

# Razonamiento de los ESTUDIANTES de bachillerato FRENTE a un problema de regresión

Miguel Napoleón Medina Delgado/  
Gabriel Esteban Olay Blanco/  
Ernesto Alonso Sánchez Sánchez

Recibido: 17/02/2016  
Aprobado: 20/03/2016

**Resumen:** En este artículo se hace una reflexión, a partir de las respuestas de los estudiantes a un problema de regresión lineal, sobre la importancia de destacar el aspecto estadístico en el diseño de la enseñanza de los temas de la materia. En particular se describen las enseñanzas que nos arroja analizar las respuestas de los estudiantes al problema: su creencia de que el problema es de cálculo y determinista; el papel de los recursos gráficos y la sub-utilización de los datos. Se concluye enfatizando la necesidad de ver a los problemas de estadística ubicados en situaciones de incertidumbre y no como problemas de cálculo.

**Palabras clave:** regresión lineal, razonamiento, enseñanza, incertidumbre.

**Abstract:** This article is a reflection, from the student responses to a problem of linear regression, on the importance of highlighting the statistical aspect in the design of the teaching of the subject topics. In particular the lessons we analyze throws student responses to the problem are described: his belief that the problem is calculation and deterministic; the role of graphics resources and the underutilization of data. It concludes by emphasizing the need to see the problems of statistics located under uncertainty rather than calculation problems.

**Keywords:** linear regression, reasoning, learning, uncertainty.

## Introducción

La segunda Unidad de la asignatura de Estadística y Probabilidad I del Colegio de Ciencias y Humanidades (CCH) se llama "Datos Bivariados" e incluye la *regresión lineal* y la *correlación*. Estos temas forman parte de casi cualquier programa de estudios de Probabilidad y Estadística de Bachillerato. Por otro lado, en la vida diaria y profesional las personas deben entender y tomar decisiones ante problemas relacionados con regresión y correlación; por ejemplo, frecuentemente se presentan preguntas como las siguientes: ¿el poder adquisitivo de la gente es la causa de la disminución en las ventas de una compañía?, ¿las bajas calificaciones de

## Nosotros

un estudiante en un examen son el resultado de pocas horas dedicadas al estudio?, ¿de qué forma se relaciona el tratamiento médico suministrado a un grupo de pacientes, con su recuperación de cierto padecimiento? o ¿existe relación entre la estatura de los hijos y la de los padres?; esta última pregunta representa un problema que históricamente dio origen a la idea de regresión (Gea, 2013: 5). El estudio de los temas de regresión y correlación debe ayudar a entender y responder este tipo de preguntas, como lo plantea Crocker (1981: 272): "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".

Desde hace tiempo algunos estudios (Estepa y Batanero, 1996: 36; Batanero, Estepa y Godino, 1997: 194) han mostrado que los estudiantes se dejan llevar por sesgos o concepciones erróneas ante problemas de asociación estadística, dichos autores mencionan las siguientes: a) Concepción determinista de la correlación, b) Concepción local de la correlación, c) Concepción unidireccional y d) Concepción causal; es de esperar que los alumnos de bachillerato eventualmente caigan en errores derivados de ellas. Por lo anterior, es importante que los profesores busquen formas alternativas de diseñar actividades con el fin de mejorar los procesos enseñanza-aprendizaje de sus estudiantes; pero para que esto sea posible de manera efectiva, es importante que conozcan más cómo razonan espontáneamente frente a problemas de correlación y regresión. En este artículo informamos y reflexionamos acerca de los resultados de un estudio cuyo objetivo era explorar el razonamiento de los estudiantes frente a un problema de regresión.



Fotografía: Archivo Histórico Fotográfico del Colegio de Ciencias y Humanidades, S.C.I., 2013

## La regresión

En el análisis de regresión se estudia la posible relación entre una variable dependiente y otra u otras variables independientes, relación que se representa mediante un modelo matemático llamado la *ecuación de regresión*. Cuando esta ecuación es lineal respecto a un conjunto de parámetros se habla de *regresión lineal*. Por otro lado, la *correlación* es una medida de la intensidad de la relación entre dos variables la cual se expresa con un número llamado el *coeficiente de correlación*. Las relaciones precisas entre variables estadísticas generalmente son desconocidas y sólo se pueden estimar a partir de conjuntos de datos.

