

COMPRENSIÓN DE LAS IDEAS DE COVARIANCIA, CORRELACIÓN Y REGRESIÓN EN ESTUDIANTES DE NIVEL SUPERIOR

Ignacio Delgado Escobar; Ana María Ojeda Salazar
U.A.E.M., Morelos. (México). DME, Cinvestav, I.P.N. (México)

nachock@hotmail.com ; amojeda@yahoo.com

Campo de investigación: pensamiento relacionado con probabilidad, estadística. Nivel educativo: superior

Resumen

La presente investigación estudió el papel de tres estrategias de enseñanza —tradicional, basada en el análisis exploratorio de datos, y con uso de un programa de cómputo— en las respuestas a preguntas sobre covariancia, correlación y regresión a la media, planteadas a estudiantes de nivel superior. El análisis consideró los registros semióticos (Duval, 1998) utilizados y se refirió a los tipos de conocimiento (Pollatsek et al, 1981) identificables en las respuestas de los estudiantes, aportadas en cuestionarios aplicados después de la enseñanza. El estudio, cualitativo, tuvo tres fases: documental, exploratoria y de investigación. Respecto a los conceptos citados, el enfoque de análisis exploratorio de datos obtuvo los mejores resultados, con evidencia de conocimientos funcional y analógico; luego, el uso del programa de cómputo resultó en conocimiento funcional, y el tradicional en el de cálculo.

Introducción

La poca importancia que se otorga al desarrollo conceptual de la Probabilidad y Estadística en todos los niveles del Sistema Educativo Mexicano, orienta el proceso de enseñanza de estas ramas a la de la operación de fórmulas para calcular algún parámetro estadístico y omite la interpretación. Las dificultades de comprensión de sus ideas fundamentales se revela con la manifestación de razonamientos incorrectos frente a situaciones aleatorias, particularmente con errores que se cometen en el medio social, debidos al uso de algunas heurísticas que desatienden normas de la lógica y de la suficiencia de los datos para inferencias (por ejemplo, Kahneman et al., 1982).

La comprensión de los conceptos de covariancia, correlación lineal y regresión es de interés por varios motivos. El primero es que los conceptos son relevantes en el currículum de Estadística desde el nivel medio superior para comprender otros métodos tales como análisis de varianza y análisis multivariado.

En este foro presentamos algunos resultados de un proyecto de investigación sobre la comprensión por los estudiantes de nivel superior y maestría de las ideas de covariancia, correlación y regresión a la media.

Elementos teóricos

El estudio de conceptos y técnicas de estadística incluye al tratamiento de datos y a las reglas y procedimientos de cálculo, pero a menudo los profesores que imparten esta materia no ponen énfasis en la diferenciación entre los procedimientos y reglas de cálculo y la significación de los datos, lo que provoca una confusión en el estudiante que redundará en la operación de fórmulas sin comprensión del concepto que les subyace (Pollatsek et al, 1981). Por otra parte, Batanero et al (1996) argumenta que el concepto de asociación y causalidad es de relevancia para la educación matemática, porque es fundamental para muchos métodos estadísticos que permiten modelar numerosos fenómenos en diferentes ciencias.

El fin principal en muchas de estas aplicaciones es encontrar explicaciones causales que nos ayuden a comprender nuestro entorno. Sin embargo, la asociación no implica necesariamente una relación causal. A veces existe "correlación espuria" entre variables debida a la influencia de factores concurrentes sin que haya un vínculo causal. Aparte de esta dificultad epistemológica, investigadores en la rama psicológica, como Kahneman et al (1982), han mostrado que la habilidad para emitir juicios de asociación no se desarrolla intuitivamente, que las personas adultas a veces basan sus juicios en creencias previas sobre el tipo de asociación que debería existir entre las variables, en lugar de en las contingencias empíricas presentadas en los datos, y finalmente que la existencia de tales preconcepciones en las situaciones aplicadas es otra de las dificultades de la enseñanza de la asociación.

Desde un punto de vista más general, Frawley (1999) asevera que la ciencia cognitiva argumenta abierta y descaradamente sobre representaciones mentales internas, de cálculos y de otras operaciones realizadas con base en estas representaciones internas (pág. 38).

