

**CARRERA: MAESTRÍA GENERACIÓN Y ANÁLISIS DE INFORMACIÓN
ESTADÍSTICA**

AÑO: 2024 CUATRIMESTRE: 2°

Ciclo¹:

MATERIA Imputación de datos faltantes en encuestas

CARGA HORARIA: 32hs

DOCENTE TITULAR: Hoszowski, Augusto

DOCENTE ADJUNTO: Fernández, Luis Nahuel

CORRELATIVIDADES: Teoría y Técnicas de Muestreo

1-PRESENTACIÓN Y OBJETIVO DEL CURSO

Presentar un panorama de la problemática de los valores faltantes en las encuestas y en general en los operativos estadísticos, así como un abanico de técnicas para atacar este problema. Entre las herramientas para su resolución, se pretende utilizar tanto técnicas estadísticas “tradicionales” como la reponderación, o más novedosas como las que provienen del aprendizaje automático. Se hace énfasis en la implementación informática en R de las diversas técnicas con bases de datos del sistema estadístico nacional.

OBJETIVOS PARTICULARES:

Se espera que al completarse el curso los alumnos estén capacitados para evaluar qué técnica de imputación se adapta al problema concreto que se presenta y cómo implementarlo informáticamente en R.

Realizar simulaciones que guíen en la elección de la estrategia de imputación más conveniente

Estimar el error introducido al imputar los datos faltantes

2- DESARROLLO DE LOS CONTENIDOS

Problemática de la no respuesta en los operativos estadísticos. Diferentes clasificaciones. Impacto de la no respuesta en los estimadores: varianza y sesgo. Ejemplificación con simulaciones con R

Corrección de la no respuesta parcial y total. Imputación y reponderación. Ejemplos de

¹ CICLO Común u Orientación

estrategias desarrolladas en las estadísticas oficiales de Argentina

Imputación. Esquemas básicos y variantes. Ejemplificación con el paquete VIM.

Imputación por **hotdeck** y **kNN**. Simulaciones con **VIM** aplicados a las bases de microdatos públicas de EPH y Aprender

Arboles aleatorios. Breve introducción en el marco del *machine learning*. Implementación mediante **missForest**. Simulaciones con las bases de microdatos de Aprender.

Imputación multivariada. La metodología MICE y su implementación con el paquete **mice** de R. Aplicación con la base de microdatos de la encuesta MICS (UNICEF)

3-BIBLIOGRAFÍA PRINCIPAL

Muestreo: Diseño y Análisis, Sharon Lohr. Thomson (2000)

Imputación de datos: teoría y práctica, Medina, F., Galván, M., Naciones Unidas, CEPAL, División de Estadística y Proyecciones Económicas (2007)

Diseño de Muestras para encuestas a hogares. Directrices prácticas. Departamento de Asuntos Económicos y Sociales. División de Estadística. Estudio de Métodos. Serie F. Nro 98. Naciones Unidas (2009)

A review of hot deck imputation for survey non-response. International Statistical Review, 40-64. Andridge, R. and Little, R. (2010).

Ponderación de la muestra y tratamiento de valores faltantes en las variables de ingreso en la Encuesta Permanente de Hogares. Instituto Nacional de Estadística y Censos (INDEC). Metodología nº 15 (sin fecha).

La No Respuesta en las Encuestas. Jean- Jacques Droebeske – Pierre Lavallée. Centro de Estudios de Opinión. Universidad de Antioquía.

4-BIBLIOGRAFÍA COMPLEMENTARIA

Missing data imputation using regression and classification tree software **GUIDE**. JSM Proceedings. Survey Research Methods Section, American Statistical Association. Lee, H., Jeong, D. (2017).

Are deep learning models superior for missing data imputation in large surveys? Evidence from an empirical comparison. Wang, Z., Akande, O., Poulos, J., Li, F. (2021).

4-METODOLOGÍA y MODALIDAD DE CURSADA

Las clases comenzarán con una exposición del tema para luego realizar ejercitación en laboratorio. En el aula virtual se distribuirá material de consulta así como la ejercitación domiciliaria.

5-REQUISITOS PARA LA CURSADA Y PROMOCIÓN

a- ASISTENCIA A CLASES

75% de asistencia

b- EVALUACIÓN

Regularidad de la materia: 75% de asistencia y el primer Trabajo Práctico

Aprobación de la materia: aprobación de los dos Trabajos Prácticos

6-ORGANIZACIÓN DE CLASES

- I. No respuesta en los operativos estadísticos
- II. Clasificación de la no respuesta. Corrección de la no respuesta total y la no respuesta parcial
- III. Imputación, Definición y propiedades básicas. Introducción a las simulaciones con R.
- IV. Hotrdeck. Implementación en R.
- V. Imputación por regresión. Variantes. Implementación en R. Ejemplo con bases de microdatos de la Encuesta Permanente de Hogares y del operativo Aprender.
- VI. Knn. Variantes. Implementación en R. Ejemplo con bases de microdatos de MICS.
- VII. Imputación mediante árboles aleatorios. Estimación del error de imputación. Ejemplo con las encuestas PISA.
- VIII. Imputación multivariada. Implementación con el paquete de R **mice**.

	Temario / Unidad	Modalidad		
		Presencial		Virtual
		Aula común	Aula laboratorio	
Clase 1	Problemática de la no respuesta en las encuestas por muestreo. Clasificaciones.		X	
Clase 2	Corrección de la no respuesta. Imputación y reponderación.		X	
Clase 3	Efectos de la no respuesta. Varianza y sesgo.		X	
Clase 4	Imputación por hotdeck		X	
Clase 5	Imputación por regresión		X	
Clase 6	Imputación vecinos más cercanos		X	
Clase 7	Imputación por árboles aleatorios		X	
Clase 8	Imputación multivariada		X	

7-FECHA DE EXÁMEN O ENTREGA DE TRABAJO FINAL

a- PRIMERA FECHA:

b- SEGUNDA FECHA:

Encuentros virtuales y presenciales los días miércoles de 18 a 22 hs. Fecha de inicio:

02/10/24 Fecha de finalización: 20/11/24 Total de horas: 32

Fecha límite de pago de matrícula: 28 DE SEPTIEMBRE DE 2024

ARANCEL: 96.000 pesos

Inscripción y consultas: maestriaestadistica@untref.edu.ar