujeres que trabaja.
- Concordancia o coincidencia en preferencia u ordenación de una misma serie de datos, como por ejemplo, si dos miembros de un jurado, de modo independiente, califican a una misma persona, ya que las calificaciones están correlacionadas pero la de uno no influye en la del otro.
- Covariación casual o espúrea, cuando aparece una correlación entre dos variables, pero es efecto del azar.

2.3. Actitudes y creencias

Finalmente, el sentido de la correlación requiere unas buenas actitudes hacia la estadística y

en concreto, hacia el estudio de la correlación y regresión. El profesor debe lograr que los estudiantes se interesen por el tema y su aprendizaje, y valoren su importancia para la modelización matemática en diversas áreas de conocimiento. Ello se consigue mediante una serie de ejemplos adecuados de aplicación y el uso del trabajo con datos reales, que son hoy abundantes en Internet. Se debe fomentar también una actitud crítica hacia el uso inadecuado de la estadística y el interés por detectar sesgos e informaciones erróneas sobre estadística que con frecuencia aparecen en los medios de comunicación.

Asimismo han de superar sus creencias erróneas sobre la correlación, entre las que encontramos las siguientes [27]:

- Concepción causal: Son muchos los estudiantes que confunden correlación y causalidad, pero mientras una relación causal lleva siempre a una correlación, el caso contrario no siempre es cierto.
- Concepción unidireccional: no considerando la correlación negativa o estimando una intensidad menor de la misma; en los casos en que existe este sesgo las correlaciones negativas se estiman a la baja.
- Concepción local: basar la estimación de la correlación en el análisis de sólo una parte de los datos, sobre todo fijándose en casos sobresalientes o atípicos.
- Concepción determinista: considerar sólo la dependencia funcional y no la dependencia aleatoria.

3. Recursos y actividades para desarrollar el sentido de la correlación y regresión

Los componentes descritos del sentido de la correlación y regresión deberían ser desarrollados en los estudiantes a partir del trabajo con múltiples problemas y ejemplos, así como llevando a cabo una metodología de enseñanza basada en proyectos como el descrito en [28]. El trabajo con proyectos hace que los estudiantes recorran todos los pasos del método estadístico, desde el planteamiento de un problema, la decisión sobre qué datos recoger, su recogida, análisis e interpretación y comunicación de las conclusiones.

Otro modo de ayudar a desarrollar este sentido en los estudiantes es el trabajo con recursos interactivos. El auge actual de la tecnología y medios de comunicación personal o redes sociales aumenta la responsabilidad de las personas en su propia formación. Son muchos los recursos disponibles en Internet, como por ejemplo, conjuntos de datos que pueden utilizarse en la clase de estadística. Para aprovechar su potencial, se deberían aprovechar las posibilidades que brindan las nuevas tecnologías, de manera que se innovase en la enseñanza mediante estos recursos que pudiesen ayudar a interpretar y razonar críticamente datos y conceptos estadísticos [29].

Como se expone en [30], estos recursos permiten crear micromundos virtuales en el que los conceptos y objetos matemáticos abstractos pueden concretizarse. Igualmente, proporcionan a los estudiantes una experiencia estocástica que es más difícil de conseguir que con otros medios. Están hoy día al alcance de estudiantes y profesores mediante el uso de ordenadores y dispositivos móviles, y en la mayoría de los casos no tienen coste. Proponemos, en lo que sigue, algunos ejemplos de estos recursos así como la explicación de cómo podrían utilizarse en el aula de matemáticas. Desarrollaremos el análisis completo de un ejemplo, siguiendo la metodología desarrollada en [31] y describimos otros en forma más breve.

3.1. Análisis detallado de un recurso

En la clase de matemáticas se suele exponer a los estudiantes que la recta de regresión hace mínima la suma de los cuadrados de las desviaciones de los puntos en el diagrama de dispersión al modelo utilizado (recta de regresión de mínimos cuadrados).

