



Análisis descriptivo

Sesión N° 6

13 septiembre 2021

Análisis de datos estadísticos en R

Profesora Valentina Andrade de la Horra
Ayudantes Dafne Jaime y Nicolás Godoy

Contenidos Sesión 6



Tendencia central

Dispersión

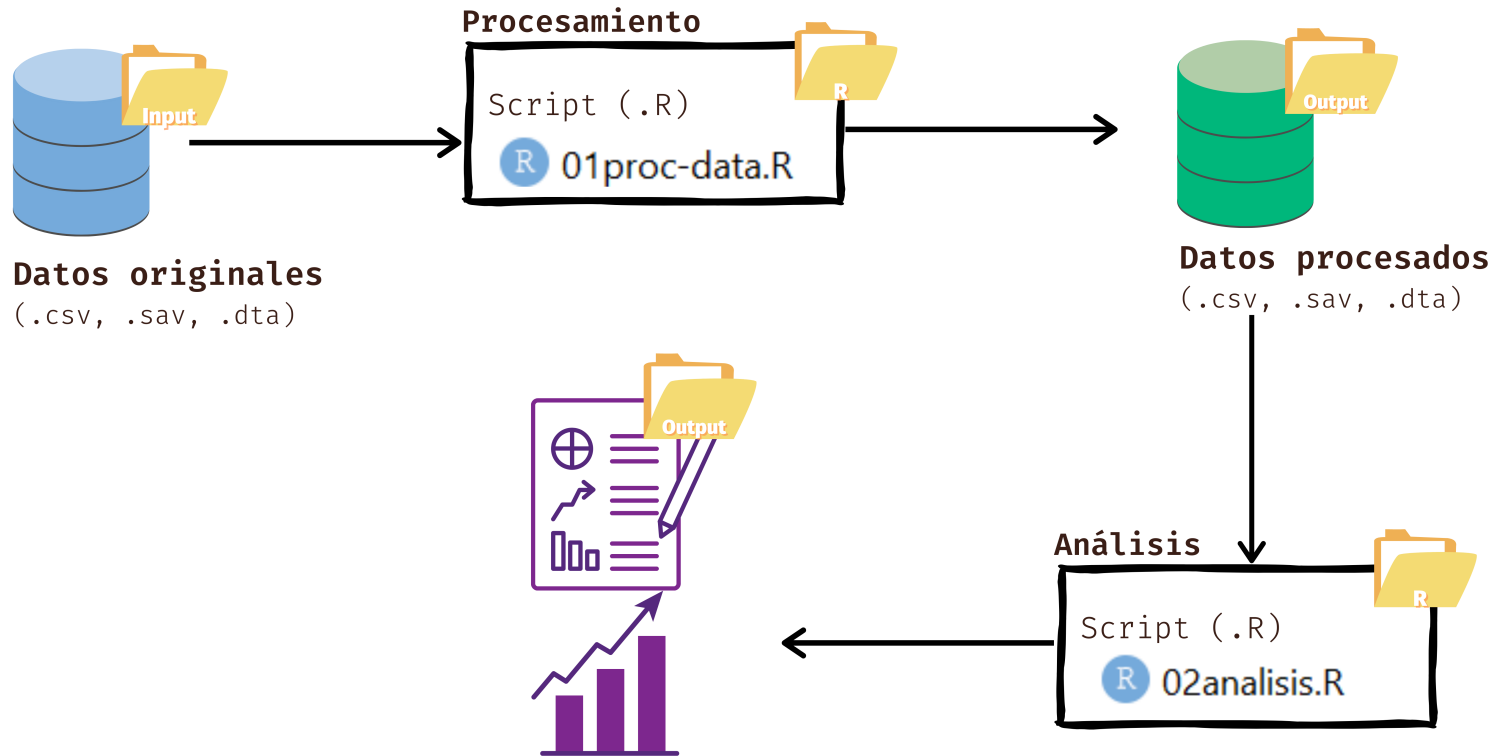
Proporciones

Visualización



1: Flujo del Rproject

Etapas del flujo



- *Hoy partimos con análisis descriptivo*

Orden de un script de análisis



- # 1. *Cargar paquetes*
- # 2. *Cargar datos*
- # 3. *Explorar datos*
- # 4. *Análisis descriptivo*
- # 5. *Output para reporte*



