

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 la variable respuesta se encuentra definida a partir de una distribución para datos proporcionales, como la distribución beta o la distribución simplex, dichos datos pueden ser obtenidos a partir de variables medidas como porcentajes, tasas y proporciones, donde su resultado por lo general se encuentra ubicado en el intervalo cero uno, sin embargo, en ocasiones pueden dar resultados en cero y/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, la mayoría basados en estadística bayesiana, 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** y en **GitHub**, 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 las distribuciones y los modelos de regresión de efectos fijos y mixtos para las distribuciones beta en tres diferentes parametrizaciones y la distribución simplex, inflados con ceros y/o unos, basado en la distribución ZOIP, la estimación de los parámetros se hace vía máxima verosimilitud y la cuadratura de Gauss-Hermite utilizando diferentes alternativas, como la cuadratura de Gauss-Hermite adaptativa con o sin *pruning*. 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, a través de los estudios de simulación se presenta la alternativa de estimación que obtiene mejor desempeño, para cada modelo presentado. Además, se ajustan diferentes modelos de regresión ZOIP a un caso real sobre el porcentaje de uso de las tarjetas de crédito en una entidad bancaria colombiana y la dependencia a

diferentes variables de negocio que fueron definidas previamente como efectos fijos y aleatorios, dependiendo del caso de estudio.

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