

R para la investigación cuantitativa

Especialización en Investigación Científica UNEMI

Laboratorio de Investigación para el Desarrollo del Ecuador

Instructor: Daniel Sánchez, MA **Conocimientos previos:** Habilidades básicas en computación, Git, y shell.

1 Resumen del módulo

Este módulo introduce los conceptos básicos del lenguaje de programación R para el análisis de datos y la computación estadística. El material está diseñado para quienes buscan utilizar R como una herramienta para métodos de investigación cuantitativa.

Asumo que los estudiantes tienen poca experiencia previa en programación: familiaridad con habilidades básicas de computación, Git y la línea de comandos. Cubriremos muchos conceptos importantes del lenguaje, sin embargo, el enfoque no está en la programación técnica, sino en entender el lenguaje como una herramienta para el análisis estadístico.

2 Contenidos del módulo

El siguiente es un esquema planeado del curso. Esto puede cambiar dependiendo del ritmo de la clase.

2.1 Introducción a R (14, 16 y 17 de agosto)

Introducción al lenguaje de programación R, con un enfoque a investigación cuantitativa y programación estadística. Los estudiantes explorarán la interfaz RStudio y las ventajas que ofrece para trabajar con R. Los estudiantes aprenderán los conceptos básicos de programación como clases de objetos, estructuras de datos, bucles y funciones. Adicionalmente, aprenderán las bases de cómo importar y limpiar bases de datos en diferentes formatos mediante R base y los diferentes paquetes disponibles.

2.2 Limpieza y manipulación de datos - Parte 1 (21, 23 y 24 de agosto)

El módulo priorizará la utilización de R como herramienta de análisis de datos, utilizando la metodología tidy de programación. Se introduce el lenguaje para leer archivos de diferentes formatos, crear variables, filtrar observaciones, manipular índices, seleccionar columnas de tablas, y crear resúmenes descriptivos de bases de datos para investigaciones cuantitativas.

2.3 Limpieza y manipulación de datos - Parte 2 (28, 30 y 31 de agosto)

El módulo continúa en el uso de R como herramienta de análisis de datos. En esta segunda parte, se profundizan los contenidos aprendidos, demostrando procedimientos de limpieza de datos más complejos como la conversión de tipos de datos, manejo avanzado de filas duplicadas y vacías, manipulación de datos categóricos o de texto, vinculación de registros, entre otros. Se introduce la noción de las reconfiguraciones (reshape o pivot) de datos y su uso, dependiendo del contexto.

2.4 Estadística descriptiva y visualización de datos en R (4, 6 y 7 de septiembre)

Este módulo se enfocará en el uso de R para el análisis estadístico y visualización de datos. Se presentará el uso de las funciones de estadística descriptiva (contar, media, mediana, varianza y desviación estándar). A la par, los estudiantes aplicarán la metodología tidy para aprovechar las posibilidades de agrupación de bases de datos por variables categóricas. Se presentará el uso de R base y el paquete ggplot2 para visualización de datos a través de gráficos de barra, diagramas de caja, histogramas, distribuciones acumuladas, entre otros.

3 Software

3.1 Instalación de R y RStudio

Este curso requiere la instalación de R y RStudio. Ambos son software libre y de código abierto, y están disponibles para Windows, macOS y Linux. Necesitarás instalarlos **antes** de la primera clase.

- **R:** Descarga la última versión de R desde el Comprehensive R Archive Network (CRAN) en <https://cran.r-project.org/>. Sigue las instrucciones para tu sistema operativo.
- **RStudio:** Descarga la última versión de RStudio Desktop desde [aquí](#). Sigue las instrucciones para tu sistema operativo.

Si tienes problemas con la instalación, por favor publica tu consulta en el canal de Slack.

A continuación encontrarás un tutorial de instalación para Windows en video:

https://www.youtube.com/watch?v=_LBJvSZ2JxY

3.2 Teclado

Rutinariamente necesitaremos escribir símbolos como “/”, “<-”, “%>%” y otros. Asegúrate de estar cómodo con la distribución de tu teclado y de que puedas escribir estos símbolos fácilmente. Esto puede parecer trivial, pero es importante para el curso, ya que absolutamente no podemos permitirnos perder tiempo buscando símbolos en el teclado. Es posible que necesites cambiar la distribución del teclado al idioma correcto para que la computadora siga la disposición física de tu teclado. Para los usuarios de Windows, esto se puede hacer fácilmente presionando **Win + Espacio** y seleccionando la distribución correcta (consulta [aquí](#)).

A continuación encontrarás un tutorial de configuración de teclado en video:

https://www.youtube.com/watch?v=Yt_HGJRM5vA.

4 Materiales del curso

Se proveerán todos los materiales del curso en el Google Drive del programa. Los materiales incluyen diapositivas, código y bases de datos.

5 Referencias

A continuación se presentan algunas referencias útiles para el curso. No es necesario leer estos libros, pero pueden ser útiles para aquellos que deseen profundizar en R.

Chang, Winston. 2018. *R Graphics Cookbook: Practical Recipes for Visualizing Data*. O'Reilly Media. <https://r-graphics.org/>.

Heiss, Florian. 2020. *Using R for Introductory Econometrics*. Düsseldorf: Independently published. <https://www.urfie.net/>.

Huntington-Klein, Nick. 2022. *The Effect: An Introduction to Research Design and Causality*. 1st edition. Boca Raton: Chapman and Hall/CRC. <https://theeffectbook.net/>.

Wickham, Hadley, Mine Çetinkaya-Rundel, and Garrett Golemund. 2023. *R for Data Science*. "O'Reilly Media, Inc.". <https://r4ds.hadley.nz/>.

Xie, Yihui. 2017. *Dynamic Documents with R and Knitr*. Chapman and Hall/CRC. <https://duhi23.github.io/Analisis-de-datos/Yihue.pdf>.

Xie, Yihui, Joseph J Allaire, and Garrett Golemund. 2018. *R Markdown: The Definitive Guide*. Chapman and Hall/CRC. <https://bookdown.org/yihui/rmarkdown/>.