EVIDENCIAS DE RAZONAMIENTO COVARIACIONAL DE ESTUDIANTES DE BACHILLERATO EN PROBLEMAS DE CORRELACIÓN Y REGRESIÓN LINEAL

Miguel Medina, Ernesto Sánchez
ingcivi_6@yahoo.com.mx,
esanchez0155@gmail.com
Centro de Investigación y de Estudios Avanzados del IPN
Ciudad de México, México

Resumen

Se informa sobre el razonamiento covariacional que exhiben estudiantes de bachillerato en problemas de correlación y regresión lineal, proponiendo a partir del análisis de las respuestas dos categorías de razonamiento para un problema de Estimación y una para el de Recta de Mejor Ajuste.

Abstract

The covariational reasoning that high school students exhibit in correlation and linear regression problems is reported. Based on the analysis of the data, two categories of reasoning for an estimation problem and one for the Best Fit Line are proposed.

Problema de investigación

Cuando dos variables representan el comportamiento de fenómenos sujetos a ciertos grados de incertidumbre, o bien, representan datos obtenidos experimentalmente cuyo comportamiento futuro tiene grados de incertidumbre, se dice que las variables se encuentran relacionadas o asociadas mediante covariación estadística.

El razonamiento estadístico covariacional de un sujeto consiste en los procesos que le permiten percibir, describir y justificar las relaciones entre variables estadísticas; estos procesos se presentan de manera dual, por un lado ocurren en la mente del sujeto, y por otro, en el discurso hablado o escrito cuando se describen o justifican relaciones entre variables (Moritz, 2004). Por ejemplo Zimmerman (2005) declara que el razonamiento covariacional es la base para realizar inferencias causales inductivas en ciencias, reconociendo sin embargo, que la covariación entre eventos es necesaria, pero no suficiente para inferir una relación causal, por ejemplo. Mientras que, McKenzie y Mikkelsen (2007) consideran que el razonamiento covariacional es una de las más importantes actividades cognitivas que han desarrollado los humanos.

En los programas de estadística de bachillerato, el estudio de la covariación en el que debería desarrollarse el razonamiento covariacional se reduce al cálculo del coeficiente de correlación lineal y a la determinación de la recta de mínimos cuadrados. Sin embargo, la tendencia tradicional en este nivel ha sido enseñar de manera expositiva una serie de procedimientos algorítmicos para resolver ejercicios rutinarios (DGB, 2009). Hace falta diseñar lecciones de enseñanza de manera que su objetivo central sea el desarrollo del razonamiento y no sólo el aprendizaje de técnicas. Por ejemplo, en Medina, Olay y Sánchez (2016) se menciona que entender y resolver los problemas de Estadística no es tarea de simplemente conocer el procedimiento de cálculo, 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. Por otro lado, para diseñar lecciones adecuadas es necesario conocer y entender mejor los razonamientos que ponen en juego los estudiantes de bachillerato cuando enfrentan problemas de regresión y correlación por primera vez y sin enseñanza expositiva explícita, teniendo como base sus conocimientos matemáticos y estadísticos desarrollados previamente. Sabiendo cómo razonan ante este tipo de problemas se pueden proponer diseños de enseñanza con mejores probabilidades para promover el logro de objetivos sobre el tema para este nivel. En esta perspectiva se formula la siguiente pregunta de investigación: ¿Qué patrones de razonamiento covariacional exhiben las respuestas de estudiantes de bachillerato en problemas de estimación y recta de mejor ajuste?

Antecedentes

Ante problemas de correlación los estudiantes de bachillerato manifiestan algunos rasgos característicos: a) Las concepciones previas de los estudiantes sobre la relación entre las variables tienen una gran influencia en sus juicios acerca de la correlación entre estas variables, b) Los estudiantes frecuentemente suponen que existe correlación entre dos eventos que no están correlacionados (correlación ilusoria), c) Tienen dificultad para razonar acerca de la covariación, cuando la relación entre las variables es negativa (Zieffler y Garfield, 2009).

En el caso de los problemas de recta de ajuste, Casey (2014) y Casey y Wasserman (2015) reportan: a) Dado que la determinación informal de la recta de mejor ajuste es un proceso visual, una necesidad de los estudiantes sería convertir los datos recopilados en tablas, en una representación gráfica (diagrama de dispersión), b) Los estudiantes tienen dificultad para observar la tendencia global de un conjunto de datos, cuando leen una gráfica de dispersión, debido a que enfocan su atención en puntos aislados y perciben los datos como una serie de casos individuales, en lugar de considerarlos holísticamente, c) Muchos estudiantes al trazar una recta de ajuste, enfocan su atención en algunos puntos de la gráfica como: el primero, el último, el más alto o más bajo, o en un subconjunto de estos.