## El problema

Una aplicación de un modelo de regresión es predecir aproximadamente el valor de una variable cuando se conoce el valor que toma la otra variable. Un problema de predicción accesible para los estudiantes es el siguiente (adaptado de Moore, 1998:106-107):

El consumo de gas: La familia Morales está a punto de instalar paneles solares en su casa para reducir el gasto en la calefacción. Para conocer mejor el ahorro que puede significar la instalación de dichos paneles, los Morales han ido registrando su consumo de gas durante el último año y medio. En la tabla se muestran los datos con el promedio del consumo de gas y de la temperatura media ambiental de cada mes:

| Mes              | Nov  | Dic  | Ene  | Feb  | Mar  | Abr   | May  | Jun  | Jun  |
|------------------|------|------|------|------|------|-------|------|------|------|
| Temperatura (°C) | 5.2  | -9.8 | -5.4 | 0.2  | 4.1  | 11.3  | 16.3 | 18.5 | 18.5 |
| Gas (m3)         | 17.6 | 30.5 | 24.9 | 21   | 14.8 | 11.2  | 4.8  | 3.4  | 3.4  |
| Mes              | Ago  | Sep  | Oct  | Nov  | Dic  | Ene   | Feb  | Mar  | Abr  |
| Temperatura (°C) | 18   | 15.2 | 11.8 | 1.8  | 0.7  | -10.4 | 1.8  | 17.1 | 10.7 |
| Gas (m3)         | 3.4  | 5.9  | 8.7  | 17.9 | 20.2 | 30.8  | 19.3 | 6.3  | 10.7 |

Según los datos mostrados en la tabla, responde:

**a** Si la temperatura media registrada en un mes es de 8 °C ¿Cuál es el consumo de gas esperado por la familia Morales en dicho mes? (Situación 1a)

**b** La familia registró un consumo de gas de 23 m<sup>3</sup> en otro mes, ¿cuál fue la temperatura media ambiental de ese mes? (Situación 1b)

La solución formal de este problema consiste en encontrar la recta de regresión  $y = mx + b$  y entonces para resolver el inciso "a", sustituir  $x = 8$  en la ecuación y encontrar el valor de  $y$ . En cambio, dado un valor de  $y = 23$  (inciso b) encontrar el valor de  $x$  no se obtiene simplemente sustituyendo, porque la relación no es necesariamente invertible (la temperatura es una causa de consumo de gas, pero el consumo de gas no lo es de la temperatura) (ver pormenores en Moore, 1998: 142). No obstante, asumiremos que es posible deducir una aproximación, a manera de indicador, simplemente encontrando el valor de  $x$  correspondiente a ese valor de  $y$ . Concretamente, el modelo de regresión lineal obtenido, es decir, la recta de regresión de los datos del problema, que se obtuvo con ayuda del software dinámico *Fathom*, es:  $G = -0.938T + 20.7$ . Entonces para el inciso "a" se tiene:  $G \text{ m}^3$  y para el inciso "b":  $T$  de donde al despejar la temperatura:  $T = -2.45^\circ \text{C}$ . Entonces si la temperatura media del mes es de 8°C habrá un consumo aproximado de 13.19 m<sup>3</sup> de gas. Cuando hay un consumo de 23 m<sup>3</sup> se puede conjeturar que bajo ciertas condiciones es probable que la temperatura fuera muy baja, -2.45°C es sólo un indicador.

Los profesores sabemos la gran dificultad que representa la realización, por parte de los estudiantes, del anterior procedimiento y ubicamos, con buenas razones, que el núcleo de las dificultades es el cálculo de la recta de regresión. Es claro que cada profesor piensa y ensaya formas diferentes de hacer menos difícil para el estudiante el aprendizaje de los temas de asociación estadística. En nuestra opinión, la mayoría de esos intentos, que por supuesto son parte esencial de las actividades del profesor, carecen de hipótesis acerca de cómo razonarían los estudiantes acerca de la situación y cómo utilizarían los datos para dar una respuesta si simplemente se enfrentan a ella sin ninguna enseñanza previa específica sobre el tema. A muchos profesores les sorprende la pregunta ¿Cómo razonarían

sus estudiantes si simplemente se les formula el problema sin ninguna enseñanza previa?. Muy probablemente la respuesta de cualquier profesor sería que los estudiantes no sabrán resolver el problema correctamente. No obstante, un enfoque constructivista de la enseñanza de las matemáticas considera que para que el profesor pueda diseñar la instrucción de algún tema de manera productiva, debe saber cómo razonarían sus estudiantes ante los problemas con los recursos que hasta ese momento han adquirido; éstos incluyen conocimientos específicos, pero también habilidades de razonamiento y actitudes ante los problemas y la materia. Los estudiantes del nivel ya han acumulado herramientas que deben estar en la base de los nuevos procedimientos que van a aprender y conviene saber de qué manera afloran cuando se enfrentan a los problemas. En resumen, la pregunta que debemos formular es ¿qué podemos aprender los profesores de los estudiantes?

