

Curso de R para el análisis de datos



Resumen del curso

El enfoque principal del curso será aprender el uso de R con énfasis en el análisis de datos.

Se espera que al finalizar el curso el participante haya obtenido la capacidad de hacer un primer análisis de datos en R y tenga la autonomía para profundizar en el uso de R en su área de interés.

Objetivo

Proporcionar a los participantes del curso una sólida formación básica en los fundamentos de análisis de datos con R y RStudio, abarcando las fases de limpieza, exploración, modelado de datos e interpretación de resultados.

Se hará especial énfasis en la comprensión de la salida del programa R, así como en el desarrollo de la capacidad de utilizar de forma crítica estos métodos en actividades de su área de trabajo.

Pre-requisitos:

Este curso presupone conocimiento teórico de los temas estadísticos a tratar en las diferentes clases. Es recomendable que, en caso que el estudiante lo considere necesario, realice un repaso de los aspectos teóricos del tema a tratar en clase.

No se supondrá ninguna familiaridad con el programa R ni con ningún otro lenguaje de programación.

Cronograma

El curso consta de 16 clases teórico-prácticas de 2:30 hs. cada una, los días lunes y miércoles, de 10:30 a 13:00 hs. a partir del 23/05/2022.

Luego los participantes dispondrán de dos semanas para realizar y entregar un trabajo de cierre de curso que constará de un análisis hecho en R con datos de su área de interés.

El curso finalizará el 27/07/2022 con la presentación de las distintas aplicaciones realizadas por los participantes.

Cronograma

Clase	Fecha	Horario	Tema
Clase 1	Lunes 23 de mayo	10:30 a 13:00	Introducción
Clase 2	Miércoles 25 de mayo	10:30 a 13:00	Estructuras básicas de R
Clase 3	Lunes 30 de mayo	10:30 a 13:00	Códigos eficientes
Clase 4	Miércoles 1 de junio	10:30 a 13:00	Importación y exportación de datos en R
Clase 5	Lunes 6 de junio	10:30 a 13:00	Análisis exploratorio de datos
Clase 6	Miércoles 8 de junio	10:30 a 13:00	tidyverse
Clase 7	Lunes 13 de junio	10:30 a 13:00	Visualización estadística de datos
Clase 8	Miércoles 15 de junio	10:30 a 13:00	Reportes en Rmarkdown
Clase 9	Lunes 20 de junio	10:30 a 13:00	Correlación
Clase 10	Miércoles 22 de junio	10:30 a 13:00	Regresión lineal simple
Clase 11	Lunes 27 de junio	10:30 a 13:00	Modelos de respuesta binaria
Clase 12	Miércoles 29 de junio	10:30 a 13:00	Introducción a series de tiempo
Clase 13	Lunes 4 de julio	10:30 a 13:00	Análisis de series de tiempo
Clase 14	Miércoles 6 de julio	10:30 a 13:00	Análisis de clústers
Clase 15	Lunes 11 de julio	10:30 a 13:00	Aplicaciones web interactivas
Clase 16	Miércoles 13 de julio	10:30 a 13:00	Clase de dudas
Libre	Miércoles 20 de julio		
Clase 17	Lunes 25 de julio	10:30 a 13:00	Presentación de trabajos
Clase 18	Miércoles 27 de julio	10:30 a 13:00	Presentación de trabajos

Evaluación

La evaluación del aprendizaje se realizará de dos formas:

- Evaluación continua: consignas sobre los distintos contenidos temáticos a realizar con datos de interés o proporcionados por los docentes

- A través de la evaluación del trabajo final del curso

Vías de comunicación

Se dispondrá de una carpeta en Google Drive donde se dejarán disponibles para cada semana del curso, las presentaciones, los scripts, los conjuntos de datos utilizados y la bibliografía utilizada.

También podrán encontrar los materiales en <http://alea.uy/index.php/cursor/>

Ingresando con la clave **cursoRbps**

La forma de comunicación será vía mail a través de contacto@alea.uy
contacto.alea.uy@gmail.com

Bibliografía

- R para principiantes (<https://bookdown.org/jboscomendoza/r-principiantes4/>)
- El Arte de programar en R (https://cran.r-project.org/doc/contrib/Santana_El_arte_de_programar_en_R.pdf)
- Robert Kabacoff - R in Action Data Analysis and Graphics with R
- R for Data Science (<https://r4ds.had.co.nz/>) (<https://es.r4ds.hadley.nz/>)
- Alejandro Quintela del Rio - Estadística Básica Edulcorada (<https://bookdown.org/aquintela/EBE/>)

¿Qué es la Estadística?



Estadística

La Estadística es una disciplina científica cuyo propósito es apoyar, favorecer o facilitar la realización de inferencias inductivas basadas en datos mediante:

- i) el resumen de la información contenida en los datos (a efectos de que los usuarios puedan realizar sus inferencias a base de tal resumen), y
- ii) el dimensionamiento o cuantificación o control del error inherente a toda inferencia inductiva.

CARLOS ARAÚJO



Sabiduría

Conocimiento

Información

Datos



¿Qué es R?

Además de ser un lenguaje de programación estadístico R es:

- Una herramienta muy poderosa para todo tipo de manipulación y procesamiento de datos
- Una comunidad de programadores, usuarios, académicos y profesionales
- Una herramienta que nos permite hacer variedad gráficos y visualizaciones de datos
- Una colección gratuita de paquetes que extienden su funcionalidad básica

Beneficios de R

- Es gratis
- Es libre ("Think free as in free speech, not free beer." Richard Stallman)
- Funciona en distintas plataformas
- Comunidad activa con muchos tutoriales y en permanente actualización
- Gráficos estadísticos de gran calidad y versatilidad
- Ayuda a profundizar en aspectos teóricos que fundamentan los resultados.

¿Por qué R?

John Chambers opina que el tiempo que usted invierta en desarrollar y extender sus habilidades con el sistema R lenguaje, paquetes y ambiente de programación, le redundará en la capacidad de cuestionar y contestar con confianza en el cálculo con datos. "La exploración de los datos con las preguntas correctas y respuestas confiables son fundamentales para el análisis de datos"

Más ventajas

- R fue diseñado por estadísticos (Ross Ihaka y Robert Gentleman) para hacer el procesamiento de datos más fácil.
- En el desarrollo de R están involucrados científicos de primer nivel tanto en el lado estadístico como de cómputo lo que garantiza un software de excelencia.
- R es sin duda el software estadístico más usado en investigación estadística, pero también en otras áreas como finanzas, medicina, biología, etc.
- El carácter de colaboración abierta por medio de paquetes se refleja en la posición que ocupa R en la frontera de la investigación.
- R le ofrece un ambiente que permite llevar a cabo sus ideas sin limitarlas únicamente a los procedimientos incluidos en su aplicación estadística.

Aplicaciones realizadas en R

- <http://informe-mortalidad-2020.cardiosalud.org/>
- <https://bbc.github.io/rcookbook/>
- <https://www.bps.gub.uy/1920/boletin-estadistico.html>
- `demo()`



```
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

> |
```