

Curso extracurricular: Introducción a la imputación de datos faltantes con R

Profesor: Augusto E. Hoszowski

Licenciado en Ciencias Matemáticas por la Universidad de Buenos Aires, es Especialista en Muestreo con amplia experiencia en análisis de datos y asesor estadístico en la Encuesta MICS de UNICEF. Entre 2001 y 2018 fue docente de Teoría y Técnicas de Muestreo en la Licenciatura en Estadística de la Universidad Nacional de Tres de Febrero. Fue coordinador de la Maestría en Diseño, Gestión y Análisis de Encuestas en la Universidad Nacional de La Matanza, coordinador de muestreo en el Instituto Nacional de Estadística y Censos y coordinador de metodología de la Dirección de Evaluación e Información Educativa del Ministerio de Educación de la Nación. Es expositor en la Diplomatura en Información y Evaluación Educativa (UNTREF-UNIFE), asesor estadístico en la evaluación del instrumento Value Drivers (Crow) y consultor estadístico en modelos aplicados al Machine Learning en DATAPREDICTOR (www.datapredictor.net).

Presentación curso

Este curso virtual se propone introducir a los/as alumnos/as en las herramientas básicas necesarias para abordar la problemática de los valores faltantes en las encuestas haciendo uso del software R.

Objetivos

Permitir a los/as alumnos/as:

- Conocer los diferentes tipos de no respuesta que aparecen en las encuestas y su impacto en la confiabilidad de los resultados.
- Capacitarse en el manejo informático que permita la corrección de la no respuesta mediante las diferentes técnicas de imputación.

Destinatarios

Estudiantes de grado, posgrado, profesionales e investigadores que deseen adquirir herramientas para el abordaje de la problemática de los datos faltantes. Se recomienda haber tomado un curso de muestreo y conocimientos básicos de programación R.

Temario

- La problemática de la no respuesta en los operativos estadísticos.
- Causas de la no respuesta. Diferentes clasificaciones. Los datos faltantes en el software R.
- Imputación y reponderación. Objetivos de la imputación. Evaluación de los métodos de imputación. Métodos usuales. Imputación por el modo, la media y mediana.
- Imputación por *hot deck*. Variantes.

- Imputación por regresión y vecino más cercano.
- Imputación mediante árboles aleatorios.
- Imputación multivariada. Método MICE para imputación multivariada.

Bibliografía

- Lohr, S. *Muestreo: Diseño y Análisis*, CENGAGE (2000)
- Lumley, T. *Complex Surveys: A Guide to Analysis Using R*. Wiley (2010)
- Medina, F., Galván, M., *Imputación de datos: teoría y práctica*, Naciones Unidas, CEPAL, División de Estadística y Proyecciones Económicas (2007)

Requisitos para la cursada y aprobación

Cumplir con el 80% de asistencia a los encuentros sincrónicos y aprobación de un trabajo práctico final.

Organización del curso

La cursada se organiza en 6 jornadas, una por semana. Cada jornada tendrá una duración de tres horas. En cada clase, luego de una exposición del tema, se ejemplificará en R las diversas técnicas, en bases de datos públicas de la Argentina. Se complementarán los temas con ejercitación domiciliaria

Modalidad: online sincrónica y asincrónica

Días y Horario: encuentros virtuales los días jueves de 18.00 a 21.00 horas

Fecha de inicio: 17/08/2023

Fecha de finalización: 05/10/2023

Total de horas: 24