

Clase 4.0

Análisis

Marcos Rosetti y Luis Pacheco-Cobos

Estadística y Manejo de Datos con R (EMDR) — Virtual

Estadística descriptiva

Estadística descriptiva

- Medidas de tendencia central.

```
# Media  
mean(mtcars$mpg) # emplea na.rm = T si hay valores faltantes
```

```
## [1] 20.09062
```

```
# Mediana  
median(mtcars$mpg)
```

```
## [1] 19.2
```

```
# Moda  
getmode <- function(v) {  
  uniqv <- unique(v)  
  uniqv[which.max(tabulate(match(v, uniqv)))]  
}  
getmode(mtcars$mpg)
```

```
## [1] 21
```

Estadística descriptiva

- Medidas de dispersión.

```
# Desviación estándar  
sd(mtcars$mpg)
```

```
## [1] 6.026948
```

```
# Cuantiles  
quantile(mtcars$mpg)
```

```
##      0%      25%      50%      75%     100%  
## 10.400 15.425 19.200 22.800 33.900
```

```
# Rango intercuantil  
IQR(mtcars$mpg)
```

```
## [1] 7.375
```

```
# Rango total  
range(mtcars$mpg)
```

```
## [1] 10.4 33.9
```

Estadística descriptiva

- Tamaños de muestra.

```
length(mtcars$mpg) # Número de observaciones
```

```
## [1] 32
```

```
length(mtcars) # Si es un df, número de variables (columnas)
```

```
## [1] 11
```

```
dim(mtcars) # Si es matriz o df, núm. de observaciones x núm. de variables
```

```
## [1] 32 11
```

Estadística descriptiva

- `dplyr::group_by` y `dplyr::summarize`

```
library(dplyr)
library(magrittr)

mtcars %>%
  group_by(cyl) %>%
  summarise(n = n(), mean_mpg = mean(mpg), sd_mpg = sd(mpg))
```

```
## # A tibble: 3 × 4
##   cyl     n mean_mpg sd_mpg
##   <dbl> <int>   <dbl> <dbl>
## 1     4    11    26.7   4.51
## 2     6     7    19.7   1.45
## 3     8    14    15.1   2.56
```

Estadística descriptiva

- `dplyr::summarize_all`

```
library(dplyr)
library(magrittr)

mtcars %>%
  group_by(cyl) %>%
  summarise_all(list(~mean(.)))
```

```
## # A tibble: 3 × 11
##   cyl  mpg  disp  hp  drat   wt  qsec  vs  am  gear  carb
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1     4  26.7  105.  82.6  4.07  2.29  19.1  0.909  0.727  4.09  1.55
## 2     6  19.7  183.  122.   3.59  3.12  18.0  0.571  0.429  3.86  3.43
## 3     8  15.1  353.  209.   3.23  4.00  16.8  0      0.143  3.29  3.5
```

Licencia CC BY



Estadística y Manejo de Datos con R (EMDR) por Marcos F. Rosetti S. y Luis Pacheco-Cobos se distribuye bajo una [Licencia Creative Commons Atribución 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/).