1 Introduccion

referencias

- A handbook of statistical analysis using R
- Instalación

```
http://CRAN.R-project.org
```

- Es un interprete de comandos

```
R> x <- sqrt(25) + 2
```

paquetes

- Son colecciones de funciones y datos
- Necesitan ser activadas
- La instalacion trae un conjunto base
- Instalaciones nuevas

```
install.packages("sandwich")
```

```
library(sandwich)
```

```
library(HSAUR2)
```

```
library(help=sandwich)
```

Help & documentacion

- Ayuda

```
help("mean")
?mean

help(package = "sandwich")

vignette("sandwich", package = "sandwich")
```

Objetos de datos

- Lista Forbes

```
data("Forbes2000", package = "HSAUR2")
ls()
 print(Forbes2000)
  help("Forbes2000")
?Forbes2000
```

- Caracteristicas basicas

```
class(Forbes2000)
dim(Forbes2000)
nrow(Forbes2000)
ncol(Forbes2000)
```

```
names(Forbes2000)
class(Forbes2000[,"rank"])
length(Forbes2000[,"rank"])
```

Tipos de objetos

- Vectores

```
> 1:3
> c(1,2,3)
> seq(from = 1, to = 3, by = 1)
> class(Forbes2000[,"name"])
> length(Forbes2000[,"name"])
> Forbes2000[,"name"][1]
```

- Valores nominales: factores

```
class(Forbes2000[,"category"])
nlevels(Forbes2000[,"category"])
levels(Forbes2000[,"category"])
table(Forbes2000[,"category"])
```

- Numericos

```
R> median(Forbes2000[,"sales"])

R> mean(Forbes2000[,"sales"])

R> range(Forbes2000[,"sales"])

R> summary(Forbes2000[,"sales"])
```

Lectura de datos

- Leer desde fichero csv

```
csvForbes2000 <- read.table("Forbes2000.csv”, header =
TRUE, sep = ",", row.names = 1)
class(csvForbes2000[,"name"])
```

- Un poco mejor:

```
> csvForbes2000 <- read.table("Forbes2000.csv",
+ header = TRUE, sep = ",", row.names = 1,
+ colClasses = c("character", "integer", "character",
+ "factor", "factor", "numeric", "numeric", "numeric",
+ "numeric"))
> class(csvForbes2000[,"name"])

> all.equal(csvForbes2000, Forbes2000)
```

escritura

- Escribir csv

```
 > write.table(Forbes2000, file = "Forbes2000.csv", sep = ",",
+ col.names = NA)
```

Manipulaciones basicas

```
 >companies <- Forbes2000[,"name"]
 >companies[1]
 >companies[1:3]
 >companies[-(4:2000)]   # los que no estan en el vector
 >Forbes2000[1:3, c("name", "sales", "profits", "assets")]
> companies <- Forbes2000$name
 >order_sales <- order(Forbes2000$sales) # orden por ventas
 >companies[order_sales[1:3]]   #inferiores
> Forbes2000[order_sales[c(2000, 1999, 1998)], c("name",
"sales", "profits", "assets")]
> Forbes2000[Forbes2000$assets > 1000, c("name", "sales",
"profits", "assets")]
 > table(Forbes2000$assets > 1000)
```

- Missing values

```
> na_profits <- is.na(Forbes2000$profits)
> table(na_profits)
> Forbes2000[na_profits, c("name", "sales", "profits", "assets")]

>table(complete.cases(Forbes2000))
```

- Subseting

```
> UKcomp <- subset(Forbes2000, country == "United Kingdom")
> dim(UKcomp)
```

- Estadisticas sumarias

```
summary(Forbes2000)

lapply(Forbes2000, summary)

mprofits <- tapply(Forbes2000$profits,
+ Forbes2000$category, median, na.rm = TRUE)

median(Forbes2000$profits)

median(Forbes2000$profits, na.rm = TRUE)

rev(sort(mprofits))[1:3]
```

Graficos simples

- Histogramas

```
> layout(matrix(1:2, nrow = 2))
> hist(Forbes2000$marketvalue)
> hist(log(Forbes2000$marketvalue))
```

- Formulas

```
> fm <- marketvalue ~ sales
> class(fm)
```

- Scatter plot

```
>plot(log(marketvalue) ~ log(sales), data = Forbes2000, pch = ".")
```

```
>plot(log(marketvalue) ~ log(sales), data = Forbes2000, col=rgb(0,0,0,0.1),  pch = 16)
```

- Boxplot

```
> tmp <- subset(Forbes2000,
+ country %in% c("United Kingdom", "Germany",
+ "India", "Turkey"))
> tmp$country <- tmp$country[,drop = TRUE]
> plot(log(marketvalue) ~ country, data = tmp, ylab =
"log(marketvalue)", varwidth = TRUE)
```