Una demostración matemática (incluso elemental) de esta propiedad de la recta de regresión requiere el uso de derivadas parciales; por tanto, está fuera de la comprensión de los estudiantes de Secundaria o Bachillerato. Una alternativa de “justificación informal”, que permite dotar al estudiante del sentido del método de mínimos cuadrados, lo proporciona el

applet disponible en el servidor de Educación de la Comunidad de Navarra (Ver Figura 4), disponible en: [//docentes.educacion.navarra.es/msadaall/geogebra/figuras/e3regresion.htm](http://docentes.educacion.navarra.es/msadaall/geogebra/figuras/e3regresion.htm). Los estudiantes pueden trabajar con el conjunto de puntos que proporciona el recurso, o bien introducir sus propios datos, que quedan representados en un diagrama de dispersión. También aparece representada el cuadrado de las desviaciones de cada punto de la nube a la recta. El programa permite “buscar a ojo la recta de regresión”. Si se pincha esta opción, aparece una recta (azul en la Figura 4), que el estudiante puede desplazar; según lo hace, van cambiando los cuadrados representados y su suma. De este modo, el estudiante puede buscar una estrategia, por ejemplo, variando la inclinación de la recta y observando el efecto sobre la suma de cuadrados, tratando de hacerla mínima.

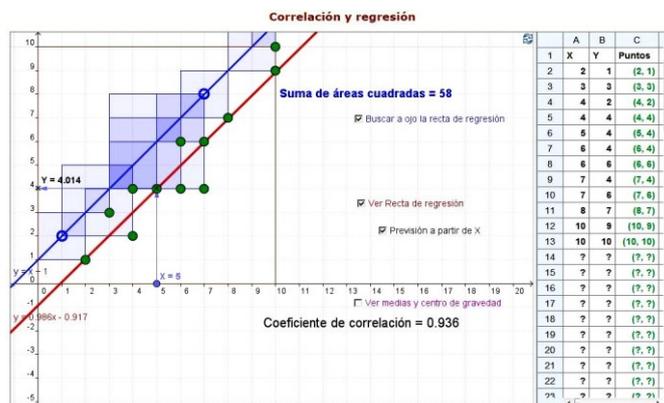


Figura 4. Applet sobre correlación y regresión en el servidor de Educación de Navarra.

Cuando el estudiante ya no es capaz de disminuir esta suma, se le puede pedir al applet que dibuje la recta de regresión y así el estudiante la puede comparar con su predicción de la misma. Con la información que se muestra en pantalla, el profesor puede ayudar a los estudiantes a interpretar los resultados con preguntas como: ¿por qué la pendiente de la recta es positiva (o negativa)? ¿Qué significa el valor a (ordenada en el origen)?, ¿cuánto aumentará el valor de la variable dependiente Y , si aumentamos el valor de la variable independiente X en dos unidades?

Además, uno de los elementos que aparece marcado en el gráfico es el centro de gravedad, que viene determinado por los valores medios de cada una de las variables. El estudiante comprobará que la recta de regresión pasa por ese punto. Una vez comprende esta propiedad, puede representar el centro de gravedad cuando comienza a buscar la recta de regresión “a ojo” para ayudarle en su tarea. Por otro lado, como el recurso representa también el coeficiente de correlación lineal de Pearson, el estudiante puede experimentar cómo varía este coeficiente modificando la posición de los puntos. Así, variando la situación de puntos y cambiándolos de lugar, se facilita la interpretación del efecto de la mayor o menor dispersión de la nube sobre este coeficiente.

En la Tabla 1 resumimos los principales objetos matemáticos implícitos en el uso del recurso mostrado en la Figura 4. Observamos la complejidad que implica trabajar con este recurso por la riqueza de objetos matemáticos que se manejan. Al mismo tiempo, el recurso permite concretizar estos objetos matemáticos. Algunas dificultades previsibles en el uso de este recurso, descritas en la investigación previa sobre la resolución de una tarea covariacional ([27], [32], [11], [33], [1], [14], [16], [34], [35]), podrían ser las siguientes:

- No distinguir entre distribución bidimensional y dos conjuntos de datos no relacionados;
- Centrarse en datos bidimensionales aislados sin considerar la tendencia global de estos;
- La ausencia de un adecuado razonamiento proporcional, tratando de deducir la correlación con sólo el estudio de frecuencias absolutas;
- Dificultades en la interpretación de los diagramas de dispersión en relación al tipo de dependencia, signo, y ajuste de una línea de regresión;
- Dificultades al relacionar el signo del coeficiente de correlación y el tipo de asociación

- (directa, inversa, independencia). Igual ocurre con el signo de la covarianza;
- Dificultad en cuanto al razonamiento sobre la asociación negativa, que suele tener un mayor índice de dificultad que la positiva;
 - La oposición entre el razonamiento numérico covariacional y el razonamiento gráfico covariacional: prevalece la estrategia de razonamiento gráfico en la estimación del coeficiente de correlación ante una descripción verbal de la situación;
 - Inducir siempre la relación de causalidad entre las variables de estudio con motivo de la existencia de una correlación significativa.