La comprensión es un término que connota la síntesis, la habilidad para hacernos de una representación concreta y analógica. Es el modo fundamental del conocimiento para toda situación que implique subjetividad y afectividad; es el conocimiento que ejerce todo aquel que se percibe como individuo, o sea como sujeto. Se ha definido la comprensión como un proceso en que la representación de cada cosa como un pensamiento se asemeja a algo que ya conocemos. Siempre que el funcionamiento interno de una nueva cosa es bastante extraño o complicado, representamos sus partes en términos de signos más familiares. De este modo, hacemos que la novedad parezca similar a algo más común (Minsky, 1986, pág. 57; citado en Esté, 1997, págs. 189 y 190); así los sujetos pueden utilizar y aplicar los conceptos en situaciones más complicadas.

Pollatsek y sus colaboradores (1981), luego de un estudio realizado respecto a la comprensión de la media, han propuesto tres tipos de conocimiento relacionados con la comprensión de un concepto, a saber, el conocimiento de cálculo, el funcional y el analógico. El primero concierne a la identificación en la expresión simbólica de un concepto de los elementos y operaciones implicados y al desarrollo de algoritmos; el segundo atañe a la advertencia de los elementos implicados en un concepto, sus relaciones y su función como unidad; el tercero interpreta la función de un concepto en diferentes contextos.

En tanto el despliegue de la actividad matemática requiere de un soporte semiótico, Duval (1998) ha caracterizado ese soporte para que constituya un registro semiótico de representación, con el señalamiento de tres actividades cognitivas de las que debe ser expresión: su producción, el tratamiento al interior de un registro dado, y la conversión de un tipo de registro a otro.

Erickson y Nosanchuk han señalado, en el primer capítulo de su obra (1977), las ventajas de la exploración de datos, *leyéndolos* mediante el uso de operaciones aritméticas básicas, de las cuales la gran arma es la substracción, y el uso de gráficas (pág. 2); se realiza esta exploración para *dar sentido* a los datos en mano antes de proceder a los cálculos o al uso de programas de cómputo, lo cual ya tendrá razón de ser.

Una de las ventajas del uso de la hoja electrónica de cálculo es la identificación de los elementos en la expresión simbólica de una entidad conceptual (Landín, 1996). Por otro lado, la presentación de gráficas con un programa de cómputo tal sugiere el uso de éste en la enseñanza luego de la propuesta de Duval citada.

Preguntas de investigación

¿Qué papel juega el tipo de enseñanza —con el uso de la hoja electrónica de cálculo, la tradicional y la basada en el análisis exploratorio de datos— en la comprensión de la covariancia, la correlación y la regresión a la media?

¿Qué dificultades manifiestan los estudiantes en tareas de producción, tratamiento y conversión de registros semióticos de representación, referidas a la covariancia, a la correlación y a la regresión a la media?

¿Qué dificultades presentan los estudiantes con referencia a los tipos de conocimiento funcional, analógico y de cómputo de la covariancia, la correlación y la regresión a la media luego de su enseñanza?

Los objetivos perseguidos fueron:

Identificar bondades y desventajas de tres enfoques de enseñanza de la covariancia, la correlación y la regresión a la media para su comprensión por los estudiantes.

Identificar los tipos de conocimiento que los estudiantes manifiestan de la covariancia, correlación y regresión a la media luego de su enseñanza.

Indagar sobre el recurso a, y entre, diferentes registros semióticos en referencia a la covariancia, correlación y regresión a la media, y sobre métodos y estrategias que los estudiantes emplean ante situaciones dadas que atañen a esos conceptos.

Métodos y procedimientos

Las interrogantes de la investigación se estudiaron, de manera consecuente, en tres escenarios de enseñanza de la estadística. El escenario tradicional, es decir, el correspondiente a la enseñanza desarrollada por el docente como expositor, con el uso del pizarrón para plantear los elementos teóricos y la resolución de problemas a modo de ejemplos, con la propuesta a los estudiantes de otros adicionales a manera de ejercicios con el uso de la calculadora. El escenario con el recurso a la hoja electrónica de cálculo, que tiene lugar en el aula de cómputo. El escenario con enfoque en el análisis exploratorio de datos (Erickson et al, 1977).

