

LA REGRESIÓN LOGÍSTICA ORDINAL EN LAS CIENCIAS DE LA SALUD

La regresión logística acumulativa, también conocida como regresión logística ordinal, es una extensión o un caso especial de la regresión logística binaria, cuando una variable de resultado ordinal tiene más de dos niveles. El modelo de regresión logística acumulativa simple ha sido ampliamente implementado como el modelo predeterminado para el análisis de datos ordinales en *software* estadístico como SAS, SPSS, Stata, S Plus y R. Evalúa la relación entre un conjunto de variables predictoras y una variable de resultado ordinal a través de una función de enlace *logit*. Se conoce como el *logit* acumulativo, porque calcula las probabilidades acumulativas de estar en o por debajo de un nivel particular de la variable de respuesta. Además, para cada variable predictora se asume que las probabilidades acumulativas estimadas son las mismas en todas las categorías ordinales, lo que se conoce como el supuesto de *odds proporcionales*¹⁻².

El modelo de regresión logística acumulativa convencional asume que los datos son recolectados a través de un muestreo aleatorio simple, por lo cual cada unidad de muestreo tiene la misma probabilidad de ser seleccionada a partir de una población. No obstante, en estudios poblacionales, por lo general, los investigadores utilizan técnicas de muestreo más complejas, como la combinación del muestreo estratificado, muestreo por conglomerados y muestreo aleatorio simple; es decir, se usan diseños de muestras complejas³.

Debido a que los diseños de muestras complejas implican el uso de diferentes estratos (por ejemplo: áreas geográficas), con técnicas de muestreo agrupadas y probabilidades de selección desiguales pero conocidas, no es apropiado llevar a cabo el análisis del modelo de regresión logística acumulativa para variables de respuesta ordinal sin tener en cuenta los diseños de muestras complejas. De no hacerlo, pueden obtenerse estimaciones sesgadas de los parámetros, las estimaciones de la varianza pueden ser incorrectas y los resultados alejados de la realidad. Por lo tanto, es fundamental para los investiga-

dores entender las técnicas para el análisis de datos con diseños de muestras complejas³.

Entre los *software* estadísticos que pueden llevar a cabo un análisis estadístico dentro del contexto de diseños de muestras complejas se encuentra Stata, que es un paquete estadístico capaz de realizar numerosos análisis para datos provenientes de muestras complejas con su prefijo *svy*. No obstante, el uso de diseños de muestras complejas en el análisis de regresión logística para variables de respuesta ordinal es escaso¹⁻². Además, aunque los investigadores están cada vez más interesados en la realización de análisis secundarios de datos obtenidos a gran escala, la falta de las habilidades analíticas hace la tarea complicada. Por lo anterior, es necesario ayudar a los investigadores a comprender mejor la regresión logística para variables ordinales con datos de muestreo complejo y aplicarlo en datos reales.

Referencias

1. Agresti A. Analysis of ordinal categorical data. New Jersey: John Wiley & Sons, 2010.
2. Bender R, Grouven U. Ordinal logistic regression in medical research. Clin Med. 1997;31:546-51.
3. Lohr SL. Sampling: Design and Analysis. (2a Ed). Boston: Brooks/Cole; 2010.

Adriana Angarita-Fonseca

Fisioterapeuta.

Magíster en Epidemiología y Estadística

Estudiante Doctorado Universidad de Saskatchewan

Miembro Grupo de Investigación

en Manejo Clínico - CliniUDES

Universidad de Santander UDES,

Bucaramanga, Colombia

Correo electrónico: adriangarita@udes.edu.co