

Encuesta Sanitaria y Social de Hogares: análisis y visualización mediante R y Python

Andrés Cabrera León

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

Luis Castro Martín

Universidad de Granada
luiscastro193@gmail.com

Iria Enrique Regueira

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

Ramón Ferri García

Universidad de Granada
rferri@ugr.es

Jorge Hidalgo Calderón

Universidad de Granada
jorgehcal@ugr.es

María del Mar Rueda García

Universidad de Granada
mrueda@ugr.es

Carmen Sánchez Cantalejo

Universidad de Granada
carmen.sanchezcantalejo.easp@juntadeandalucia.es

PALABRAS CLAVE

COVID-19, Encuestas, Machine Learning, Calibración, Visualización de Datos

RESUMEN

Las encuestas son ampliamente usadas por los servicios sanitarios de todo el mundo para obtener información que permita tomar las decisiones adecuadas. Con este propósito nace la Encuesta Sanitaria y Social de Hogares a partir de la colaboración entre el Instituto de Estadística y Cartografía de Andalucía, la Consejería de Salud y Familias de la Junta de Andalucía, las Universidades de Granada y Girona, y la Escuela Andaluza de Salud Pública. Su objetivo es contribuir al conocimiento de la situación y evolución de la población andaluza, respecto a su estado de salud, determinantes sociales, condiciones de vida y convivencia, desde el primer estado de alarma declarado con motivo de la COVID-19 por el Gobierno Español el 14 de Marzo.

La ESSOC tiene un diseño probabilístico de panel por superposición con 4 mediciones ya realizadas (abril, mayo y octubre de 2020, y abril de 2021). Por un lado, el

preprocesamiento y análisis estadístico de los datos de las ESSOC se realiza con R, obteniendo estimaciones ajustadas mediante técnicas de Machine Learning (tratamiento de la no respuesta) y calibración (tratamiento de la representatividad). Para cada variable (unas 150 en cada medición), se construyen diferentes tablas con estadísticos transversales para cada medición y estadísticos de cambio y diferencia de unas mediciones con respecto a las anteriores. Estas tablas son recuperadas en Python para su formateo y visualización mediante menús y son mostradas en Web mediante gráficos interactivos a través de la biblioteca Altair, muy potente para la visualización de datos con Python.