Tabla 1. Objetos matemáticos que se ejercitan con el recurso

Tipo de problema	<ul style="list-style-type: none"> - Organización/representación de datos bidimensionales - Análisis de la relación entre variables (tipo, intensidad, signo) - Ajuste de una función a los datos - Predicción utilizando la función de ajuste - Valorar la bondad de ajuste a los datos - Comparar dos modelos de ajuste (“a ojo” y mínimos cuadrados)
Concepto	<ul style="list-style-type: none"> - Datos bivariantes; coordenada - Distribución unidimensional y bidimensional - Dependencia (funcional, estadística/aleatoria, independencia) y sus tipos (directa/inversa, intensa/media/baja/nula, lineal o no) - Dispersión; Covarianza; Correlación; Coef. de correlación de Pearson - Regresión; Variable dependiente/independiente; modelo de regresión - Valor atípico; Predicción
Propiedades	<ul style="list-style-type: none"> - Los datos bidimensionales se refieren a cada variable de estudio, conforman una nueva variable y se representa en el diagrama de dispersión su distribución bidimensional - Las rectas de regresión coinciden en el centro de gravedad de la distribución bidimensional - Tipos de dependencia (funcional, estadística/aleatoria, independencia) - Covarianza y correlación: signo - Coeficiente de correlación lineal de Pearson: adimensionalidad; intensidad; tipo de dependencia; signo y dirección de la relación; signo y pendiente de la recta de ajuste - La recta de regresión hace mínima la suma de los cuadrados de las diferencias de los puntos (abscisas u ordenadas) de la distribución bidimensional a la recta - La recta de regresión permite realizar estimaciones; predicción y correlación
Lenguaje	<ul style="list-style-type: none"> - Gráfico, numérico, simbólico y tabular.
Procedimiento	<ul style="list-style-type: none"> - Representar datos bivariantes (gráfico y tabular) e interpretarlos - Interpretar la covarianza, correlación y coef. de correlación - Ajustar una recta (modelo de regresión) “a ojo” (método informal) - Interpretar la recta de ajuste mínimo cuadrática; intersección de las rectas de regresión en el centro de gravedad de la distribución - Determinar una predicción utilizando la recta de regresión
Argumento	<ul style="list-style-type: none"> - Ejemplos y contraejemplos; generalización

3.2. Otros ejemplos de recursos en Internet

Existen muchos otros recursos que pueden utilizarse para el estudio de la correlación y regresión, con actividades semejantes a las expuestas o complementarias. Por ejemplo, en la Figura 5 mostramos un applet disponible en el servidor del National Council of Teacher of

Mathematics (<http://www.nctm.org/standards/content.aspx?id=26787>), que representa más claramente las desviaciones de los puntos a la recta de regresión mínimo cuadrática. Permite igualmente elegir la recta de regresión “a ojo”. Pero, además, es posible utilizar tres métodos diferentes de ajuste: mínimos cuadrados, valor absoluto y mínima distancia. Se podría pedir a los estudiantes comparar los criterios y las rectas obtenidas por los tres métodos, y ver la ventaja de cada uno de ellos.

