

Curso de Introducción a R

Pablo Cabrera Álvarez

17/6/2019

Objetivos

Este curso tiene como objetivo introducir la dinámica de trabajo en R para investigadoras que previamente usaban programas de tipo GUI, como SPSS o Stata.

Estructura

El curso consta de cinco módulos básicos (MB) en los que se expondrán los principios generales del lenguaje R y cinco módulos optativos (MO) para profundizar en diferentes posibilidades que ofrece el programa. Se recomienda realizar todos los módulos básicos y añadir los optativos que se consideren necesarios. Cada módulo tiene una duración de 4 horas e incluyen aproximadamente una hora y media de exposición y dos horas de práctica.

Introducción a R (MB)

Paquetes: base *Datos:* mtcars

- ¿Por qué R?
- Trabajar con R-Studio
- Cómo pedir ayuda `help()`, `?` y otras fuentes
- Aspectos básicos del lenguaje orientado a objetos
- El *workspace* en R
- Tipos de objetos en R

Práctica: Generación e interacción con los diferentes tipos de objetos en R. Búsqueda de ayuda y lectura de la documentación.

Operaciones básicas con R (MB)

Paquetes: base, forcats, lubridate, hms, stringr *Datos:* mtcars, titanic

- Transformar objetos `class()` y `as.*()`
- Operadores básicos
- Factores `factor()`, `levels()`, `ordered()`, `fct_reorder()` y `fct_relevel()`
- Cadenas de texto `str_locate()`, `str_remove()`, `str_replace()` y `str_sub()`
- Fechas `dmy_hms()` y `parse_date_time()`
- Valores perdidos NA

Práctica:

Manipulación de datos I (MB)

Paquetes: base, haven, readr, dplyr, tidyr *Datos:* mtcars, iris, Barómetro CIS

- Leer datos en R `read_csv()`, `read_spss()`, `read_rds()` y `load()`
- Explorar un conjunto de datos `head()`, `tail()`, `str()` y `colnames()`
- Ordenar `arrange()`
- Seleccionar variables `select()`
- Filtrar casos con `filter()`
- Crear nuevas variables `mutate()`
- Recodificar con `recode()`

Práctica: Manipular un conjunto de datos de encuesta.

Manipulación de datos II (MB)

Paquetes: base, dplyr, tidyr *Datos:* Base de datos electorales, Barómetro CIS, iris

- Introducción al uso de *pipes* %>%
- Agrupar `group_by()`
- Resumir `summarise()`
- Combinar bases de datos: filas `rbind()` y `bind_rows()`
- Combinar bases de datos: columnas `cbind()`, `bind_cols()`, `left_join()` y `right_join()`
- Cambiar de formato largo a ancho `spread()`
- Cambiar de formato ancho a largo `gather()`

Práctica: Manipular varios conjuntos de datos para crear una base de datos conjunta.

Tabulaciones y descriptivos (MB)

Paquetes: base, dplyr, tidyr, readr, haven, sjmisc, expss *Datos:* Barómetro CIS

- Tablas de frecuencias `frq()`
- Estadísticos descriptivos `descr()`
- Tablas de contingencia `crosstab()`
- tablas para preguntas de respuesta múltiple `tab_cells()`
- Tablas personalizadas `tab_*`
- Exportar resultados `openxlsx`

Práctica: Crear tabulaciones para análisis y tablas para ser exportadas.

Informes integrados en R (MO)

Paquetes: base, tidyverse, rmarkdown, kable *Datos:* mtcars

- Rmarkdown en el proceso de trabajo con datos
- Posibilidades de Rmarkdown: outputs
- Markdown
- Insertar datos y código
- Documentos interactivos
- Generar documentos

Práctica: Generar un informe en rmarkdown que incluya manipulación de los datos y análisis.

Visualización de datos con ggplot2 (MO)

Paquetes: base, tidyverse *Datos:* mtcars, barómetro CIS

- Introducción a la gramática de gráficos
- Preparación de datos `ggplot()` y `aes()`
- Geometrías para definir el gráfico `geom_*`
- Gráficos de estadísticas `stat_*`
- Gráficos múltiples y aspectos estéticos `facet_grid()`, `facet_wrap()` y `theme()`

Práctica: Generar diferentes representaciones gráficas a partir de diversos conjuntos de datos.

Análisis de datos de encuesta (MO)

Paquetes: base, tidyverse, survey *Datos:* ESE

- Declarar un diseño de datos complejo `svydesign()`
- Estadísticos y errores ajustados `svymean()`, `svytotal()` y `svyratio()`

- Análisis de subpoblaciones `svyby()`
- Tests de asociación `svytable()` y `svychisq()`
- Modelos estadísticos con diseño complejo `svyglm()`

Práctica: Declarar un diseño de datos complejos y utilizarlo para analizar la Encuesta Social Europea (ESE).

Muestreo y ponderaciones (MO)

Paquetes: base, tidyverse, sampling, survey *Datos:* Registro de datos electorales, encuesta electoral

- Muestreo SRS en R `sample()`
- Muestreo complejo: estratos
- Muestreo complejo: conglomerados
- Computación de coeficientes de ponderación `postStratify()` y `calibrate()`

Práctica: Realizar un muestreo desde una base de datos de registros electorales y ajustar una encuesta con datos electorales.

Modelos estadísticos (MO)

Paquetes: base, tidyverse, stats *Datos:* mtcars, iris, barómetro CIS

- Regresión lineal `lm()`
- Regresión logística `glm()`
- Modelos de clasificación `kmeans()`
- Análisis factorial `princomp()` y `princomp()`

Práctica: Realizar los diferentes tipos de modelos.

Metodología

Cada sesión tendrá una duración aproximada de cuatro horas y serán de carácter práctico. Las explicaciones irán mostrando código que se podrá ir ejecutando por parte de las investigadoras. Cada sesión tendrá una práctica que consistirá en un cuadernillo, un script comentado y un script con las soluciones de cada ejercicio.

Bibliografía básica

Recursos