

Análisis de Datos Categóricos

Ayudantía 1 - parte 2

Felipe Olivares

Explorando RStudio

Inicialmente es importante identificar las principales herramientas que usaremos en la consola de RStudio. A continuación se encuentra una notación con sus nombres y funciones, enumeradas en la siguiente figura.

Explorando RStudio

The screenshot shows the RStudio desktop environment with the following components highlighted by numbered circles:

- 1**: File menu
- 2**: Source editor (script file)
- 3**: Run button
- 4**: Environment pane
- 5**: History pane
- 6**: Files pane
- 7**: Plots pane
- 8**: Packages pane
- 9**: Help pane

The source editor contains the following R code:

```
1 # this is my script
2 ...
3 ...
4 ...
5 ...
6 ...
7 I'm going to generate 100 pairs of *random numbers* and make a
8 scatterplot.
9 ...
10 [r]
11 # this is an R chunk
12 x <- rnorm(100)
13 y <- rnorm(100)
14 plot(x, y)
15
```

The Environment pane shows the following values:

Variable	Class	Length	Values
x	num	[1:100]	8.945 -8.923 -1.471 -8.184 8.9...
y	num	[1:100]	0.8148 0.2578 -0.122 -1.6914 1...

The Plots pane shows a scatter plot of y versus x, with data points represented by open circles. The x-axis ranges from -2 to 2, and the y-axis ranges from -2 to 3.

Explorando RStudio

- ➊ **New file:** Creación de un nuevo documento o directorio de trabajo en formatos como R Script, R Notebook y Rmarkdown. La ayudantía será realizada principalmente mediante documentos R Script.
- ➋ **Save:** Guarda los cambios del documento abierto en el panel “Edit/source.”
- ➌ **Run:** Ejecuta el código seleccionado en el panel “Edit/source.”
- ➍ **Import Dataset:** Importa el Dataset del directorio indicado.
- ➎ **Clear:** Limpia los objetos creados en el “Environment.”

Explorando RStudio

- ⑥ **Files:** Documentos del directorio de trabajo (carpeta), dentro del disco del computador.
- ⑦ **Plots:** Visualizador de gráficos.
- ⑧ **Packages:** Buscador de descarga de librerías y repositorios R.
- ⑨ **Help:** Interface de ayuda respecto a la sintaxis y librerías de R.

Librerías

Una librería de R es un conjunto de funciones y bases de datos encapsuladas en un objeto. Estas funciones nos permiten **utilizar distintas técnicas de estimación estadísticas, descargar bases de datos, graficar funciones, manipular bases de datos y obtener tablas de resultados.**

Librerías

Hay al menos **3 opciones** para instalar una librería en RStudio.

- Ejecutar el comando en el panel “Edit/source”:
`install.packages(“nombre de la librería de interés”)`.
- En la pestaña “Tools” de RStudio: Tools -> Install Packages -> escribir el nombre de la librería.
- En el panel “Assistance”: Packages -> Install -> escribir el nombre de la librería.

Una vez instalado la librería, la puedes cargar en el panel de “Edit/source” con el comando “**library(nombre)**”

Objetos

- R es un lenguaje de programación orientada a objetos. De este modo, su sintaxis define como objeto a **una estructura que tiene asociadas atributos**. Desde el punto de vista estadístico, las bases de datos son objetos donde las variables corresponden a sus atributos.
- Parte de la flexibilidad y virtud de este lenguaje de programación es que es posible guardar en objetos distintos tipos de estructuras. Un objeto puede ser un **vector (secuencia de atributos numéricos)**, un **factor (secuencia de atributos categóricos)**, un **conjunto de vectores y factores (base de datos)**, una **función matemática**, un **modelo estadístico**, una **tabla de resultados**, y un **gráfico entre otros**.
- A lo largo del curso estaremos revisando y manipulando distintos tipos de objetos.