Dada la importancia para el estudio de los procedimientos efectuados por los estudiantes, más allá de la consideración solamente de la corrección de las respuestas que otorgaron, el enfoque de la investigación propuesta fue cualitativo. El diseño de los instrumentos de recolección de datos y el tipo de análisis de ellos, por tanto, respondieron a la búsqueda de cualidades en los procedimientos efectuados para contestar cuestiones de los conceptos en mira principalmente, aunque algunos aspectos se atendieron en lo cuantitativo. Los instrumentos de investigación fueron guiones de clase, cuestionarios a contestar en lápiz y papel y guiones de entrevistas individuales audiograbadas semiestructuradas. Más específicamente, de las respuestas en cuestionarios y entrevistas identificamos tipos de conocimiento funcional, analógico y de cómputo de los conceptos de interés y las dificultades que manifestaron con la formación, tratamiento y conversión de un registro semiótico de representación a otro.

Merriam S. (1998) señaló, sobre los métodos cualitativos que, primero, la clave consiste en entender el fenómeno de interés desde las perspectivas de los participantes y no del investigador. Segundo, el investigador es el instrumento primario para la colección de datos y análisis. Tercero, la investigación cualitativa se refiere al estudio de casos. Cuarto, emplea la estrategia de investigación inductiva. Quinto, se enfoca en los procesos, significados y entendimiento; el producto del estudio cualitativo se enriquece descriptivamente.

Organización de la investigación y criterios de análisis

El estudio comprendió tres fases. En la primera se efectuó una revisión y análisis de contenidos temáticos de cursos de probabilidad y estadística en cursos de la Universidad Autónoma del Estado de Morelos y el Instituto de Estudios Superiores de Monterrey “campus” Morelos, así como de los principales textos recomendados en éstos. La segunda fue de exploración y la tercera de desarrollo de la investigación. Para la de exploración se diseñaron dos cuestionarios con cuatro y doce ítems, que fueron respondidos por tres grupos de estudiantes, veintiuno de Maestría y diecisiete de licenciatura, sin realizar entrevista alguna.

La fase de investigación se dividió en dos etapas: 1) Curso como escenario de investigación, cuestionarios con nueve y once ítems, aplicados a tres grupos de veintidós, dieciocho y dieciséis estudiantes de maestría y licenciatura, respectivamente; y 2) tres cuestionarios con once, diez y diez ítems, aplicados a cuatro grupos con trece, veintitrés, treinta y trece estudiantes, una entrevista video grabada y doce entrevistas audio grabadas.

La obtención de los datos se llevó a cabo en el aula para las enseñanzas tradicional y con análisis exploratorio de datos; y en el aula de cómputo, para la enseñanza mediada por la hoja electrónica de cálculo. Las clases fueron impartidas en seis sesiones de dos horas en un curso completo de estadística, por el presente investigador.

Para el análisis de las respuestas interesó su corrección o incorrección y tipos de errores, y en la revisión de la justificación se consideró la interpretación del enunciado hecha por el respondiente, el tratamiento a registros de representación, y los tipos de razonamiento (heurísticas, relaciones de tipo causal, entre otros) expresados. Por limitaciones de espacio nos referiremos sólo a estos dos últimos aspectos.

Análisis y resultados

Con referencia a la fase de investigación, en este trabajo solamente se informa del análisis de tres ítems, uno por cada escenario de enseñanza.

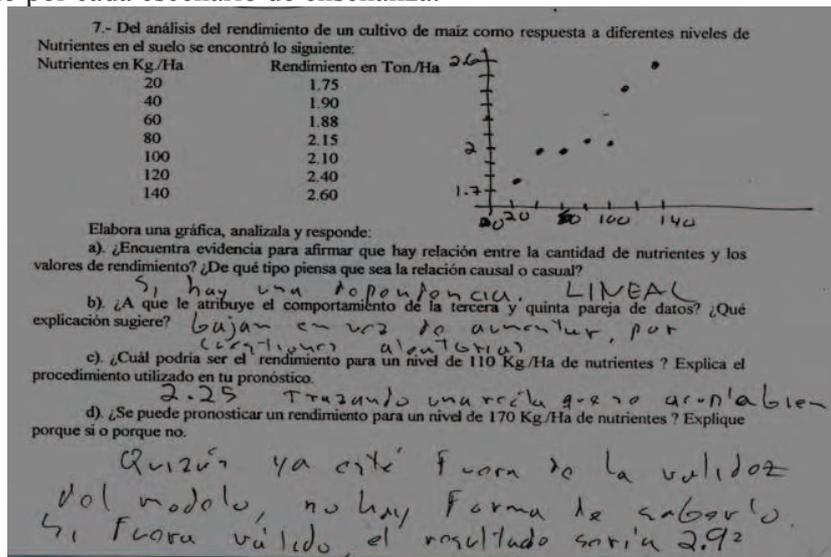


Figura 1. Respuesta de estudiante que utilizó análisis exploratorio (ítem 8).