Los autores nos hicimos la pregunta: ¿Cómo razonan los estudiantes frente al problema de predicción del consumo de gas sin que ellos hayan estudiado antes el tema de regresión lineal?. Por ejemplo, nos preguntamos si los estudiantes recurren a la idea de dibujar la nube de puntos y ver la tendencia de los datos. Si esto fuera así, constataríamos que la gráfica es una herramienta a la mano para los estudiantes. En caso de que no utilicen este recurso espontáneamente, si se les sugiere su uso ¿podrían avanzar en el problema? y de ser así ¿hasta dónde? Las respuestas a estas preguntas podrían ayudar en el proceso de enseñanza-aprendizaje.

Fotografía: Archivo Histórico Fotográfico del  
Colegio de Ciencias y Humanidades,  
S.C.I., 2013



Se llevó a cabo una investigación que consistió en aplicar el problema del gas a dos grupos de Bachillerato: uno con 31 estudiantes de la escuela pública CBTIS, no. 204 en el estado de Michoacán y otro con 40 alumnos del CCH, plantel Vallejo en la Ciudad de México. Los estudiantes cursaban 5° y 6° semestre, respectivamente.

La aplicación se hizo en dos momentos (dos sesiones de 2 horas cada una); en el primero, se les proporcionó el problema del consumo de gas (como se presentó arriba) y se pidió que lo resolvieran echando mano de los recursos que creyeran conveniente; en el segundo momento, se les sugirió que trazaran la nube de puntos y ahora, con base en la gráfica, volvieran a responder las preguntas de los incisos “a” y “b”.

# Resultados

Expondremos brevemente los patrones generales de respuestas de los estudiantes sin describir de manera particular cada una de ellas. En el primer momento, ningún estudiante intentó graficar la nube de puntos para dar una respuesta. La casi totalidad de los estudiantes realizaron operaciones aritméticas (en el grupo del CBTIS se ubicaron 81.8% de respuestas y para el grupo del CCH un 82.5% en la situación y con respecto a la situación Ib, se tiene un 89.5% en el CBTIS y 81.1% en el CCH) para encontrar el valor pedido o bien respondieron verbalmente, pero de manera equivocada. Un hecho relevante es que, salvo dos excepciones, los estudiantes no involucraron en sus cálculos la totalidad de los datos. La mayoría de los procedimientos consistieron en elegir una pareja de datos para formar junto con el dato dado (8 en el inciso "a", y 23 en el inciso "b") un conjunto de 3 datos y entonces formular y resolver una proporción o regla de tres.

En el segundo momento, la totalidad de los estudiantes no tuvo problemas en trazar la nube de puntos. A pesar de esto, muchos volvieron a presentar en sus respuestas rasgos similares a las del primer momento, es decir, eligen una pareja y junto con el dato dado forman una terna de números con los que forman una proporción. Hubo, sin embargo, una clase nueva de respuestas en las que se pone en evidencia que algunos estudiantes (18.2% en el CBTIS y 17.5% en el CCH para la situación Ia y 10.5% en el CBTIS y 18.9% en el CCH para la situación Ib) ven en la gráfica el sentido de la relación, dicen por ejemplo "entre más calor menos consumo de gas". Esta relación no necesariamente la descubren los estudiantes mediante la gráfica, pero ninguno la formuló en el primer momento, sino sólo se presentaron casos en el segundo momento.

# Discusión

¿Qué podemos aprender de las respuestas de los estudiantes? Las respuestas obtenidas propician varias reflexiones que pueden ayudar a diseñar mejor las actividades de enseñanza. Hemos destacado los siguientes puntos:

