

Propuesta de enseñanza de estadística descriptiva basada en la Paradoja de Simpson

M.Cecilia De Cortazar¹, Anyelen Di Paolantonio¹, María Valeria Calandra^{1,2}

¹UIDET GAMEFI, Departamento de Ciencias Básicas, Facultad de Ingeniería, Universidad Nacional de La Plata, calle 116 y 48, La Plata (1900).

²maria.calandra@ing.unlp.edu.ar

Resumen

Se presenta una secuencia didáctica aplicada en una materia optativa de Sexto Año del Colegio Nacional Rafael Hernández dependiente de la UNLP llamada “Big Data”, en la cual se desarrollan temas de Estadística. Se trata de una propuesta de enseñanza en la que se aborda el estudio y análisis de datos, sus lecturas e interpretaciones utilizando como recurso didáctico la Paradoja de Simpson con el soporte tecnológico del software R, frecuentemente utilizado en la enseñanza universitaria y en investigación.

La intencionalidad docente es potenciar las habilidades de pensamiento analítico y crítico de los estudiantes al tomar decisiones basadas en el análisis estadístico que a menudo se ven profundamente afectados por las interpretaciones erróneas de los datos así como también por la omisión de variables que pueden resultar significativas según la situación analizada.

El pensamiento crítico se caracteriza por ser, según Ennis (1985): “un pensamiento racional y reflexivo interesado en qué hacer o creer acerca de algo” (p.44). Para desarrollar este tipo de pensamiento, coincidimos con Contreras et al. (2012) en que “Una herramienta didáctica posible es utilizar algunas de las paradojas clásicas para crear situaciones didácticas que sirvan para enfrentar a los estudiantes con sus intuiciones incorrectas y hacerlas evolucionar en forma positiva.” (p.1).

Para su desarrollo se decidió abordar un caso de estudio propuesto por Taylor y Mickel (2014) en el cual se utilizó un conjunto de datos reales del Departamento de Servicios de Desarrollo del estado de California responsable de asignar fondos que apoyan a más de 250.000 residentes con discapacidades del desarrollo (denominados “consumidores”). La base de datos consta de 1.000 de estos consumidores con la siguiente información: número de seguro social, edad, edad dividida en cohortes, género, origen; y los datos del gasto que le insume al Estado cada consumidor, es decir,

el monto anual en dólares que el mismo gasta en cada consumidor para apoyar a estas personas y sus familias. De un primer estudio se desprendería una clara discriminación por parte del estado de California de los ciudadanos de acuerdo a su origen; sin embargo, un análisis más profundo revela que no existiría discriminación y que se habría producido un efecto de la Paradoja de Simpson. La paradoja de Simpson (o efecto Yule-Simpson) es una paradoja en la cual una tendencia, que aparece en varios grupos de datos, desaparece cuando estos grupos se combinan con otra variable y en su lugar aparece la tendencia contraria.

En la escuela media, la enseñanza y el aprendizaje de estadística generalmente han sido abordada mediante secuencias que proponen la realización de cálculos manuales o utilizando calculadoras científicas, Excel o GeoGebra. Consideramos que las Tecnologías de la Información y la Comunicación juegan un papel fundamental para realizar un análisis más profundo de las variables estadísticas, disminuyendo el tiempo de proceso de la información, pudiendo centrar el trabajo en la lectura e interpretación de la misma. Para ello, se trabajó con el Software R de uso libre específico para estadística, a través de la aplicación para dispositivos móviles (Android e iOS) y/o web (www.rcommand.com): Run R Script - Online Statistical Data Analysis -. Esta aplicación genera una cuenta en la nube para cada usuario permitiendo ser ejecutados sin poseer una instalación del programa.

La secuencia estuvo dividida en etapas planificadas de manera que se les permitiera a los estudiantes asumir el rol de “investigadores” (demandando 3 clases de 3hs cátedra), formulando hipótesis, analizando los datos y obteniendo conclusiones que pudieran ser modificadas a través del recorrido por la misma. Una de las primeras tareas tuvo la intención de que los estudiantes se familiaricen con la aplicación y el manejo del lenguaje R. Se les propuso el siguiente cuestionario guía: ¿cuál es el gasto medio anual que le demanda al estado cada grupo de consumidores de acuerdo a su origen? Si realiza un diagrama de barras con esas variables, ¿cuáles comunidades representan mayores gastos? ¿Consideran que hay algunas comunidades que son favorecidas? Si realizan un gráfico de sectores para de la cantidad de consumidores de acuerdo a su origen, ¿cuál tiene más frecuencia absoluta? En una segunda etapa realizaron un gráfico que permitió valorar los gastos medios del estado de acuerdo al género de los consumidores. En una tercera etapa analizaron los gastos para cada comunidad en función de la edad y llegaron a una paradoja, ya que en un primer análisis del promedio de gastos los consumidores sudamericanos eran significativamente inferiores a los de los

norteamericanos, sin embargo, en todas menos una cohorte de edad (18-21) el promedio de los gastos de los consumidores sudamericanos fueron mayores que los de los norteamericanos.

Esta propuesta permitió el uso del software R en un aula en una escuela de enseñanza media, sin necesitar computadoras. El uso de la aplicación móvil hizo posible profundizar en los métodos estadísticos abordados en clase. Posibilitó influir de un modo positivo en la formación estadística de nuestros estudiantes mediante la inclusión en las clases de casos prácticos derivados de problemas reales. La experiencia originó una discusión acerca de la Paradoja de Simpson y las implicancias de no realizar un análisis estadístico riguroso de los datos teniendo en cuenta todas las variables y sus interacciones. Se mostró la importancia del aprendizaje de estadística para formar ciudadanos analíticos y críticos. Los estudiantes se familiarizaron con el uso de la aplicación durante el desarrollo de la secuencia.

Palabras clave: estadística; paradoja de Simpson; software R.

Referencias bibliográficas

- Contreras, J. M., Batanero, C., Cañadas, G. R. y Gea, M. M. (2012). La paradoja de Simpson. *Suma*, 71, 27-34.
- Ennis, R. H. (2005). Pensamiento crítico: un punto de vista racional. *Revista de psicología y educación*, 1(1), 47-64.
- Maronna, R. A. (2021). *Probabilidades y Estadística Elementales para Ciencias e Ingeniería*. La Plata: Editorial Exacta.
- Taylor, S. A. y Mickel, A. E. (2014). Simpson's Paradox: A Data Set and Discrimination Case Study Exercise. *Journal of Statistics Education*, 22(1), Recuperado de: www.amstat.org/publications/jse/v22n1/mickel.pdf.
- Simpson, E. H. (1951). The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society (Series B)*, 13, 238–241.