Clame Acta Latinoamericana de Matemática Educativa

SECCIÓN 1 / ANÁLISIS DEL DISCURSO MATEMÁTICO ESCOLAR

VOL 31, NÚMERO 1, AÑO 2018

PREDICCIÓN INFORMAL EN UN DIAGRAMA DE PUNTOS

Ana García, Gabriel Yáñez
Universidad Industrial de Santander. (Colombia)
anna.garciamado@gmail.com, gyanez@uis.edu.co

Resumen

En este artículo se reporta la segunda fase de una investigación en curso que busca responder a la pregunta ¿Cuáles son las formas de Razonamiento Covariacional Informal (RCI) de los estudiantes alrededor de la recta de mejor ajuste en un ambiente computacional? En primer lugar se aplicó una prueba a 39 estudiantes de octavo grado que no habían estudiado previamente conceptos alrededor del análisis de datos bivariados. Del análisis de los resultados de esta prueba surgieron una serie de hipótesis que permitieron diseñar una segunda prueba la cual fue aplicada a 40 estudiantes con las mismas características. El análisis de los resultados de estas dos pruebas nos proporcionó información sobre el RCI actual de los estudiantes, evidenciándose estrategias y concepciones ya reportadas en investigaciones anteriores y también algunas novedades. En particular, se analizan las estrategias de predicción que utilizan los estudiantes, así como las de ubicación de la recta de mejor ajuste.

Palabras clave: razonamiento informal, predicción, recta de mejor ajuste

Abstract

This article reports the second phase of a current research that seeks to answer the question: Which are the forms of students' Informal Co-variational Reasoning (ICR) on the best fit straight line in a computational environment? First, we applied a test to 39 eighth-grade students who had not previously studied concepts on the analysis of bi-variation data. From the analysis of the results of this test a series of hypotheses emerge. It allowed us to design a second test which was applied to 40 students with the same characteristics. The analysis of the results of these tests provided us with information about the students' current ICR, showing strategies and conceptions previously reported, as well as some new ones. Particularly, we analyze the prediction strategies that students use, as well as those of the location of the best fit straight line.

Key words: informal reasoning, prediction, best fit line

Introducción

En investigaciones alrededor del razonamiento de los estudiantes surge la necesidad de explorar y aprovechar sus conocimientos previos, ya sean formales o informales, así como sus creencias, experiencias, intuiciones e inclusive errores para diseñar el camino que los conduzca hacia la comprensión de conceptos matemáticos. En particular, en Estadística, en los años recientes, se ha venido investigando alrededor del Razonamiento Informal (RI) que utilizan los estudiantes con el ánimo de impulsar su desarrollo ya que se asume que es el paso previo para la comprensión posterior de los modelos formales.

Respecto al Razonamiento Covariacional (RC) se han hecho investigaciones desde diferentes disciplinas como la psicología, las matemáticas, la estadística, la educación matemática y la educación estadística (Zieffler, 2009; Garfield y Ben-Zvi, 2008) en las que se han identificado sesgos, concepciones, dificultades, errores e incluso trampas alrededor del análisis de datos bivariados. También se han identificado formas de razonamiento covariacional que han permitido generar instrumentos y trayectorias para la enseñanza y el aprendizaje de este tema.

Dentro de las concepciones previas que tienen los estudiantes en cuanto a la asociación estadística en general, está la *Causalista, Determinista, Unidireccional* y *Localista* (Estepa y Batanero, 1995), inclusive en estudiantes que ya han estudiado el tema. Por otro lado, la dependencia estadística de dos variables no es tan familiar como sí lo es la dependencia funcional o determinista; inclusive entre los profesores, ya que muchos de ellos tienen un pensamiento determinista sobre datos bivariados. De manera que, para una comprensión adecuada de la correlación y la regresión se necesita una comprensión no solo acerca de las funciones sino también de la variación (Engel y Sedlmeier, 2011).

Investigaciones más recientes muestran cómo los estudiantes razonan sobre la “forma” de la relación entre dos variables cuantitativas, en particular sobre la ubicación de la recta de mejor ajuste. Casey (2015) investiga las concepciones que tienen 33 estudiantes de octavo grado sobre la recta de mejor ajuste antes de su instrucción formal. Entre los resultados están los criterios y métodos que usan los estudiantes para la posición de la recta de mejor ajuste, así como la precisión en que lo hacen.

Criterios y métodos para la posición

Se identifican siete métodos: i) a través de tantos puntos como sea posible; ii) igual número de puntos a ambos lados; iii) cerca de todos los puntos como sea posible; iv) reflejar la relación que tienen las variables con base al conocimiento del contexto; v) a mitad de camino entre los puntos mínimo y máximo; vi) a través de los puntos primero y último y; vii) comenzando desde el primer punto y luego trazar a través del máximo número de puntos.

