

Curso Introducción a R en investigación biomédica

Solución ejercicios prácticos Modulo3 - Estadística descriptiva con R

Unidad de Apoyo Metodológico IIS Biogipuzkoa - Lore Zumeta & Jone Renteria

Noviembre 2025



1. Explora el dataset

Carga los datos descargados en formato CSV. Usa `str()` y `summary()` para inspeccionar la estructura y obtener un resumen de las variables.

```
# Cargar datos desde CSV (ajusta la ruta al archivo descargado)
diabetes <- read.csv("diabetes.csv")

# Inspección básica
str(diabetes)      # estructura del dataset

## 'data.frame':   403 obs. of  19 variables:
## $ id      : int  1000 1001 1002 1003 1005 1008 1011 1015 1016 1022 ...
## $ chol    : int  203 165 228 78 249 248 195 227 177 263 ...
## $ stab.glu: int  82 97 92 93 90 94 92 75 87 89 ...
## $ hdl    : int  56 24 37 12 28 69 41 44 49 40 ...
## $ ratio   : num  3.6 6.9 6.2 6.5 8.9 ...
## $ glyhb  : num  4.31 4.44 4.64 4.63 7.72 ...
## $ location: chr  "Buckingham" "Buckingham" "Buckingham" "Buckingham" ...
## $ age     : int  46 29 58 67 64 34 30 37 45 55 ...
## $ gender  : chr  "female" "female" "female" "male" ...
## $ height  : int  62 64 61 67 68 71 69 59 69 63 ...
## $ weight  : int  121 218 256 119 183 190 191 170 166 202 ...
## $ frame   : chr  "medium" "large" "large" "large" ...
## $ bp.1s   : int  118 112 190 110 138 132 161 NA 160 108 ...
## $ bp.1d   : int  59 68 92 50 80 86 112 NA 80 72 ...
## $ bp.2s   : int  NA NA 185 NA NA NA 161 NA 128 NA ...
## $ bp.2d   : int  NA NA 92 NA NA NA 112 NA 86 NA ...
## $ waist   : int  29 46 49 33 44 36 46 34 34 45 ...
## $ hip     : int  38 48 57 38 41 42 49 39 40 50 ...
## $ time.ppn: int  720 360 180 480 300 195 720 1020 300 240 ...

summary(diabetes)  # resumen estadístico

##      id          chol         stab.glu        hdl
## Min.   : 1000   Min.   : 78.0   Min.   : 48.0   Min.   : 12.00
## 1st Qu.: 4792   1st Qu.:179.0   1st Qu.: 81.0   1st Qu.: 38.00
```

2. Tablas de frecuencia

```

## Median :15766   Median :204.0   Median : 89.0   Median : 46.00
## Mean    :15978   Mean   :207.8    Mean   :106.7    Mean   : 50.45
## 3rd Qu.:20336   3rd Qu.:230.0   3rd Qu.:106.0   3rd Qu.: 59.00
## Max.    :41756   Max.   :443.0    Max.   :385.0    Max.   :120.00
## NA's    :1        NA's   :1       NA's   :1       NA's   :1
##      ratio      glyhb      location      age
## Min.   : 1.500   Min.   : 2.68   Length:403     Min.   :19.00
## 1st Qu.: 3.200   1st Qu.: 4.38   Class  :character 1st Qu.:34.00
## Median : 4.200   Median : 4.84   Mode   :character  Median :45.00
## Mean   : 4.522   Mean   : 5.59   NA's   :1       Mean   :46.85
## 3rd Qu.: 5.400   3rd Qu.: 5.60   NA's   :1       3rd Qu.:60.00
## Max.   :19.300   Max.   :16.11   NA's   :1       Max.   :92.00
## NA's   :1        NA's   :13    NA's   :1
##      gender      height      weight      frame
## Length:403     Min.   :52.00   Min.   : 99.0   Length:403
## Class  :character 1st Qu.:63.00   1st Qu.:151.0  Class  :character
## Mode   :character Median :66.00    Median :172.5  Mode   :character
##               Mean   :66.02    Mean   :177.6
##               3rd Qu.:69.00   3rd Qu.:200.0
##               Max.   :76.00    Max.   :325.0
##               NA's   :5       NA's   :1
##      bp.1s      bp.1d      bp.2s      bp.2d
## Min.   : 90.0   Min.   :48.00   Min.   :110.0   Min.   : 60.00
## 1st Qu.:121.2   1st Qu.:75.00   1st Qu.:138.0  1st Qu.: 84.00
## Median :136.0   Median : 82.00   Median :149.0   Median : 92.00
## Mean   :136.9   Mean   : 83.32   Mean   :152.4    Mean   : 92.52
## 3rd Qu.:146.8   3rd Qu.:90.00   3rd Qu.:161.0   3rd Qu.:100.00
## Max.   :250.0   Max.   :124.00   Max.   :238.0   Max.   :124.00
## NA's   :5       NA's   :5       NA's   :262    NA's   :262
##      waist      hip      time.ppn
## Min.   :26.0    Min.   :30.00   Min.   : 5.0
## 1st Qu.:33.0   1st Qu.:39.00   1st Qu.: 90.0
## Median :37.0   Median :42.00   Median :240.0
## Mean   :37.9   Mean   :43.04   Mean   :341.2
## 3rd Qu.:41.0   3rd Qu.:46.00   3rd Qu.:517.5
## Max.   :56.0   Max.   :64.00   Max.   :1560.0
## NA's   :2       NA's   :2       NA's   :3