En este ítem el estudiante puso en juego el tratamiento del registro tabular, al poner en relación los datos de una variable con la otra, luego la conversión al diagrama de dispersión y también exhibió tratamiento de éste; mediante la lengua natural evidenció identificación de las variables, de la escala, la variación aleatoria y la asociación (dispersión y pendiente) (Duval, 1998). Esta información indica que el sujeto posee un conocimiento funcional, por lo menos, acerca de la correlación y regresión lineal, ya que puso en relación el concepto en una situación específica. El conocimiento de cálculo lo exhibió en el tratamiento gráfico.

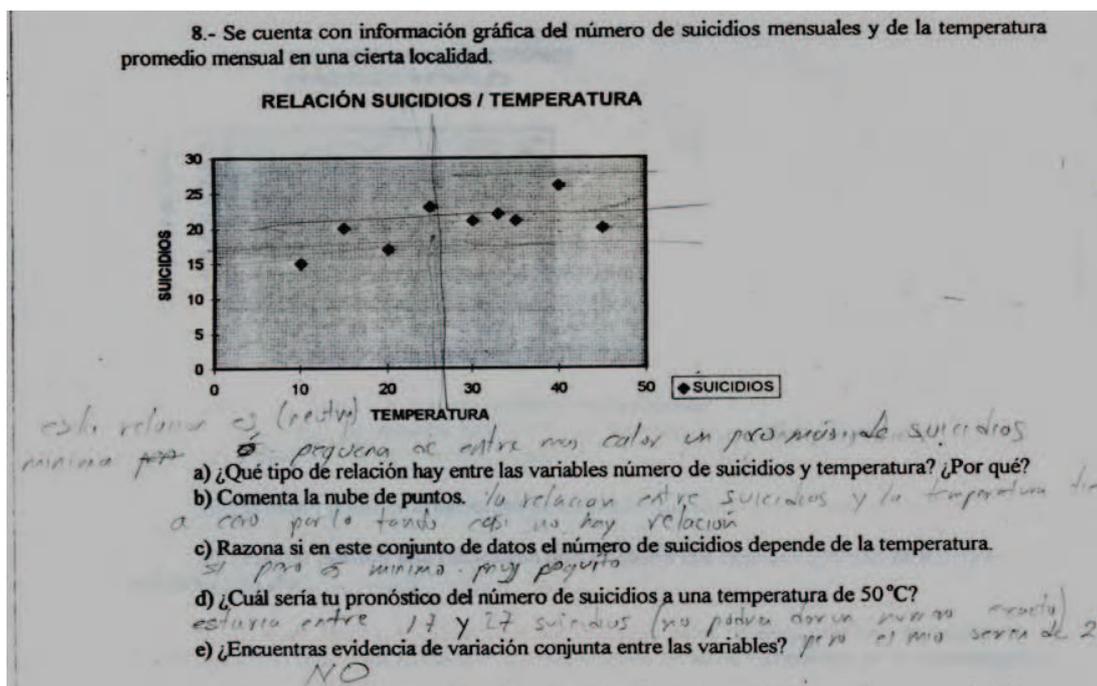


Figura 2. Respuesta de estudiante que utilizó la hoja electrónica de cálculo (ítem 11).

El estudiante, cuyas respuestas proporcionó en lengua natural principalmente, ubicó de manera aproximada el centro de gravedad de la nube de puntos en el diagrama de dispersión para responder las preguntas; por otro lado, efectuó la conversión entre el registro gráfico y la lengua natural. De acuerdo con Pollatsek et al (1981) evidenció el conocimiento de cómputo del diagrama de dispersión y el conocimiento funcional, ya que, identificó las variables dependiente e independiente en el diagrama de dispersión, y advirtió el grado de asociación mediante la estimación de la pendiente y la dispersión de los puntos a una recta.