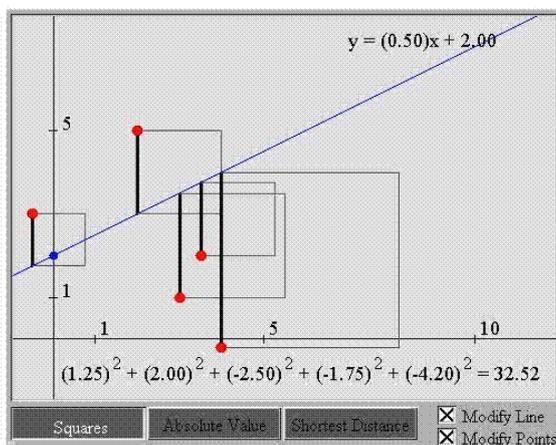


Figura 5. Applet sobre diferentes métodos de ajuste de una recta

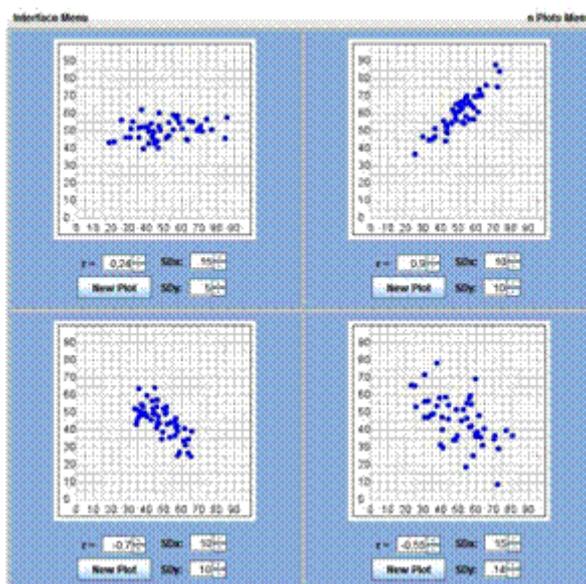


Figura 6. Applet sobre correlación

El recurso mostrado en la Figura 6 (<http://www.math.usu.edu/~schneit/CTIS/scorrelation/>) presenta también varias posibilidades:

- Variando el valor del coeficiente de correlación se obtienen nubes de puntos con dicho valor del coeficiente, bien en forma de animación (varía aleatoriamente el valor del r) o a petición del usuario. Se puede también fijar un valor de r , cambiando la desviación típica de las variables X e Y , observando cómo afectan estas a la tendencia de la nube de punto, para r fijo.
- Se pueden comparar sobre la misma pantalla hasta cuatro diagramas de dispersión simultáneos
- Otras opciones permiten asignar un coeficiente de correlación dado a un gráfico entre varias opciones, o viceversa.

En la Tabla 2 listamos otros recursos útiles para la enseñanza de la correlación y regresión.

Con ellos el profesor puede proponer actividades similares a las descritas o reforzar el aprendizaje de conceptos y propiedades.

Tabla 2. Otros recursos sobre correlación y regresión

RECURSO	ENLACE/DIRECCIÓN WEB
Biblioteca Nacional de Manipuladores Virtuales	http://nlvm.usu.edu/es/nav/frames_asid_144_g_4_t_5.html?open=activities&from=category_g_4_t_5.html
7Thales-CICA	thales.cica.es/rd/Recursos/rd99/ed99-0527-04/regresion.html
W. H. Freeman	bcs.whfreeman.com/ips4e/cat_010/applets/CorrelationRegression.html
Rossman y Chance	Correlación: http://www.rossmanchance.com/applets/guesscorrelation/GuessCorrelation.html Regresión: http://www.rossmanchance.com/applets/RegShuffle.htm
Aula virtual bioestadística	http://e-estadistica.bio.ucm.es/mod_regresion/regresion_applet_ghost.html
Roger Webster	http://www.stat.sc.edu/~west/javahtml/Regression.html
NCTM	http://illuminations.nctm.org/Activity.aspx?id=4187 http://www.nctm.org/Classroom-Resources/Interactives/Line-of-Best-Fit/
Rice Virtual Lab in Statistics	Regresión “a simple vista”: http://onlinestatbook.com/stat_sim/reg_by_eye/index.html Regresión a la media: http://onlinestatbook.com/stat_sim/reg_to_mean/index.html Efecto de transformaciones en la relación entre dos variables: http://onlinestatbook.com/stat_sim/transformations/index.html http://onlinestatbook.com/stat_sim/comp_r/index.html Fiabilidad y Regresión: http://onlinestatbook.com/stat_sim/reliability_reg/index.html

4. Notas finales

Si asumimos que los objetos matemáticos emergen del conjunto de prácticas que el estudiante realiza, como una actividad humana de “resolución de problemas” mediada por el lenguaje y “socialmente compartida”, debemos tratar que el objeto matemático trabajado en el aula adquiriera significado personal para el estudiante.

En muchas ocasiones, las nociones estadísticas se presentan a los estudiantes de un modo limitado, con un enfoque particular y encasillado, impidiendo adquirir un razonamiento superior, que permita ser generalizado. Por otra parte, el conjunto de situaciones-problema de tipo estadístico, que el estudiante trata de resolver en la clase de Matemáticas, dan la impresión de que las técnicas estadísticas son más importantes que los resultados obtenidos. En cualquier caso, el estudiante se limita a aplicar la Estadística como un conjunto de fórmulas (que no suele entender), y que ofrecen conclusiones que, en la mayoría de los casos, tristemente, no saben interpretar. En muchas ocasiones, este es el caso de la correlación y regresión.

