

Temario

Herramientas estadísticas en R

Esteban Degetau

28 de febrero de 2025

Objetivo

R es un lenguaje de computación diseñado específicamente para análisis estadístico y gráfico (Crawley 2012). El objetivo de este curso es brindar las herramientas estadísticas para resolver problemas de datos cotidianos en organizaciones grandes como el IMSS. Al final del curso podrás:

1. Resolver una variedad de preguntas empíricas usando R.
2. Implementar flujos de datos automáticos desde la fuente de datos hasta medios de comunicación diversos (tablas, figuras, reportes, presentaciones, páginas web, etc.).
3. Usar Quarto para escribir reportes reproducibles y GitHub para versión de control.

Prerequisitos: Este curso asume que tienes conocimientos básicos de estadística y programación. No asumo que tengas experiencia previa con R.

Contenido

1. R y el Tidyverse
 - El lenguaje estadístico R
 - Paquetes del Tidyverse
 - Manipulación de datos con `dplyr`
 - Visualización de datos con `ggplot2`
 - Programación funcional con `purrr`
 - Reportes reproducibles con Quarto

Referencias: Crawley (2012); Wickham, Grolemund, et al. (2017); Wickham (2010); Wickham (2019).

2. Inferencia estadística

- Propiedades asintóticas de los estimadores
- Pruebas de hipótesis
- Bootstrap y métodos de remuestreo
- Simulaciones

Referencias: Aguilar Esteva (s. f.); Schmelzer (s. f.); Çetinkaya-Rundel y Hardin (2024).

3. Regresión lineal

- Mínimos cuadrados ordinarios
- Inferencia estadística
- Interpretación de coeficientes
- Sesgo por variables omitidas
- Efectos fijos con `fixest`

Referencias: Aguilar Esteva (s. f.); Schmelzer (s. f.); Cunningham (2021).

4. Pronóstico de series de tiempo

- Suavización exponencial
- ARIMA
- Pronóstico automatico con `fable`

Referencias: Hyndman y Khandakar (2008); Hyndman y Athanasopolous (2021).

5. Aprendizaje de máquina

- Supervisado: OLS y clasificación
- No supervisado: Componentes principales y clasificación
- Elevar la práctica del modelaje con `Tidymodels`

Referencias: Athey y Imbens (2019); James et al. (2021); Silge y Julia (s. f.).

Fechas y programa

Las clases son de manera presencial en la Sala de Juntas de la CPE los Miércoles de 11:30 a 13:30. Tomaremos cinco clases en total del 12 de marzo al 9 de abril.

Evaluación

Tareas: 50%

A lo largo del curso, asignaré una tarea por cada tema. Las tareas se entregarán en la fecha indicada en clase. Pueden discutir las tareas con sus compañeros, pero cada estudiante debe entregar su propio trabajo.

La entrega de la tarea consta de un documento PDF con los análisis requeridos y el código de Quarto que genera el documento. Parte importante de la calificación de las tareas es que el código de Quarto corra sin errores en mi computadora.

Las tareas serán difíciles y requerirán tiempo para completarlas. No dejen las tareas para el último día.

Proyecto final: 50%

El proyecto final debe cumplir con las siguientes características:

1. Analizar datos del IMSS.
2. Comunicar el análisis en un reporte reproducible de Quarto.
3. Explorar alguna pregunta importante en tu trabajo.

Los requisitos del proyecto final son intencionalmente vagos para permitir la creatividad y la exploración de temas que te interesen o relacionados con tu trabajo. El objetivo principal del proyecto final es que aprendas cómo puedes usar R y Quarto para resolver problemas cotidianos en tu trabajo.

Algunos ejemplos de proyectos finales son:

- La automatización de (alguna porción de) algún reporte o presentación mensual que ya tengas.
- Un ejercicio de predicción de series de tiempo (por ejemplo, fijación de metas).
- Un análisis de regresión de alguna variable de interés.
- Un ejercicio de clasificación supervisada o no supervisada.

Cualquier otro proyecto que cumpla con los requisitos es más que bienvenido. Si tienes una idea vaga de qué podrías hacer, podemos discutirlo en clase.

! No trabajos en Excel

Puesto que el objetivo del curso es aprender a programar en R y establecer flujos de datos automatizados, no se aceptarán tareas o proyectos que usen Excel para pasos intermedios.

Recursos adicionales

Además de las referencias en este documento, también puedes encontrar ligas de interés en la página de [recursos](#).

- Aguilar Esteva, Arturo. s. f. «Notas de Microeconometría». <https://bookdown.org/viclrz/notasmicro/>.
- Athey, Susan, y Guido W. Imbens. 2019. «Machine Learning Methods That Economists Should Know About». *Annual Review of Economics* 11 (1): 685-725. <https://doi.org/10.1146/annurev-economics-080217-053433>.
- Çetinkaya-Rundel, Mine, y Johanna Hardin. 2024. *Introduction to Modern Statistics*. Vereinigte Staaten: OpenIntro, Inc. <https://openintro-ims.netlify.app/>.
- Crawley, Michael J. 2012. *The R Book*. John Wiley & Sons.
- Cunningham, Scott. 2021. *Causal Inference: The Mixtape*. Yale University Press. <https://www.jstor.org/stable/j.ctv1c29t27>.
- Hyndman, Rob J., y George Athanasopolous. 2021. *Forecasting: Principles and Practice*. Melbourne, Australia: OTexts. <https://otexts.com/fpp3/>.
- Hyndman, Rob J., y Yeasmin Khandakar. 2008. «Automatic Time Series Forecasting: The **Forecast** Package for *R*». *Journal of Statistical Software* 27 (3). <https://doi.org/10.18637/jss.v027.i03>.
- James, Gareth, Daniela Witten, Trevor Hastie, y Robert Tibshirani. 2021. *An Introduction to Statistical Learning: With Applications in R*. 2nd ed. 2021 edition. New York NY: Springer. <https://www.statlearning.com/>.
- Schmelzer, Christoph Hanck, Martin Arnold, Alexander Gerber, and Martin. s. f. *Introduction to Econometrics with R*. <https://www.econometrics-with-r.org/>.
- Silge, Max Kuhn, y Julia. s. f. *Tidy Modeling with R*. <https://www.tmwr.org/>.
- Wickham, Hadley. 2010. *Ggplot2: Elegant Graphics for Data Analysis*. 1st ed. 2009. Corr. 3rd printing 2010 edition. Switzerland: Springer. <https://ggplot2-book.org/>.
- . 2019. *Advanced R, Second Edition*. 2nd edition. Boca Raton London New York: Chapman; Hall/CRC. <https://adv-r.hadley.nz/>.
- Wickham, Hadley, Garrett Grolemund, et al. 2017. *R for data science*. Vol. 2. <https://r4ds.had.co.nz/>.