Precisión en la posición

Existe una considerable variabilidad entre los estudiantes para ubicar la recta. Algunos la ubican muy cerca y paralela a la recta de regresión. Otros utilizan el criterio de que pase por la mayoría de puntos, inclusive un estudiante llama la atención de que la recta debe pasar por el origen. Algunos, reflejando un razonamiento univariado, trazan rectas horizontales para dejar igual número de puntos a cada lado, es decir buscando una aplicación para la media o mediana.

Como se mencionó anteriormente, la identificación de errores, concepciones y formas de razonamiento, así como el análisis de contenido alrededor del análisis de datos bivariados, ha tenido implicaciones en la enseñanza de este tema. Dentro de las implicaciones están las investigaciones que tienen como resultado final el diseño de una trayectoria de aprendizaje o de una secuencia de actividades que permiten desarrollar el razonamiento covariacional de los estudiantes. Ejemplos de este tipo de investigación son los trabajos de Cobb, McClain y Gravemeijer (2003), Garfield y Ben-Zvi (2008) y Casey (2015).

Marco Conceptual

Conocer cómo los estudiantes razonan y piensan estadísticamente ha sido el foco de muchos investigadores en educación estadística; lo anterior, argumentándose en que la enseñanza tradicional de la estadística se reduce a procedimientos y cálculos en donde los estudiantes no desarrollan estos procesos: pensar y razonar estadísticamente (Ben-Zvi y Garfield, 2004).

> El *razonamiento estadístico* puede definirse como la forma en que las personas razonan con ideas estadísticas y le dan sentido a la información estadística. Lo que implica realizar interpretaciones basadas en conjuntos de datos, representaciones de datos o resúmenes estadísticos de datos. Implica también la conexión de un concepto con otro (por ejemplo el centro y la variabilidad), o combinar ideas sobre datos y el azar. Razonar significa entender y ser capaz de explicar los procesos estadísticos y ser capaz de interpretar completamente los resultados estadísticos (Garfield y Ben-Zvi, 2004, p. 7).

Por otra parte, el razonamiento en un sentido informal, RI, se entiende como la forma en que el estudiante usa su Conocimiento Informal (CI), el cual hace referencia, por una parte, a ese conocimiento cotidiano y a las experiencias que se adquieren fuera del salón de clases, y, por otra parte, al conocimiento que resulta luego de una instrucción formal previa, en otras palabras, “es el punto de partida para el desarrollo del conocimiento formal” (Zieffler et al., 2008, p.3). Como nuestro interés es la asociación estadística entre dos variables cuantitativas, resulta necesario definir lo que aquí se entiende por Razonamiento Covariacional (RC) y RCI.

Después de examinar las definiciones que se le han dado al Razonamiento Covariacional desde los diferentes campos en que se ha estudiado, Zieffler (2006, p.6) concluye que “el razonamiento covariacional se refiere a la forma en que la gente piensa acerca de, o razona acerca de la relación entre dos o más variables”, por ejemplo a partir de la lectura de un diagrama de dispersión, o en la interpretación de las correlaciones y en otras tareas que conlleven al análisis de dos variables así como la interpretación de los resultados arrojados en ese análisis.

Finalmente, en un intento por aunar las definiciones que Zieffler et al. (2008) hacen sobre CI y RI con la definición de RC, adoptamos la definición de RCI como la forma en que la gente razona y argumenta acerca de la relación entre dos o más variables haciendo uso solo de su RI y CI. Claramente, parte del RCI está el razonar sobre un diagrama de dispersión: la existencia o no de una relación entre dos variables, la predicción del valor de la variable respuesta para un valor de la variable explicativa, así como la fuerza y la forma con que se relacionan, por ejemplo, la recta de mejor ajuste.

Metodología

Participantes

La primera prueba se aplicó a 39 estudiantes de octavo grado, entre los 12 y 13 años, quienes no habían visto previamente conceptos de estadística como diagramas de dispersión y recta de mejor ajuste, inclusive sus antecedentes sobre el concepto como tal de recta eran nulos, en términos formales. La segunda prueba se aplicó a 40 estudiantes con las mismas características descritas para la primera.

Con las dos pruebas lo que se pretendía evaluar era: i) si los estudiantes establecen o no una relación entre las dos variables; ii) identificar si tenían un sentido de variabilidad; iii) los criterios que utilizan para predecir el valor de la variable respuesta dado un valor para la variable explicativa y; iv) conocer los criterios con que trazan la recta que mejor se ajusta a los datos y con qué precisión lo hacen. En pocas palabras, su nivel de RCI actual.