Es aquí donde los recursos tecnológicos, como los descritos en este trabajo, pueden tener una función en el aula, proporcionando a los estudiantes micromundos en que las ideas abstractas se representen y manipulen de una forma más concreta.

También se podrían complementar estas actividades con los ejercicios propuestos en los libros de texto, así como pequeños proyectos estadísticos en los que los estudiantes recogiesen sus

propios datos para analizar la correlación entre ellos. Por ejemplo, los datos de medidas físicas de los estudiantes como peso, talla, perímetro de cintura, longitud del antebrazo o la pierna son fáciles de tomar y les permitiría realizar una introducción al tema; analizando las diferencias entre chicos y chicas se podría conseguir el trabajo con otros contenidos, como las distribuciones univariantes e incluso una introducción intuitiva a la inferencia.

En este sentido, conviene también proponer tareas con datos tomados de Internet (por ejemplo, de institutos de estadística), ya que permitirán al estudiante adquirir un sentido profundo de lo que implica un estudio estadístico de fenómenos sociales, tal y como se propone en las concreciones curriculares [3].

En concreto, creemos que puede ser de utilidad el análisis realizado de uno de los applets, incluyendo los objetos matemáticos trabajados y las posibles dificultades de los estudiantes. Esta información puede ser de interés en la planificación del trabajo en el aula y de la formación de profesores. Coincidimos con [36] en que *“el reto para el formador de profesores es encontrar formas de mejorar el conocimiento pedagógico y el conocimiento del contenido estadístico del futuro profesor”* ([36], p. 5). Es decir, proporcionar al futuro profesor un conocimiento del contenido para la enseñanza [37], que le haga competente para desarrollar las nociones que aparecen en los currícula.

Así, un aspecto importante de la formación del profesorado, en cuanto a la enseñanza de la noción de asociación estadística, consiste en el dominio del conocimiento de las posibles dificultades y errores de tipo filosófico, epistemológico y didáctico, que ligadas a la interpretación y aplicación del concepto puedan presentarse en los estudiantes. Al respecto, ofrecemos en este trabajo diferentes elementos a considerar para una completa formación del profesorado. Esperamos que la descripción realizada del sentido de la correlación y sus componentes pueda motivar a los profesores para tratar de desarrollarlos en los estudiantes, de forma que adquieran una comprensión profunda del tema. Pensamos que los recursos descritos disponibles en Internet y otros semejantes, junto con las actividades sugeridas, puedan ayudar al profesorado en esta labor.

Agradecimientos: Proyecto EDU2013-41141-P (MEC), Grupo FQM126 (Junta de Andalucía) y Beca CONICYT PFCHA 72150306.

Referencias

- [1] Batanero, C., Estepa, A., Godino, J. D. y Green, D. R. (1996). “Intuitive strategies and preconceptions about association in contingency tables”. *Journal for Research in Mathematics Education*, 27, 151-169.
- [2] 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.
- [3] 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.
- [4] Moore, D. S. (1999): “Discussion: What shall we teach beginners?” *International Statistical Review*, 67(3), 250-252.
- [5] Moore, D. S. (2004): “Foreword”. En D. Ben-Zvi y J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*, (pp. ix-x). Dordrecht (The Netherlands): Kluwer Academic Publishers.
- [6] Pfannkuch, M. y Wild, C. (2004): “Towards an 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 Netherlands: Kluwer Academic Publishers.
- [7] Batanero, C., Díaz, C., Contreras, J. M. y Roa, R. (2013): “El sentido estadístico y su desarrollo”. *Números*, 83, 7-18.
- [8] Burrill, G. y Biehler, R. (2011): “Fundamental statistical 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. 57-69). Dordrecht: Springer.