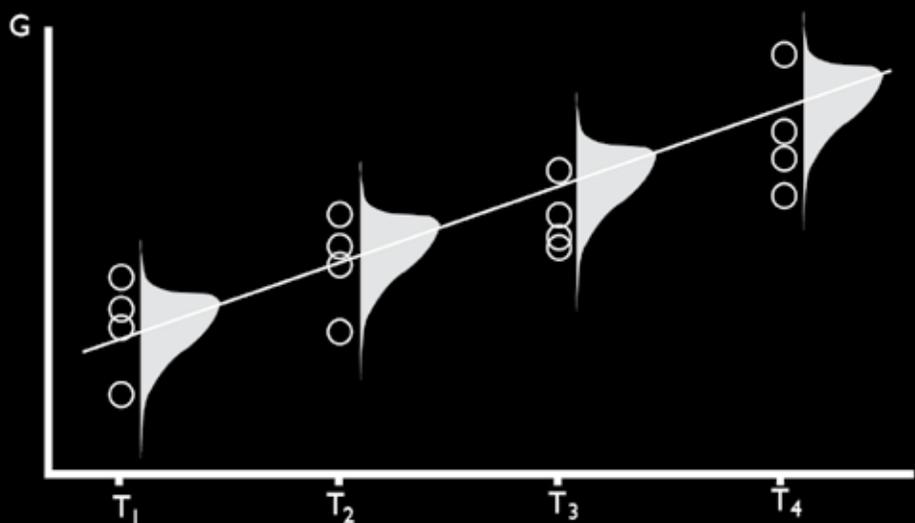
- Los estudiantes ven al problema de regresión como de proporcionalidad o, más en general, como problema de cálculo (incluso algunos profesores lo piensan así).
- La gráfica abre un poco las posibilidades de acercarse al problema propiciando que se formule en algunos casos la asociación general entre las variables.
- Los estudiantes tienen la dificultad de imaginar procedimientos que utilicen todos los datos.

A partir de que las respuestas de los estudiantes muestran que utilizan procedimientos de proporcionalidad para encontrar la solución del problema se puede

## Nosotros

deducir que ellos creen que es un problema sólo de cálculo y, por lo tanto, que la respuesta está determinada y es única. Debe tenerse en cuenta que esta idea permanecería en el estudiante, aunque supiera calcular la recta de regresión. Esto nos permite formular una pregunta y una reflexión; la primera es ¿cómo se debe entender el problema desde un punto de vista estadístico?; la segunda, consiste en revisar la creencia generalizada de que la dificultad del problema para los estudiantes se reduce a saber calcular la recta de regresión.

Con relación a la pregunta diremos que se debe tener en cuenta que cualquier resultado estadístico mantiene incertidumbre y, en consecuencia, no hay soluciones determinadas y únicas. El contexto del gasto de gas en relación con la temperatura se presta para desarrollar esta idea. El consumo de gas de una familia, en particular de los Morales, es influida por la temperatura, pero no tiene por qué estar determinada por ella; hay consumo de gas por otras circunstancias contingentes que no dependen de la temperatura, por ejemplo, al usar el horno para cocinar un pastel por el cumpleaños de un miembro de la familia. También puede haber ahorro de gas cuando la familia o algunos de sus miembros salen de vacaciones y no están en la casa por varios días. De esta manera se debe entender que la temperatura de  $8^{\circ}\text{C}$  no determina del todo cuánto gas se va a consumir; de manera análoga, que se hayan consumido  $23\text{ m}^3$  de gas no determina del todo la temperatura (Una fuga de gas no implica nada acerca de la temperatura). No obstante, uno de los factores que más influye en el consumo de gas es la temperatura, pues si hace mucho frío se utiliza más el agua caliente y se utilizan calentadores de gas; esto lo indican los datos dados. Se puede pensar que en condiciones ideales hay una relación afín entre Gasto de gas ( $G$ ) y Temperatura ( $T$ ), es decir, de la forma:  $G_i = aT_i + b$ ; pero debido a las contingencias que en gran medida se deben al azar la ecuación anterior se transforma en:  $G_i = aT_i + b + e_i$ , donde  $e_i$  es un valor aleatorio que se denomina "error". Este error suele variar de manera normal alrededor del cero, por lo que la siguiente representación ofrece una idea de cómo debiera pensarse la relación entre las dos variables perturbada por errores.



Ahora nos preguntamos: ¿Lo más difícil para los estudiantes es calcular la recta de regresión o entender el sentido del problema? Si el cálculo de la recta de regresión fuera la esencia del problema, la solución sería enseñar a los estudiantes el uso de una calculadora graficadora o un software estadístico, los cuales tienen la función de calcular la recta de regresión de un conjunto de datos. Al creer que lo sustancial de un problema como el anterior es encontrar la recta de regresión estamos suponiendo que la discusión conceptual detrás de ésta es transparente para los estudiantes. Si así fuera, el tipo de respuestas que habríamos recogido serían consideraciones como “no se puede saber con exactitud por...”, o se hubieran calculado aproximaciones más plausibles, como la de elegir dos puntos alrededor del punto 8 y calcular el otro valor mediante interpolación (lo mismo para 23). Los resultados nos llevan a pensar que las verdaderas dificultades son conceptuales y no de cálculo; aunque ciertamente éstos no sean fáciles.

