

Paquete ZOIP de R para modelo de regresión mixto con datos proporcionales inflados con ceros y/o unos

Juan Camilo Díaz Zapata¹, Freddy Hernández Barajas²

¹ Estudiante de maestría en estadística, Universidad Nacional de Colombia, Sede Medellín.

² Profesor asistente, Universidad Nacional de Colombia, Sede Medellín.

Resumen El modelo de regresión mixto para datos proporcionales inflados con ceros y/o unos, es un modelo de regresión donde las covariables se encuentran dadas por efectos fijos y aleatorios; y donde la variable respuesta se encuentra definida a partir de una distribución para datos proporcionales, tal como la distribución beta o la distribución simplex. Los datos proporcionales pueden ser obtenidos a partir de variables medidas en porcentajes y razones, donde su resultado por lo general se encuentra ubicado en el intervalo entre cero y uno, sin embargo, en ocasiones los datos proporcionales pueden dar resultados en cero o en uno, representando la ausencia o presencia total de cierta característica medida. Diferentes autores han trabajado en el desarrollo de diferentes modelos de regresión y metodologías de estimación para este tipo de datos, sin embargo, no se ha desarrollado un modelo de regresión mixto para datos proporcionales inflados con ceros y/o unos, que reúna los principales modelos de regresión de este tipo y que la estimación de los parámetros sea vía máxima verosimilitud y la cuadratura de Gauss-Hermite. En este trabajo se presenta el paquete **ZOIP** del sistema computacional R, alojado en el CRAN de R, en él se implementa la distribución ZOIP (Zeros Ones Inflated Proporcional), que se compone de la mezcla de tres distribuciones, dos de ellas discretas y una distribución continua que además es una función de densidad de probabilidad para datos proporcionales, también es implementado el modelo de regresión para efectos fijos y mixtos, ZOIP, que reúne diferentes distribuciones y modelos de regresión de efectos fijos y mixtos para datos proporcionales inflados en cero y/o uno, este modelo es basado en la distribución ZOIP, la estimación de los parámetros se hace vía máxima verosimilitud y la cuadratura de Gauss-Hermite utilizando diferentes alternativas. Se realizan tres estudios de simulación que muestran la convergencia de los parámetros para el ajuste de una distribución ZOIP y los diferentes casos de uso de los modelos de regresión ZOIP con efectos fijos y mixtos. Además, se ajustan diferentes modelos de regresión ZOIP que, en un caso real de una entidad bancaria colombiana, permiten explicar el porcentaje de uso de las tarjetas de crédito basándose en diferentes variables de negocio.

Palabras clave: modelos lineales mixtos, datos proporcionales inflados, cuadratura de Gauss-Hermite, máxima verosimilitud.