Del análisis de la primera prueba surgieron algunas hipótesis que se irán mencionando junto con la descripción de cada ítem de esta segunda prueba, la cual se puede ver en https://www.dropbox.com/s/kazphxilhnmisgo/Prueba%20diagn%C3%B3stica%202.pdf?dl=0.

Hipótesis 1

Respecto a si los estudiantes establecían o no una relación o asociación entre las dos variables, se observó que los estudiantes, al parecer, pueden establecer dicha relación sin la necesidad de observar el gráfico. De manera que, tomaremos como primera hipótesis el hecho de que *usar variables sin contexto, o por lo menos un contexto no cercano o no comprensible para los estudiantes los obliga a usar un razonamiento "puro" en sus interpretaciones.*

Por lo anterior, en el ítem 1 de la segunda prueba las variables estaban descontextualizadas, la variable independiente se correspondía con X y la variable dependiente con Y. Queríamos conocer si los estudiantes establecían o no una relación entre las variables X e Y. Lo anterior ubicando un punto en el diagrama de dispersión dado para un valor específico en X, esperábamos que el estudiante identificara el "comportamiento" de los puntos en el diagrama y como consecuencia ubicara el punto que se le pedía para el correspondiente valor de X siguiendo ese comportamiento previamente identificado.

Hipótesis 2

Respecto a la fuerza de la asociación entre las dos variables, se estableció la siguiente hipótesis: *los estudiantes asumen la intensidad de asociación de las dos variables de acuerdo al engrosamiento de la nube de puntos.* Por lo anterior, se diseñó el ítem 2 con cuatro diagramas de dispersión cada uno con diferente coeficiente de correlación.

Hipótesis 3

Otra hipótesis que establecemos, como resultado de las respuestas dadas a las preguntas de predicción en la primera prueba es que *los estudiantes asumen de alguna manera la recta, por lo menos entre vecinos cercanos.* Para ratificar lo anterior y explorar otras maneras de usar esta estrategia, se propuso el ítem 3 (Figura 1, tomado y adaptado de Casey, 2015), donde las variables sí estaban contextualizadas. Queríamos ratificar las estrategias de los estudiantes para predecir un valor. El ítem consta de tres situaciones distintas: con dos vecinos cercanos, uno a cada lado; otra con tres "vecinos cercanos" uno a un lado, y dos al otro lado.

Hipótesis 4

Siguiendo en línea con la hipótesis 3, nuestra cuarta hipótesis es que *la extensión de una concepción local podría conducirnos, en un principio, al ajuste de una recta poligonal, o ajuste parcial, que bien podría "enderezarse" y convertirse en la recta que queremos encontrar, la que mejor se ajusta.* Por lo anterior, en el ítem 4, donde las variables también estaban contextualizadas, partimos preguntando sobre la recta que mejor se ajusta a dos puntos (¿será "trivial" para los estudiantes que esta recta corresponde a la recta que pasa por los dos puntos?) En seguida quisimos indagar sobre cómo trazan la recta que mejor se ajusta para tres puntos no colineales y finalmente se les presenta todo el conjunto de puntos. En últimas lo que queremos conocer con este ítem es qué entienden por "la recta de mejor ajuste".

Hipótesis 5

Un criterio muy usado es que la recta que mejor se ajusta es la que pasa por más puntos. Lo anterior se evidenció en una tarea propuesta por Casey (2015) en la que se daba un diagrama de puntos y los estudiantes ubicaban la recta, representada por un alambre, que mejor se ajustaba para ellos. Con el ítem 5 de esta prueba nos preguntamos qué pasa con "el regreso", es decir suponemos una recta que se ajusta a seis puntos e indagamos sobre cómo deberían ir los puntos. La hipótesis 5 que establecemos acá es que el criterio mencionado anteriormente no va a cambiar, aunque expresado de otra manera: *la nube de puntos que se corresponde con la recta que mejor se ajusta, es la que más puntos toque la recta*.
De manera que en el ítem 5 se mostraban cuatro diagramas (A, B, C y D) con la misma recta pero seis puntos distribuidos de manera diferente en cada gráfico. Se pedía ordenar, de mayor a menor, los gráficos de acuerdo al nivel de ajuste de la recta a los seis puntos.

