COMPARACIÓN DE MÉTODOS DE IMPUTACIÓN DE DATOS FALTANTES Y APLICADO A LA BANCA UNIVERSAL Y COMERCIAL VENEZOLANA

Daniel Paredes
Escuela de Estadística,
Universidad de Los Andes,
Mérida, Venezuela
danielepm4384@yahoo.com

Rafael E. Borges
Escuela de Estadística,
Universidad de Los Andes,
Mérida, Venezuela
borgesr@ula.ve

Gerardo Colmenares L.
Instituto de Investigaciones Económicas y Sociales,
Universidad de Los Andes,
Mérida, Venezuela
gcolmen@ula.ve

Resumen

La presencia de datos faltantes es un problema que puede afectar el análisis de un conjunto de datos. En este trabajo se comparan varios métodos de imputación de datos faltantes para razones financieras indicadoras de riesgo bancario aplicados a datos de la banca universal y comercial venezolana. Los métodos de imputación analizados se enmarcan dentro de un contexto multivariado, considerándose cuatro métodos: i) imputación múltiple, ii) el algoritmo EM, iii) cadenas de Markov Monte Carlo y, iv) una metodología basada en modelos de series de tiempo multivariadas del tipo Vectores Auto-Regresivos (VAR). Para evaluar el desempeño de estos métodos utilizando datos completos o sin ausencia de información y eliminando valores de manera aleatoria en una proporción fija de pérdida que irá en aumento desde 0,05 hasta 0,5. Se plantean distintas proporciones de valores perdidos dado que el desempeño de cada método puede ser influenciado por el porcentaje de falta de información de manera distinta. Para determinar si existe diferencia significativa entre los métodos de imputación se utilizan modelos de regresión logística multinomial basados en la prueba no paramétrica de Friedman, que es análoga al Análisis de Varianza de dos vías, considerando como bloques a las distintas proporciones de datos ausentes y como tratamientos a los distintos métodos de imputación. Si se concluye que existe diferencia significativa entre los métodos de imputación respecto al sesgo en el poder de separación se aplicará la comparación múltiple por la prueba de sumas de rangos de Friedman para conocer cual método de imputación de datos genera menor diferencia en el sesgo en el poder de separación. También para comparar a las técnicas de imputación en los distintos niveles de porcentaje de ausencia de datos se realizará la prueba de Friedman pero usando como bloques a los distintos conjuntos de datos y como tratamientos a los métodos.

Palabras claves: Imputación de datos faltantes, prueba de Friedman, regresión logística multinomial, razones financieras, banca universal y comercial venezolana.