En lo que concierne a la enseñanza tradicional, ésta resultó principalmente en conocimiento de cálculo de los conceptos de interés. La figura 3 corresponde a la respuesta de un estudiante a un problema referido a la media; evidenció en su contestación conocimiento de cómputo y funcional de la media en la solución de acuerdo al contexto del problema.

Por otro lado, no manifestó el uso de la heurística de representatividad, de acuerdo con (Kahneman et al., 1982).

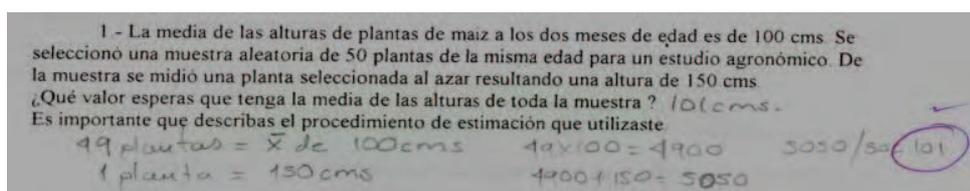


Figura 3. Respuesta de estudiante con una enseñanza tradicional (ítem 16).

Conclusiones

Se puede afirmar que, respecto a registros semióticos de representación, los estudiantes con enseñanza de análisis exploratorio (Erickson, 1977) tuvieron el mejor desempeño: el 70% con la formación, tratamiento y conversión de las representaciones semióticas (Duval, 1998); les siguieron los estudiantes cuya enseñanza recurrió al uso de la hoja electrónica de cálculo, quienes únicamente ejecutaron tratamiento y conversión de registros semióticos. Los estudiantes con enseñanza tradicional, en menor proporción realizaron el tratamiento y la conversión de registros.

En virtud de esos resultados, los estudiantes que utilizaron el procedimiento de análisis exploratorio dieron evidencia de conocimiento analógico, funcional y de cómputo en dicho orden, mientras que los que utilizaron la hoja electrónica de cálculo, sólo evidenciaron tener conocimientos funcional, analógico y de cómputo en dicho orden; finalmente en los estudiantes con enseñanza tradicional se advirtió en mayor medida conocimiento de cómputo. Los estudiantes cuya enseñanza recurrió al uso de la hoja electrónica de cálculo entendieron mejor el carácter funcional o el significado de la regresión a la media en relación con sucesos naturales; mientras que los estudiantes con enseñanza de análisis exploratorio sólo aplicaron la herramienta para la solución de problemas.

Referencias bibliográficas

- Batanero, C.; Estepa, A.; Godino, J.; Green, D. (1996). Intuitive Strategies and Preconceptions about Association in Contingency Tables. *Journal for Research in Mathematics Education*. Vol. 27. No 2, 151-169.
- Duval R. (1998). Registro de representación semiótica y funcionamiento cognitivo del pensamiento. En Hitt, F. (Ed.) *Investigaciones en Matemática Educativa II* México: Grupo Editorial Iberoamérica, (pp. 173-201).
- Erickson, B.; Nusanchuk, T. (1977). *Understanding Data*. Open University Press. GB.
- Esté, A. (1997). *Cultura replicante: el orden semiocentrista*. Ciencias cognitivas/Semiótica. Barcelona: Editorial Gedisa. (pp. 189-190).
- Frawley, W. (1999). *Vigotsky y la Ciencia Cognitiva. Cognición y Desarrollo Humano*. Vol. 36. Paidós.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment Under uncertainty: Heuristics and Biases*. Cambridge University Press.
- Landín, P. (1996). *Variable Aleatoria y Distribución Binomial: Epistemología, Estrategias Didácticas y Desempeños de Estudiantes Bachilleres*. Tesis de maestría sin publicar. ICE, Universidad Autónoma del Estado de Morelos, PNFAPM.
- Merriam, S. B. (1998). *Qualitative Research and case Study. Applications in Education*. Jossey-Bass Publishers, San Francisco.
- Pollatsek, A.; Lima, S.; Well, D. (1981). Concept or Computation: Students' Understanding of the Mean. *Educational Studies in Mathematics* 12, 191-204. Reidel, Holanda.