Marco conceptual

Este trabajo está basado en tres conceptos fundamentales, el primero es el razonamiento covariacional entendido como las actividades cognitivas incluidas en la coordinación de dos variables cuantitativas mientras se atiende las formas en que cambia una con respecto a la otra (Carlson, Jacobs, Coe, Larsen y Hsu, 2002). El segundo es la noción de agregado, que implica asumir los datos bajo una estructura de conjunto, sin privilegiar características o atributos de datos aislados (Bakker, 2004). El tercer concepto tiene que ver con la idea principal de que la conceptualización que se haga sobre los resultados obtenidos se basa en la información que arrojan los datos mismos, en este sentido el análisis no parte de un marco teórico preestablecido ni rígido, sino en un sistema de códigos y categorías sobre el razonamiento de los estudiantes bajo situaciones específicas de asociación estadística; idea congruente con los principios de la Teoría Fundamentada (Glaser y Strauss, 1967).

Método

Instrumentos y participantes

En la Figura 1 se muestran los dos problemas que se eligieron para este reporte. El primero es un problema de Estimación tomado de Moore (1988) y que se modificó a la presente versión. En el segundo problema, de recta de ajuste, se presenta un diagrama de dispersión de las variables Talla y Peso, en una situación en que están correlacionadas directamente y con una tendencia claramente lineal.

Las actividades de aprendizaje se aplicaron a un grupo de 30 estudiantes de bachillerato (16 a 18 años de edad) que cursaban la asignatura de Estadística y Probabilidad en el Colegio de Ciencias y Humanidades Plantel Vallejo de la UNAM. Las actividades se desarrollaron en cuatro sesiones de 90 minutos cada una, resueltas de manera individual y únicamente con lápiz y papel.

Problema de Estimación

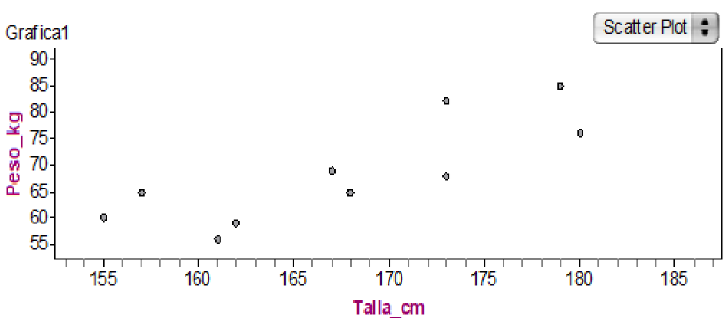
La familia Morales está a punto de instalar paneles solares en su casa para reducir el gasto en la calefacción. Para conocer mejor el ahorro que puede significar la instalación de dichos paneles, los Morales han ido registrando su consumo de gas durante el último año y medio. En la tabla se muestran 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	Jul
Temperatura media (°C)	5.2	-9.8	-5.4	0.2	4.1	11.3	16.3	18.5	18.5
Consumo de gas (m^3)	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 media (°C)	18	15.2	11.8	1.8	0.7	-10.4	1.8	17.1	10.7
Consumo de gas (m^3)	3.4	5.9	8.7	17.9	20.2	30.8	19.3	6.3	10.7

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? Explica tu respuesta:

Problema de Recta de Ajuste

La siguiente gráfica muestra los datos de la talla y el peso de 10 estudiantes de bachillerato. Traza la recta que piensas se ajusta mejor a los datos.



Explica el criterio que utilizaste para trazar la recta:

Figura 1. Problema de Estimación y de Recta de Ajuste

Análisis de datos

El análisis de los datos se realizó mediante algunos principios metodológicos de la Teoría Fundamentada (Glaser y Strauss, 1967), realizando un proceso de codificación (tipos de razonamiento) y categorización de las respuestas, refinándose gradualmente a través de un análisis comparativo constante y una saturación teórica.

