

MODELO PROBABILÍSTICO DE REGRESIÓN SIMBÓLICA PARA DATOS DE INTERVALO BASADO EN ESTADÍSTICOS DE ORDEN

Jorgelina Carrizo

Facultad de Filosofía, Humanidades y Artes. Universidad Nacional de San Juan, Argentina
jorgelinav_carrizo@yahoo.com.ar

El Análisis de Datos Simbólicos (ADS) es una extensión del análisis estadístico clásico que permite representar grupos de observaciones mediante variables multivaluadas, como intervalos conservando la variabilidad y la estructura intrínseca de los datos originales. El objetivo del presente trabajo es presentar un nuevo modelo de regresión lineal simbólica para datos de intervalo, en el que los intervalos se construyen a partir de los estadísticos de orden: mínimo y máximo. Partiendo de esta representación se establece una estructura probabilística para las variables de intervalo y se desarrolla una función de verosimilitud que permite obtener estimadores de máxima verosimilitud para los parámetros del modelo. Se analiza la derivación de la matriz de información de Fisher y las propiedades asintóticas de los estimadores. Estos resultados proporcionan un marco inferencial para el modelo, permitiendo evaluar la significancia de los parámetros y cuantificar la incertidumbre asociada a las estimaciones, además de garantizar que el límite superior del intervalo estimado sea siempre mayor que el inferior sin imponer restricciones adicionales. El modelo se aplica a un conjunto de datos reales, su performance se compara con el método del Centro introducido por Billard y Diday (2000), el método MinMax desarrollado por Billard y Diday (2002) y el método de Centro y Rango propuesto por Lima Neto y De Carvalho (2004). La evaluación comparativa se realiza mediante el error cuadrático medio para los límites inferior y superior de los intervalos y los coeficientes de correlación asociados, permitiendo analizar la precisión y la capacidad de ajuste. Se concluye que la metodología propuesta presenta una mayor precisión en la estimación de los coeficientes de regresión y constituye una alternativa inferencial sólida para el análisis de datos de intervalo.

Palabras Claves: Análisis de datos simbólicos, regresión lineal simbólica intervalar, estadísticos de orden, máxima verosimilitud, matriz de información de Fisher, inferencia estadística .

Referencias

- [1] Diday, E., Billard, L. (2006). Symbolic Data Analysis: Conceptual Statistics and Data Mining. Wiley.
- [2] Billard, L., Diday, E. (2003). From the statistics of data to the statistics of knowledge: Symbolic Data Analysis. JASA, 98(462), 470–487.
- [3] Billard, L., Diday, E. (2000). Regression analysis for interval-valued data.
- [4] Lima Neto, E. A., De Carvalho, F. A. T. (2008). Constrained linear regression models for symbolic interval-valued variables. Computational Statistics Data Analysis, 52(10), 4626–4641.
- [5] David, H. A., Nagaraja, H. N. (2003). Order Statistics. Wiley.
- [6] Lin, L.-C. (2026). Symbolic Interval-Valued Regression Model.