- [9] Wild, C. y Pfannkuch, M. (1999): "Statistical thinking in empirical enquiry". *International Statistical Review*, 67 (3), 221-248.
- [10] Gea, M., Batanero, C. y Roa, R. (2014): "El sentido de la correlación y regresión". *Números*, 87, 25-35.
- [11] Estepa, A. (2008): "Interpretación de los diagramas de dispersión por estudiantes de Bachillerato". *Enseñanza de las Ciencias*, 26 (2), 257-270.
- [12] Estepa, A. y Batanero, C. (1995): "Concepciones iniciales sobre la asociación estadística". *Enseñanza de las Ciencias*, 13 (2), 155-170.
- [13] Zieffler, A. y Garfield, J. (2009): "Modeling the growth of students' covariational reasoning during an introductory statistics course". *Statistics Education Research Journal*, 8 (1), 7-31.
- [14] 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: Internacional Statistical Institute.
- [15] Truran, J. (1994): "Examination of a relationship between children's estimation of probabilities and their understanding of proportion". En J. P. Ponte y J. F. Matos (Eds.), *Proceedings of the XVIII Psychology of Mathematics Education Conference*, (Vol. 4, pp. 337-344). Lisboa: Universidad de Lisboa.
- [16] Sánchez Cobo, F. T. (1998): "Significado de la correlación y regresión para los estudiantes universitarios". Tesis doctoral. Universidad de Granada.
- [17] 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.
- [18] Moore, D. S. (1992): "Teaching Statistics as a respectable subject". En F. S. Gordon y S. P. Gordon (Eds.), *Statistics for the Twenty-First Century*, (pp. 14-25). Washington, DC: Mathematical Association of America.
- [19] Cañadas, G. R., Batanero, C., Gea M. M., y Contreras, J. M. (2014): "Comprensión de frecuencias asociadas a las tablas de contingencia por estudiantes de psicología". *Unipluriversidad*, 13(3), 97-108.
- [20] Antonio, M., González, L., Lorenzo, J., Molano, A., del Río, J., Santos, D. y de Vicente, M. (2009): *Matemáticas aplicadas a las Ciencias Sociales I*. Madrid: Santillana Educación.
- [21] Holmes, P. (2001): "Correlation: From Picture to formula". *Teaching Statistics*, 23 (3), 67-71.
- [22] Arias, J. M. y Maza, I. (2011): *Matemáticas aplicadas a las Ciencias Sociales 1*. Madrid: Grupo Editorial Bruño.
- [23] Moore, D. S. (1995): "Estadística aplicada básica". Barcelona: Antoni Bosch.
- [24] Gea, M. M., Batanero, C., Cañadas, G. R. y Contreras, J. M. (2014): "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: Sociedad Española de Investigación en Educación Matemática (SEIEM).
- [25] Vizmanos, J. R., Hernández, J., Alcaide, F., Moreno, M. y Serrano, E. (2008): "Matemáticas aplicadas a las Ciencias Sociales 1". Madrid: Ediciones SM.
- [26] Barbancho, A. G. (1973): "Estadística elemental moderna". Barcelona: Ariel.
- [27] 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. Universidad de Granada.
- [28] Batanero, C., Díaz, C. y Gea, M. M. (2011): "Estadísticas de la pobreza y desigualdad". En Batanero, C. y Díaz, C. (Eds.), *Estadística con proyectos*, (pp. 97-124). Granada:

Departamento de Didáctica de la Matemática.

[29] Ridgway, J., Nicholson, J. y McCusker, S. (2008): "Mapping new statistical literacies and illiteracies". Trabajo presentado en el 11th International Congress on Mathematics Education, Monterrey, México.

[30] Contreras, J. M. (2009): "Recursos en internet para la enseñanza de la probabilidad condicionada". Tesis de Máster. Universidad de Granada.

[31] Contreras, J. M. (2011): "Evaluación de conocimientos y recursos didácticos en la formación de profesores sobre probabilidad condicional". Tesis Doctoral. Universidad de Granada.

[32] Estepa, A. (2007): "Caracterización del significado de la correlación y regresión en estudiantes de Educación Secundaria". Zetetiké, 15 (28), 119-151. ISSN 0104-4877.

[33] 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.

[34] 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: Kluwer.

[35] Zieffler, A. S. (2006): "A longitudinal investigation of the development of college students' reasoning about bivariate data during an introductory statistics course". Tesis doctoral. Universidad de Minnesota.

[36] Pfannkuch, M. (2008): "Training Teachers to develop Statistical Thinking. En C. Batanero, G. Burrill, C. Reading y A. Rossman (Eds.), Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference.

[37] Ball, D. L., Thames, M. H. y Phelps, G. (2008): "Content knowledge for teaching: what makes it special?" Journal of Teacher Education, 59 (5), 389-407.