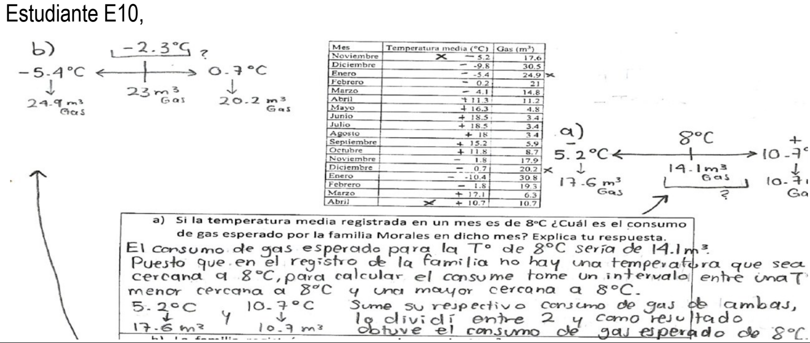
Resultados

Después de analizar las respuestas de los estudiantes se desarrolló un sistema de Códigos (tipos de razonamiento) y Categorías, que representan los patrones del razonamiento covariacional encontrados. Los tipos de razonamiento se encuentran jerarquizados según los rasgos que exhiban, es decir, la identificación de la correlación entre las variables, la utilización de todos los datos disponibles de la relación bivariada o la percepción de la incertidumbre que subyace en los datos.

Así para el problema de estimación, al nivel más alto del razonamiento lo llamamos A1-*Aritmético Interpolación* (4/28 respuestas) que representa las respuestas donde el estudiante toma un intervalo de valores donde se encuentre incluido el dato de la temperatura (8°) por el que se pregunta la estimación, y con sus correspondientes consumos de gas obtiene el promedio para encontrar el valor pedido. Se le considera en el nivel más alto porque considera datos de las dos variables y utiliza un acercamiento estadístico al obtener el promedio aritmético.

Por cuestiones de espacio solo se muestra un ejemplo de respuesta de este tipo de razonamiento en la siguiente figura.

Estudiante E10,



Mes	Temperatura media (°C)	Gas (m^3)
Noviembre	5.2	17.6
Diciembre	-9.8	30.5
Enero	-5.4	24.9
Febrero	0.2	21
Marzo	4.1	14.8
Abril	11.3	11.2
Mayo	16.3	4.8
Junio	18.5	3.4
Julio	18.5	3.4
Agosto	18	3.4
Septiembre	15.2	5.9
Octubre	11.8	8.7
Noviembre	1.8	17.9
Diciembre	0.7	20.2
Enero	-10.4	30.8
Febrero	1.8	19.3
Marzo	17.1	6.3
Abril	10.7	10.7

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? Explica tu respuesta.

Figura 2. Ejemplo de respuesta A1-Aritmético Interpolación

El siguiente tipo de razonamiento lo llamamos A2-*Aritmético Proporcional* (8/28) incluye las respuestas donde los estudiantes buscan un factor de proporcionalidad; eligiendo una pareja de datos (X-Temperatura, Y-Consumo) y con la temperatura dada de 8° forman una regla de tres, asumiendo que entre las variables existe una relación proporcional.

El siguiente es el A3-*Aritmético Siguiendo Un Patrón* (2/28), en este se eligieron las respuestas donde los estudiantes toman como referencia el valor de la temperatura dado (8°) y tratan, a partir de los datos de la tabla, de "completar" este valor mediante algún procedimiento aritmético, y una vez que lo consiguen utilizan esos datos para obtener su respuesta.

En el tipo A4-*Aritmético Sin Patrón Definido* (6/28) los estudiantes utilizaron algunas operaciones básicas (suma, resta, multiplicación y división; en un caso se utiliza la raíz cuadrada), pero sin que se pueda deducir un procedimiento bien definido.

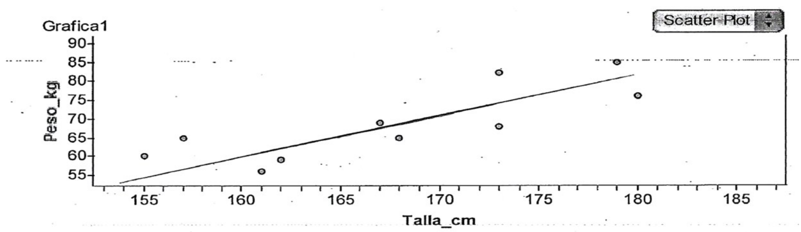
En el A5-*Percepción De La Tendencia* (6/28), el estudiante no realiza operación alguna y sólo enfoca su atención en los datos de la tabla. Sus respuestas se basan en un análisis visual de la tendencia de los datos, en particular en el sentido de su comportamiento.

Finalmente, el tipo de razonamiento A6-*Sin Argumento* (2/28) representa aquellas respuestas donde el estudiante solo aporta el resultado, sin argumentar o hacer explícito su procedimiento.

