

Uso de machine learning para ajustar pesos en encuestas panel. Aplicación a la Encuesta Sanitaria y Social.

Luis Castro Martín
Universidad de Granada
luiscastro193@ugr.es

María del Mar Rueda García
Universidad de Granada
mrueda@ugr.es

Andrés Cabrera León
Escuela Andaluza de Salud Pública
andres.cabrera.easp@juntadeandalucia.es

Carmen Sánchez Cantalejo
Escuela Andaluza de Salud Pública
carmen.sanchezcantalejo.easp@juntadeandalucia.es

Ramón Ferri García
Universidad de Granada
rferri@ugr.es

Jorge Hidalgo Calderón
Universidad de Granada
jorgehcal@ugr.es

Iria Enrique Regueira
Instituto de Estadística y Cartografía de Andalucía
iria.enrique@juntadeandalucia.es

PALABRAS CLAVE

ajuste de pesos, encuestas panel, calibración, encuestas sanitarias, machine learning

RESUMEN

Las encuestas probabilísticas son ampliamente usadas por los servicios sanitarios en todo el mundo para obtener información que les permita tomar las decisiones adecuadas. A este respecto, la ESSOC (Encuesta Sanitaria y SOCial) es un proyecto de investigación cuyo objetivo es recolectar datos fiables sobre el impacto a lo largo del tiempo de la COVID-19, y de las medidas tomadas en consecuencia, en la población general así como en aquellos especialmente vulnerables. Incluye variables de interés sanitarias, socioeconómicas, psicosociales, conductuales, ocupacionales, ambientales y clínicas; determinantes a la hora de responder correctamente a la pandemia. La encuesta se ha llevado a cabo mediante un diseño de encuesta panel por superposición. Esto implica un problema de no-respuesta, especialmente agravado en este tipo de casos debido a la fatiga de la población al ser encuestada reiteradamente.

En este trabajo, exploramos una nueva metodología de ajuste de pesos que aprovecha la información auxiliar disponible a varios niveles para corregir la no-respuesta. El proceso consta de dos pasos: en una primera fase, los pesos originales de diseño son corregidos aplicando técnicas del estado del arte del machine learning para modelar la no-respuesta con respecto a la muestra longitudinal efectiva de la medición anterior. En la segunda fase, los pesos obtenidos son calibrados usando información auxiliar a nivel poblacional. También aplicamos el método propuesto a la ESSOC para estimar totales, proporciones y cambios a lo largo del tiempo.