Análisis y resultados de la segunda prueba

Para el ítem 1, en donde se esperaba que los estudiantes ubicaran el punto respetando la asociación de las dos variables, más de la mitad de los estudiantes así lo hicieron con argumentos como *"Porque conforme asciende el valor de x el valor de y también"* o *"El valor de y que coloqué es de 440 porque creo que todo tiene un orden, entre más grande el número en x más grande el número y"*. Por lo que, parece que resulta sencillo, para la mayoría de los estudiantes, reconocer la asociación entre dos variables cuando la hay pero ¿qué pasa cuando no existe tal asociación? Para un gráfico así solo 2, de 40 estudiantes, explicitan en sus argumentos la no asociación entre las dos variables: *"Pues porque no veo que tenga un orden respectivo"*, *"Lo mismo no está aumentando ni disminuyendo solo se puede colocar donde sea"*. El esfuerzo de los demás radica en encontrar un patrón, una "forma" en la nube de puntos: *"La puse arriba porque creo que esta lleva una secuencia de u"*.

En el ítem 2, relacionando el "engrosamiento" de la nube de puntos con la intensidad o fuerza de asociación de las variables x e y, de los 40 estudiantes, 19 establecieron el orden correcto. Entre los criterios, y con mayor frecuencia, está el de *dispersión o proximidad de los puntos*: *"Yo creo que el mayor es el que los punticos están más juntos y el menor pues el que tiene los puntos dispersos por todo el gráfico"*; el *nivel de rectitud: "Creo que porque la que tenga un nivel recto esta tiene un mayor nivel de intensidad";* como se observa en este último criterios los estudiantes hablan implícita y explícitamente de recta, es decir, que la nube de puntos que se asemeje más a una recta es la que tiene mayor intensidad de asociación, lineal, entre las dos variables.

En cuanto al ítem 3 (Figura 1) en 3a, con dos vecinos cercanos, uno a cada lado, 30 de los 40 estudiantes hicieron una predicción considerablemente acertada.

Salón	Número de personas	Nivel de ruido
A	24	50
B	26	50
C	21	20
D	27	60
E	24	30
F	29	70

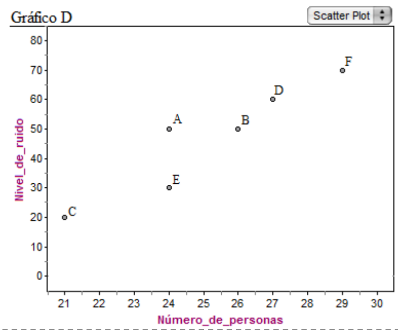

Figura 1. Tabla y diagrama de dispersión relacionado con el ítem 3 de la segunda prueba (Tomado y adaptado de Casey 2015)

Dentro de esos 30, ocho estudiantes tomaron como estrategia la media de los *vecinos cercanos*, por ejemplo: *"El nivel de ruido es 65 el valor lo encontré mirando el D y F y en la mitad de D y F encontré el valor de 65"* o de otros vecinos: *"60 porque 28 está después de 26 y 26 es de 50 y está antes de 29 es de 70 entonces la mitad de 70 y 50 es 60".* Otros, aunque no usaron la media de los vecinos cercanos, sí se guiaron por los mismos para emitir un valor: *"Nivel de ruido sería 61 ya que me guie por el que tenía 27 personas pero porque tenía uno más le subió más la altura"* o *"60 Porque el salón D se acerca a los 28 estudiantes"* o *"Nivel de ruido sería 61 ya que me guie por el que tenía 27 personas pero porque tenía uno más le subió más la altura".*

Un estudiante en su predicción puso dos puntos, entre 50 y 65, manifestando un sentido de la variabilidad: *"Pues entre más personas hay más ruido, pero puede que haya muchas personas y no tanto ruido, varía la respuesta".*

Respecto al ítem 4, entre las estrategias para responder 4a (nube con dos puntos, G y B), la más utilizada fue trazar el segmento que unía los dos puntos, algunos trazaron dos segmentos, uno partiendo del origen y luego el que unía los dos puntos del gráfico, otros trazaron líneas verticales y horizontales hasta cada punto. En total, 7 estudiantes trazaron una recta, donde 3 de ellos trazaron la recta que unía los dos puntos con argumentos como *"… se ajusta porque solo hay dos puntos y es la única recta que se puede trazar".* Otra estrategia fue trazar la recta de manera que pasara por un punto pero muy cerca al otro. 3 estudiantes trazaron la recta entre los dos puntos y partiendo del origen, argumentando de la siguiente manera: *"la pasé sin que tocara G o B".*