Una vez que se hace una comparativa entre los tipos de razonamiento y se identifican rasgos comunes se definen las Categorías, del A1 al A4 en una categoría denominada *Búsqueda de un Patrón Aritmético*. Esta categoría representa los razonamientos donde el estudiante intenta descifrar una especie de clave o pista para estimar el valor que se le pide. Realiza una búsqueda en los datos que se le proporcionaron en la tabla, sin embargo, no utiliza la totalidad de estos, pareciera que el alumno sospecha que existe un patrón o estructura ocultos en un subgrupo de datos y que lo llevarán a la respuesta. Este subgrupo de datos pueden ser los más cercanos al valor dado, un intervalo de valores que lo contengan o el llegar a conformar este dato mediante alguna operación aritmética para finalmente utilizar los correspondientes valores de la variable respuesta para obtener la estimación pedida. Por otro lado, A5 y A6 se integran en la categoría *No Procedimental*, ya que no realizan cálculo alguno y solo observan la tendencia de los datos en la tabla o bien, responden sin argumentar.

Para el problema de trazar la recta de mejor ajuste se definieron dos tipos de razonamiento característicos, el B1-*Partición* (6/21 respuestas) donde se clasificaron las respuestas en las que el estudiante traza la recta o hace referencia a que su posición debe ser tal que pase por el medio de la nube de puntos, siguiendo la dirección de ésta, es decir trazada de forma diagonal, dejando de un lado y del otro de la recta, el mismo número de puntos, como lo muestra la Figura 3.

Estudiante E14,



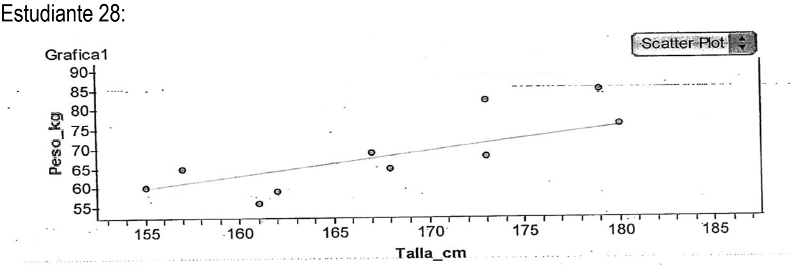
2. Explica el criterio que utilizaste para trazar la recta:

La recta pasa por en medio de todas los puntos demostrando que es como la media.

Figura 3. Ejemplo de respuesta B1-Partición

En B2-*Pertenencia* (15/21) se incluyen las respuestas que muestran dos tipos de comportamiento, por un lado se encuentran aquellas donde el estudiante traza o refiere que la recta debe pasar por el mayor número posible de puntos o por la totalidad de estos y también están las respuestas donde el estudiante traza la recta cuidando que pase a través puntos específicos de la nube, como en la respuesta mostrada en la Figura 4, donde se argumenta que debe pasar por dos puntos (el más bajo y más alto).

Estudiante 28:



2. Explica el criterio que utilizaste para trazar la recta:

Por que va del valor mas bajo al alto y pasa o se acerca a todos

Figura 4. Ejemplo de respuesta B2-Pertenencia

En este caso se define la categoría *Sentido Geométrico*, que representa los razonamientos donde el estudiante ajusta la recta de tal forma que se comporte como un modelo que tenga sentido para él, posiblemente el de la función lineal estudiada en sus cursos de geometría analítica, por esto quiere forzar que la recta pase por la mayor cantidad de puntos o por algunos específicos (dos en algunos casos o uno, el que aparenta ser el centro de la nube) e ignora el resto del conjunto. Además, la dispersión presente en el conjunto de datos lo imposibilita para ocuparlos todos y solo utiliza los que aparentan estar alineados. De igual forma, al no tener claro qué hacer con los puntos “que salen” del patrón que conoce, pareciera que elige como alternativa el trazar la recta por en medio de la nube de puntos, buscando nuevamente un procedimiento que le sea familiar, una especie de mediana. La Tabla1 muestra el resumen de los tipos de razonamiento y las categorías que surgen de los datos.

Tabla1. Tipos de Razonamiento y Categorías

Tipo de Razonamiento	Categoría
A1-Aritmético Interpolación	Búsqueda de Patrón Aritmético
A2-Aritmético Proporcional	
A3-Aritmético Siguiendo un Patrón	
A4-Aritmético Sin Seguir un Patrón	
A5-Percepción de la Tendencia	No Procedimental
A6-Sin Argumento	
B1-Partición	Sentido Geométrico
B2-Pertenencia	

