

Programación para el análisis de datos

Departamento de Ciencias Sociales, Universidad Católica del Uruguay

Docente: Martín Opertti

Fecha: Agosto 2022

Correo: operttim@gmail.com

El curso familiariza a los estudiantes con la programación para el análisis de datos, con un enfoque en las habilidades relevantes para el análisis de datos en ciencias sociales. El curso utiliza principalmente el entorno y lenguaje de programación de R. El curso abarca cómo estructurar, transformar, explorar, analizar y visualizar datos en R. También son introducidos conceptos básicos de Stata y Python.

Objetivos del curso: El objetivo del curso es que los estudiantes adquieran:

- Conocimientos sobre como estructurar, manipular, analizar y visualizar datos
- Herramientas para comprender y escribir código en R -reutilizable y interpretable- para el análisis de datos
- Conocimientos para aprender de forma independiente funciones en R
- Nociones básicas de Stata y Python para el análisis de datos

Calificaciones: Una calificación mínima de B es requerida para exonerar el curso.

Asistencia: Los alumnos deberán asistir al 75% de las clases para tener la posibilidad de exonerar el curso

Horario: El curso se dictará los días martes y jueves de 8:00 a 9:20, desde el 8 de agosto hasta el 1 de diciembre.

Evaluación:

Evaluación parcial	20%
Evaluación parcial II	40%
Trabajo final	40%

Escala de calificaciones:

6	S	Aprobado
5	MB	Aprobado
4	BMB	Aprobado
3	B	Aprobado
2	R	No aprobado
1	D	No aprobado

*Solo los estudiantes con al menos 75% de asistencia a clase podrán aprobar el curso

Estructura del curso:

- **Semana 1: Introducción al curso y al análisis de datos**
 - ¿Cómo pensar los datos?
 - ¿Qué es y para qué sirve la programación?
 - Paquetes estadísticos y lenguajes
- **Semana 2: Introducción a Stata**
 - Crear/Importar datos
 - Estructura de sintaxis
 - Manipulación básica de datos
 - Tablas
- **Semana 3: Introducción a Stata II**
 - Recodificar variables
 - Estadística descriptiva (media, mediana)
 - Estadística inferencial (t-test, regresión, etc.)
- **Semana 4: Primera evaluación parcial**
- **Semana 5: Introducción a R**
 - Interfaz, obtener ayuda
 - Operadores, objetos, funciones y paquetes
 - Funciones básicas
 - **Bibliografía:** Capítulos 1, 2, 4, 5 y 6, R for data Science, Wickham, H., & Golemund, G. (2016).
- **Semana 6: Introducción a R II**
 - Marcos de datos (dataframes)
 - Funciones básicas
- **Semana 7: Importar y limpiar datos en R**
 - Importar y exportar datos
 - Introducción al Tidyverse
 - Recodificación de variables
- **Semana 8: Estadística descriptiva en R**
 - Tablas descriptivas
 - Estadísticos descriptivos
- **Semana 9: Manipulación de datos en R**
 - Tidyverse (ordenar, agrupar, filtrar, resumir, etc.)
- **Semana 10: Manipulación de datos en R II**
 - Integración de bases de datos
 - Cambios de estructura de base de datos

- **Semana 11: Visualización de datos en R**
 - Introducción a la visualización de datos
 - Análisis exploratorios con gráficos
- **Semana 12: Estadística inferencial**
 - Pruebas de hipótesis
 - Estimar y visualizar modelos de regresión
- **Semana 13: Reportes con R Markdown y técnicas de programación avanzadas**
 - Introducción a R Markdown
 - Loops
 - Vectorización
 - Condicionales
- **Semana 14: Segunda evaluación parcial**
- **Semana 15: Introducción a Python**
- **Semana 16: Presentaciones y cierre del curso**

Bibliografía:

- Acock, A. C. (2008). A gentle introduction to Stata. Stata press.
- Grolemund, G., & Wickham, H. (2018). R for data science.
- Rodríguez, G. (2011). Stata tutorial. <https://data.princeton.edu/stata/>
- Urdinez, F. & Labrín, A. (Eds.) (2020) AnalizaR Datos Políticos