Para 4b, 4c y 4d (nubes con más de dos puntos) la estrategia con mayor frecuencia fue la de unir los puntos con segmentos. De los 7 estudiantes que trazaron una recta en el, 2 trazaron varias rectas uniendo dos puntos, mientras que los otros 5 trazaron una sola recta. Con estos 5 estudiantes se identifican algunas de las estrategias reportadas ya en la literatura (Casey, 2015) por ejemplo: la que toque la mayor cantidad de puntos; la que pasa por el "medio" dejando igual número de puntos a cada lado de la recta: *"porque así* *reparto mejor los puntos".* Respecto al inciso 4e, donde se les preguntaba que por qué creían que la recta trazada en 4d (nube con cinco puntos con una asociación positiva) era la que mejor se ajustaba, solo 2 estudiantes argumentaron en términos de "cercanía": *"porque los puntos quedan más cerca a la línea"* o *"porque es como la que más se acerca a los datos".* En general, tal como se esperaba, la mayoría de los estudiantes trazaron "rectas poligonales" al unir los puntos por segmentos.

Finalmente, en el ítem 5, 19 estudiantes ordenaron los cuatro diagramas de dispersión de acuerdo al número de puntos que toca la recta o que están muy próximos a tocarla. Sin embargo de esos 19, solo 11 hablaron en esos términos y 8 hablaron en términos de "alineamiento" u "organización": *"Porque desde A los puntos están más alineados y desde C B y D empiezan a desalinearse".*

11 estudiantes que aunque no eligieron la respuesta correcta, basaron sus argumentos en términos de "cercanía" pero al parecer entendido como el número de puntos más cercanos a la recta, por ejemplo: *"Empezamos con los seis puntos cerca a la línea luego se alejan dos, luego se alejan 3 y luego se alejan 4 quedando B, A, C, D".*

Conclusiones y reflexiones

El análisis de los resultados de estas pruebas nos proporcionó información sobre el Razonamiento Covariacional Informal (RCI) actual de los estudiantes. Se ratificaron algunas estrategias y concepciones ya reportadas en investigaciones anteriores y algunas novedades que nos orientarán para establecer más hipótesis de aprendizaje y, consecuentemente, diseñar el plan para el diseño de actividades computacionales para el aprendizaje de la recta de mejor ajuste de manera informal:

- Construir actividades sin contexto, o por lo menos no comprensible al estudiante, podría evitar que el estudiante tenga una *concepción causal*. En su lugar, se obligaría a tener un razonamiento "puro" en tareas de asociación y predicción.
- La mayoría de los estudiantes, sí reconocen de alguna manera la asociación entre dos variables y cómo varían conjuntamente al observar un diagrama de dispersión, mientras que no reconocen fácilmente la no existencia de una asociación, es decir, buscan sí o sí relacionar de alguna manera las dos variables.
- La mayoría de los estudiantes sí establecen un nivel de asociación en cuanto al grado de dispersión o proximidad de los puntos pero pocos, además de eso, en cuanto a la forma.
- Actividades como el ítem 5, es decir, "el regreso": dada la recta cómo deberían ir los puntos, nos permitirán llegar al término "cercanía". Logrado lo anterior, nos veremos enfrentados a concepciones de tipo geométricas como por ejemplo que la recta debe tocar dos puntos o que la distancia más corta entre un punto y una recta es la perpendicular.

Referencias bibliográficas

Ben-Zvi, D., & Gardfiel, J. (2004). Statistical Literacy, Reasoning, and Thinking: Goals, Definitions, and Challenges. En D. Ben-Zvi, & J. Gardfiel (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 3-15). Dordrecht, The Netherlands: Kluber Academy Publishers.

Casey, S. (2015). Examining Student Conceptions of Covariación: A Focus on the Line of Best Fit. *Journal of Statistics Education* 23 (1), 1-33.

Cobb, P., McClain, K. & Gravemeijer, K. (2003). Learning About Statistical Covariation, *Cognition and Instruction*, 21(1), pp. 1-78.

Engel, J. & Sedlmeier, P. (2011). Correlation and regression in the training of 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,247-258). New York: Springer.

Estepa, A. & Batanero, C. (1995). Concepciones iniciales sobre la asociación estadística. *Revista Enseñanza de las ciencias*, 13(2), 155-170.

Garfield, J. & Ben-Zvi, D. (2008). Learning to Reason about Covaration. In *Developing Students' Statistical Reasoning*, eds. D. Ben-Zvi y J. Garfield, Springer Sciende+Business Media B.V. pp. 289-308.

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.

Zieffler, A, & Garfield, J. (2009). Modeling the growth of students' covariational reasoning during an introductory statistics course. *Statistics Education Research Journal*, 8 (1), 7-31.

Zieffler, A., Garfield, J., Delmas, R. & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 5-19.