Figura 1: Estudiantes de Análisis de datos en R haciendo el **paso 1 y 2.**



Paso 1: Cargar paquetes

Paso 1: Cargar paquetes



```
pacman::p_load(tidyverse,  
               magrittr,  
               sjPlot,  
               sjmisc)
```

```
load("output/datos/misdatos.RData")
```




Paso 3: Explorar datos

¿Por qué creen que sería relevante este paso?

1.

2. 3.

Explorar datos para análisis



1. Base

```
dim(datos) # Nos entrega las dimensiones, es decir el numero d  
View(datos) # Visualizar objetos  
names(datos) # entrega los nombres de las variables que compon  
head(datos) # muestra las primeras filas presentes en el marco
```

2. sjPlot

```
sjPlot::view_df(datos) # Encontrar variables
```



Paso 4: ¡Análisis descriptivo!

¿Dónde?



Descargar el zip de la sesión 4 del sitio del curso

1. Recursos de la práctica



- Datos: *Encuesta de Caracterización Socioeconómica (CASEN)* (2020).
- Para ello, deben dirigirse al **siguiente enlace** y descargar los zip.
- **Libro de códigos** antes de trabajar una base de datos.



Algunas recomendaciones para el análisis

Plan de análisis



- **1. Objetivo de la investigación o reporte**
- En general, *siempre* deben tener "bajo la manga" un análisis descriptivo de sus datos
- Recomendación: tener un script propio para hacer esos análisis. También sirve un **.Rmd** (mejor, pues tendrán un pre-informe listo)

Plan de análisis



- **2. Preguntas claves**

- ¿Cuál es el nivel de medición de mis variables?
- ¿Qué variables quiero relacionar? ¿De qué forma?
- ¿Qué es lo *relevante* en mi análisis?



Trade off: los datos y las funciones

Tipos de variables



- Las que principalmente nos interesarán son los "vectores"
- En nuestro "idioma", son las **variables**.
- En general, una combinación de vectores da origen a una matriz (o **data frame** o "base de datos")

Tipos de datos: columnas



Relación entre clase y nivel de medición de la variable

- numeric
- character
- factor
- hay varios más...

Las funciones



- Existen **muchas funciones** para analizar datos
- Lo más importante es que ustedes se decidan por una que esté ad hoc a sus datos y no produzca conflictos con otros paquetes
- Eviten el "shopping" de paquetes: eso distingue entre buenos usuarios de R y aquellos que no



```
pacman::p_load(todos,  
               los,  
               paquetes,  
               quepillé,  
               en,  
               internet,  
               profenomepongaun1  
               )
```

```
# Error "su código no es eficiente"
```

Las funciones: sjPlot y sjmisc



- Una solución buena y eficiente es **sjPlot** y **sjmisc**.
- Ya sabemos que son "buenos" amigos con otras funciones de tidyverse.
- Además, como verán (y se asombrarán), no solo permiten hacer análisis descriptivo sino que modelamiento.

En síntesis



Tendencia central

Dispersión

Proporciones

Visualización

¿Y eso era?



¡Un segundo!

- Aprovecha estas semanas para repasar los prácticos
- La sesión N°7 requiere de un manejo avanzado de los prácticos

. box-2[- Una buena forma de saber "cómo están" para esta segunda parte del curso es tomar nota de **cuánto** les cuesta la tarea]

¡Ahora si que si! Nos vemos después del receso







Análisis descriptivo

Sesión N° 6

13 septiembre 2021

Análisis de datos estadísticos en R

Profesora Valentina Andrade de la Horra
Ayudantes Dafne Jaime y Nicolás Godoy