Los datos también nos revelan que para los estudiantes dibujar la nube de puntos para ver la tendencia de los datos no es un recurso que utilicen espontáneamente. No consideran que puedan avanzar en la solución del problema realizando la gráfica de los puntos. Probablemente, la utilización de la nube de puntos es frenada o inhibida por la creencia de que el problema se debe resolver mediante un cálculo. Fue necesario sugerir a los estudiantes que hicieran la gráfica, y dado que todos pudieron elaborarla, concluimos que contaban con este recurso. Una vez que trazaron la nube de puntos, algunos estudiantes percibieron cualitativamente la relación general de las variables “a menor temperatura, mayor gasto de gas”. Pero ninguno asoció un procedimiento aritmético o algebraico que le permitiera hallar un valor utilizando esa relación general.

Como se ha indicado la mayoría de estudiantes que utilizaron métodos aritméticos eligieron un dato bivariado y con el 8 (o el 23) realizaron una proporción. Conviene enunciar una dificultad que se refleja en lo anterior: la inhabilidad de estos estudiantes para utilizar todos los datos. En el enunciado se dan 12 datos bivariados, ¿por qué creen que para resolverlo sólo deben utilizar uno? La razón puede ser su creencia de que el valor debe encontrarse mediante una regla de tres. Hay, sin embargo, respuestas de algunos estudiantes donde utilizaron todos los datos; consisten en obtener la media aritmética de los valores de la variable “gasto de gas” y la media aritmética de los valores de la variable “Temperatura”. Con estos datos, junto con el 8 o el 23 se formula una proporción. Este procedimiento también es incorrecto, pero refleja un rasgo estadístico más avanzado al considerar las medias aritméticas de los valores de las variables y, con ello, incluir todos los datos en la propuesta de solución.

## Conclusiones y Propuesta

Quizá el obstáculo más importante para darle sentido a los problemas de regresión, como el explorado en este trabajo, es verlos como situaciones de incertidumbre, es decir, que entre la multiplicidad de causas (el aumento o disminución de habitantes de la casa, el costo del combustible, cambio en los hábitos de consumo, etcétera) que producen un efecto, una de ellas es la predominante (temperatura media ambiental) y produce una tendencia, pero no lo hace de manera determinista porque las otras causas producen ligeras desviaciones. Entender y resolver

los problemas en Estadística no es sinónimo de saber el procedimiento de cálculo para obtener una respuesta, sino de saber reflexionar sobre la situación en juego, manipular los datos con un objetivo preciso e interpretar los resultados teniendo en cuenta la incertidumbre. Para optimizar el proceso enseñanza-aprendizaje de los temas de estadística, el profesor además de tener en cuenta lo anterior, debe explorar el pensamiento de sus estudiantes; pues partiendo de este conocimiento será capaz de proponerles formas adecuadas para transitar a niveles más altos de razonamiento estadístico.

Finalmente, los autores proponemos como puntos relevantes para el diseño de actividades de enseñanza: 1) Fomentar el trazado y uso de representaciones gráficas (diagramas de dispersión), se recomienda el uso de softwares y simuladores dinámicos; 2) Reconocer la importancia de las respuestas informales de los estudiantes; y 3) Considerar la presencia de conceptos clave como incertidumbre, variación y error en el diseño, por encima de los procedimientos de cálculo.

## Referencias

- Batanero, C., Estepa, A. y Godino, J.D. (1997). Evolution of students' understanding of statistical association in a computer based teaching environment. En Garfield, J.B. y Burrill, G. (eds.). *Research on the Role of Technology in Teaching and Learning Statistics. IASE Round Table Conference Papers*. Voorburg: International Statistical Institute.
- Crocker, J. (1981). Judgment of covariation by social perceivers. En *Psychological Bulletin*, Núm. 90.
- Estepa, A. y Batanero, C. (1996). Judgments of correlation in scatter plots: students' intuitive strategies and preconceptions. En *Hiroshima Journal of Mathematics Education*, Núm. 4.
- Gea, M. M., Batanero, C. y Cañadas, G. R. (2013). "Un estudio empírico de los problemas de correlación y regresión en libros de texto de Bachillerato." En *Actas III Encontro de Probabilidades e Estatística na Escola*. Braga, Portugal: Centro de Investigacao em Educacao (CIEd). Universidade do Minho
- Moore, D. S. (1998). *Estadística aplicada básica*. Antoni Bosch editor.