Objetos

La asignación de un valor a un objeto mediante el código en el panel “Edit/source”, y su creación dentro del panel de “Environment”, se realiza con los signos `<-` o con `=`, de la siguiente forma.

```
objeto1 <- 1  
objeto2 <- "Urbano"  
  
objeto3 = 1  
objeto4 = "Urbano"
```

Note que a la izquierda queda el objeto creado bajo el nombre escogido, y a la derecha sus atributos.

Objetos

Cuando nuestros objetos contienen **más de 1 atributo**, sean estos numéricos o categóricos, la creación del objeto deberá seguir la siguiente sintaxis que **separa los atributos mediante comas** y **engloba los componentes entre paréntesis**.

```
vector <- c(1,2,3,4,5,6)
factor <- c("Artes Visuales","Música","Teatro","Danza")
```

Así, nuestros objetos serán las variables, que podrán ser variables continuas numéricas (**vectores**) o categóricas (**factores**).

Objetos - Ejemplo 1

A continuación se realizaremos un breve ejercicio de creación de objetos (variables) y una pequeña base de datos (data.frame).

```
# Creación de objetos variables
```

```
id <- c(50010,50020,50030,50040,50050,50060,50070)
```

```
ingreso <- c(500, 500, 250, 1000, 500, 800, 500)
```

```
edad <- c(19,31,22,57,81,18,47)
```

```
zona <- c(rep("urbano", 4), rep("rural", 3))
```

```
sexo <- c(rep("hombre", 3), rep("mujer", 4))
```

```
gusto <- c("Teatro","Cine","Teatro","Musica","Cine","Musica")
```

Objetos - Ejemplo 1

```
# Creación base de datos utilizando "data.frame"
```

```
datos <- data.frame(ID = id,  
                    ING = ingreso,  
                    ED = edad,  
                    ZON = zona,  
                    SEX = sexo,  
                    GUS = gusto)
```

Objetos - Comandos

Comandos frecuentes de R base

Comandos		Comandos	
read.csv().	carga como data.frame un data file existente	class()	identifica qué clase de objeto es
data()	carga una base de datos en el "Environment"	levels()	identifica los niveles de una variable factor
View()	abre la planilla de variables de un data.frame	table()	tabula frecuencias de valores de variable
names()	lista de variables en un data.frame	hist()	crea un histograma de la variable
dim()	no. de filas y columnas (f,c) en un data.frame	barplot()	crea unbarplot de la variable
str()	muestra la estructura interna de una variable	rm()	remueve permanentemente un objeto
summary()	calcula estadísticos descriptivos básicos	example()	muestra ejemplos de un comando
sum()	suma de todos los valores de una variable	help()	ayuda para el uso de un comando

Objetos - Comandos

Exploremos ahora alguno de los comandos más comunes en la manipulación de objetos utilizando nuestro data.frame (“datos”). Note que al referirnos a **variables específicas** deberemos señalar primero el **nombre del data frame (“datos”)**, agregar luego el **signo “\$”**, para finalmente agregar el nombre de la variable de interés (“ED”)

```
names(datos)
```

```
## [1] "ID" "ING" "ED" "ZON" "SEX" "GUS"
```

Objetos - Comandos

```
dim(datos)
```

```
## [1] 7 6
```

```
str(datos$ED)
```

```
## num [1:7] 19 31 22 57 81 18 47
```

```
summary (datos$ING)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  250.0   500.0   500.0   578.6   650.0  1000.0
```

Objetos - Comandos

```
class(datos$ZON)
```

```
## [1] "character"
```

```
levels(datos$GUS)
```

```
## NULL
```

```
table(datos$SEX)
```

```
##
```

```
## hombre  mujer
```

```
##      3      4
```


Objetos - Operadores

Tabla de operadores R (base)

Aritméticos		Comparativos		Lógicos	
+	Adición	==	Igual a	&	Y (arg)
-	Substracción	!=	Diferente de	!	No (arg)
*	Multiplicación	>	Mayor que		O (arg)
/	División	<	Menor que	is.na(x)	Excluir (arg)
^	Potencia	<=	Menor o igual que		
()	Paréntesis	>=	Mayor o igual que		