Conclusiones

Concluimos destacando las evidencias del razonamiento covariacional que llevaron a definir las categorías *Búsqueda de Patrón Aritmético* (problema de Estimación) y *Sentido Geométrico* (problema de Recta de Ajuste) donde se percibe que los estudiantes tienden a elegir estrategias procedimentales con las que se encuentran familiarizados y que les permiten hacer a un lado o bien ignorar características de los datos que no se ajustan con los patrones o modelos que tienen sentido para ellos. En el caso del problema de estimación, buscan factores de proporcionalidad, hacen uso de la regla de tres o intentan completar el número que se les proporciona, y para el problema de la recta de ajuste, alinean la mayor cantidad de puntos o eligen algunos puntos representativos y dividen la nube de puntos en dos subgrupos. También conviene destacar rasgos de razonamiento que están ausentes en las estrategias de los estudiantes, por ejemplo, que el trazo de un diagrama de dispersión no surge en ningún estudiante (la idea de que graficar los datos en un plano podría ayudar a determinar la solución no se presentó) o que se percibe un desinterés por el contexto, nadie hace alguna consideración acerca de éste, quizá creyendo que es sólo un adorno, pensando que lo importante es el procedimiento aritmético.

Con respecto a la actividad de ajustar la recta vemos que existe una ausencia del criterio de cercanía de los puntos a la recta, parece que esta idea es muy compleja dado que su conceptualización requiere definir el concepto de distancia de una recta a un conjunto de puntos, además de que no suelen considerar todos los datos; al igual que en el problema de estimación, una buena parte no concibe un procedimiento que utilice la totalidad de los puntos, además los estudiantes no consideran la variación simultánea de las dos variables involucradas, parece que fijan su atención en solo una de estas variables, generalmente en un par de puntos representativos para ellos y a partir de esos puntos realizan su estimación. La ausencia de este rasgo del razonamiento covariacional es relevante, ya que la habilidad de razonar variacional o covariacionalmente favorece la concepción y uso de las relaciones funcionales (Thompson y Carlson, 2018).

Este comportamiento puede tener su origen en una dificultad para concebir al conjunto de datos como un *agregado*, es decir, como un sistema en el que están ligados unos a otros y tienen la propiedad de ser desviaciones de un mismo modelo. En consecuencia, consideramos que el conocimiento de las tendencias de razonamiento de los estudiantes que hemos descrito, sugieren buscar diseños instruccionales que propicien la problematización de las ideas de trazar diagramas de dispersión, de utilizar todos los datos, de reflexionar sobre la relación entre lo estadísticos y el contexto, además de vincular el modelo de regresión lineal en los problemas de estimación.

Referencias

Bakker, A. (2004). Reasoning about Shape as a Pattern in Variability. *Statistics Education Research Journal*, *3*(2), 64 – 83.

Carlson, M., Jacobs, S., Coe, E., Larsen, S., & Hsu, E. (2002). Applying covariational reasoning while modeling dynamic events. *Journal for Research in Mathematics Education, 33*(5), 352-378.

Casey, S. (2014). Teachers´ knowledge of students´ conceptions and their development when learning linear regression. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics. ICOTS9 Invited Paper-Refereed*. Voorburg, The Netherlands: International Statistical Institute

Casey, S., & Wasserman, N. (2015), “Teachers’ Knowledge about Informal Line of Best Fit.” *Statistics Education Research Journal*, *14*(1), 8-30.

DGB (2009). *Programa de estudios de matemáticas*. SEMS.

Glaser, B. G. y Strauss, A. L. (1967/2008). *The discovery of Grounded Theory: Strategies for qualitative research*. New Brunswick, USA: Aldine Transaction.

McKenzie, C. R. M., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology, 54*, 33–61.

Medina, M. N., Olay, G. E., y Sánchez, E. (2016). Razonamiento de estudiantes de bachillerato frente a un problema de regresión [Reasoning of high school students in a regression problem]. *Eutopia. Revista del Colegio de Ciencias y Humanidades para el Bachillerato, 9*(24), 71-78.

Moore, D. S. (1988). Should mathematicians teach statistics? *The College Mathematics Journal*, *19*(1), 3–7.

Mortiz, J. B. (2004). Reasoning about covaration. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing statical literacy, reasoning and thinking (pp. 227-256).* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Thompson, P. W. y Carlson, P. M. (2018). Variation, Covariation, and Functions: Foundational Ways of Thinking Mathematically. In J. Cai (Ed.), *Compendium for Research in Mathematics Education* (pp. 421 - 456). Reston, Virginia: National Council of Teachers of Mathematics.

Zieffler, A. S., & Garfield, J. (2009). Modeling the Growth of Students’ Covariational Reasoning during an introductory statistics course. *Statistics Education Research Journal, 8*(1), 7-3.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27*, 172-223.