Objetos - Ejemplo 2

Los operadores constituyen la base de la sintaxis matemática de R, lo que en términos simples vuelve al software en una calculadora de alto potencial. Las operaciones simples y lógicas pueden ser también resueltas en el panel “Edit/source”.

```
#Operadores Aritméticos
```

```
##Cálculo de promedio curso con 4 notas
```

```
(4.5+5.6+6.4+7.0)/4
```

```
## [1] 5.875
```

```
promedio <- (4.5+5.6+6.4+7.0)/4
```

```
promedio
```

```
## [1] 5.875
```

Objetos - Ejemplo 2

```
#Operadores Aritméticos  
##Cálculo de cuadrado de binomio
```

```
c <- 13.457  
d <- (6.72^3)  
cbinomio1 <- (c+d)^2  
cbinomio2 <- (c^2)+(2*c*d)+(d^2)  
cbinomio1
```

```
## [1] 100439.2
```

```
cbinomio2
```

```
## [1] 100439.2
```

Objetos - Ejemplo 2

#Operadores Comparativo-Lógicos

```
cbinomio1 > cbinomio2
```

```
## [1] FALSE
```

```
cbinomio1 < cbinomio2
```

```
## [1] FALSE
```

```
cbinomio1 != cbinomio2
```

```
## [1] FALSE
```

Objetos - Funciones

Tabla de funciones R (base)

Matemáticas		Estadísticas	
<code>sqrt(x)</code>	Raíz cuadrada de x	<code>mean(x)</code>	Media
<code>exp(x)</code>	Exponencial de x	<code>sd(x)</code>	Desviación estándar
<code>log(x)</code>	Logaritmo natural de x	<code>var(x)</code>	Varianza
<code>log10(x)</code>	Logaritmo base 10 de x	<code>median(x)</code>	Mediana
<code>length(x)</code>	Número de elementos de x	<code>quantile(x,p)</code>	Quantiles
<code>sum(x)</code>	Suma de elementos de x	<code>cor(x,y)</code>	Correlación
<code>prod(x)</code>	Producto de los elementos de x	<code>max(x)</code>	Máximo
<code>sin(x)</code>	Seno de x	<code>min(x)</code>	Mínimo
<code>cos(x)</code>	Coseno de x	<code>range(x)</code>	Rango
<code>tan(x)</code>	Tangente de x	<code>sort(x)</code>	Ordenar elementos

Objetos - Funciones

Los funciones de cálculo matemático son cruciales para la estimación y análisis de datos estadísticos, y por ello es crucial también conocer su uso en R.

Objetos - Ejemplo 3

```
#Funciones Matemáticas
```

```
sqrt(c)
```

```
## [1] 3.668378
```

```
exp(d)
```

```
## [1] 6.207764e+131
```

```
log(cbinomio1)
```

```
## [1] 11.51731
```

```
log10(cbinomio2)
```

```
## [1] 5.001903
```

Objetos - Ejemplo 3

```
#Funciones Estadísticas
```

```
mean(datos$ING)
```

```
## [1] 578.5714
```

```
max(datos$ED)
```

```
## [1] 81
```

```
min(datos$ED)
```

```
## [1] 18
```

```
range(datos$ED)
```

```
## [1] 18 81
```


Consejos

- 1) El aprendizaje de R es como la práctica de un instrumento musical, entre **mayor dedicación y frecuencia de práctica mejor será la adaptación al programa, su uso y fluidez**. Recomendamos explorar RStudio y el lenguaje R todas las semanas para apoyar el aprendizaje progresivo.
- 2) Al ser parte de una **comunidad académica internacional**, muchas de las dudas que puedas tener en el camino son y han sido discutidas en múltiples recursos de internet como **blogs, foros, y tutoriales** que dan soluciones múltiples a un mismo problema. A modo de apoyo es recomendado consultar los recursos disponibles en internet además del material de ayudantías y consultas del curso.