```

2. Tablas de frecuencia

Crea una tabla de frecuencia simple con la variable LOCATION, y una tabla cruzada entre las variables GENDER y FRAME (Frame: pequeño, mediano, grande, constitución corporal).

```
# Tabla de frecuencia simple con LOCATION
```

```
table(diabetes$location)
```

```
##
## Buckingham    Louisa
##          200      203
```

```
# Tabla cruzada entre GENDER y FRAME
```

```
table(diabetes$gender, diabetes$frame)
```

```
##
##           large medium small
```

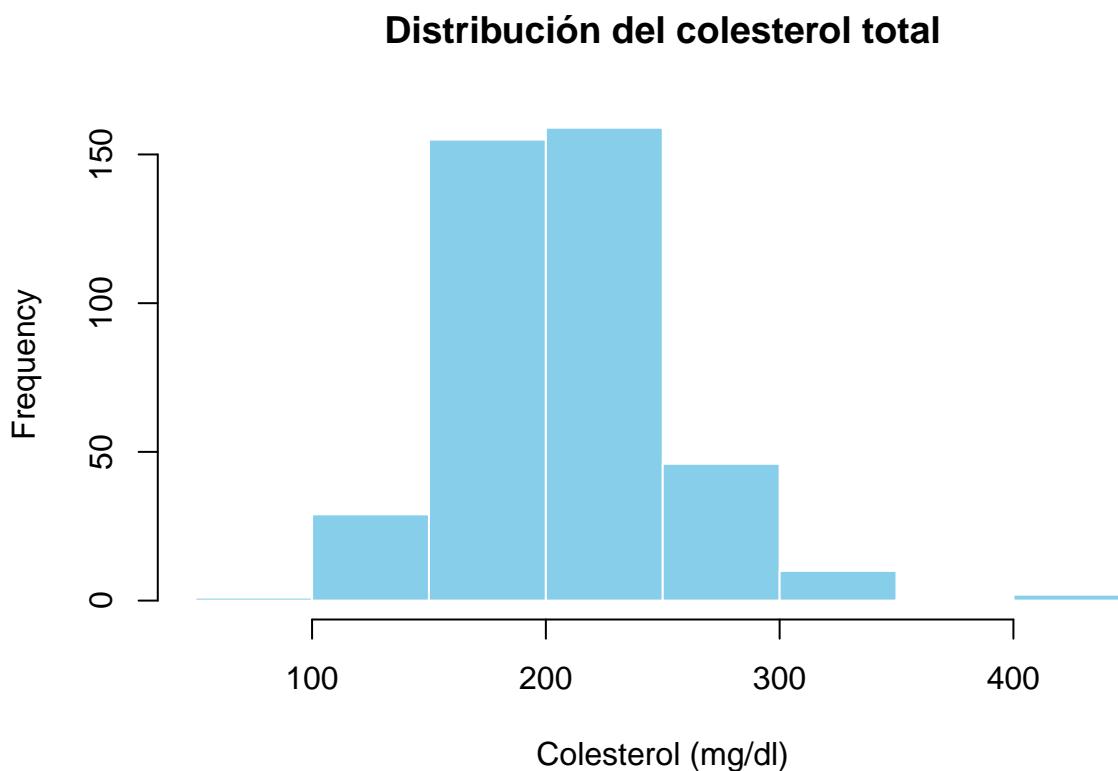
3. Histograma colesterol total

```
##   female    7    42    116    69
##   male     5    61    68    35
```

3. Histograma colesterol total

Haz un histograma del colesterol total (chol) usando hist(). Personaliza con color skyblue, título “Distribución del colesterol total” y etiqueta del eje X “Colesterol (mg/dl)”.

```
hist(diabetes$chol,
  col = "skyblue",
  main = "Distribución del colesterol total",
  xlab = "Colesterol (mg/dl)",
  border = "white")
```



Guarda el gráfico en archivo PNG con nombre hist_colesterol.png.

```
png("hist_colesterol.png", width = 800, height = 600)
hist(diabetes$chol,
  col = "skyblue",
  main = "Distribución del colesterol total",
  xlab = "Colesterol (mg/dl)",
  border = "white")
dev.off()

## pdf
## 2
```

4. Varios gráficos en una misma ventana

4. Varios gráficos en una misma ventana

Usa `par(mfrow=c(2,2))`. Incluye histograma (el anterior), boxplot, barplot y scatterplot juntos.

```
par(mfrow=c(2,2))
```

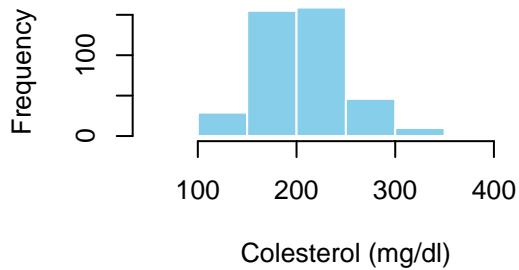
```
# a. Histograma (copia del anterior)
hist(diabetes$chol,
      col = "skyblue",
      main = "Distribución del colesterol total",
      xlab = "Colesterol (mg/dl)",
      border = "white")

# b. Boxplot del HDL
boxplot(diabetes$hdl,
        col = "lightgreen",
        main = "Boxplot de HDL",
        ylab = "HDL (mg/dl)")

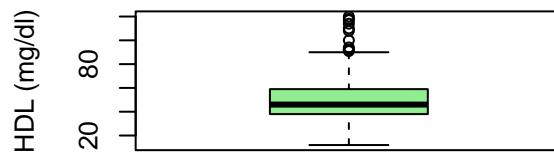
# c. Barplot genero
barplot(table(diabetes$gender),
        col = c("tomato", "lightblue"),
        main = "Frecuencia inclusión en el estudio según sexo",
        ylab = "Número de individuos")

# d. Scatterplot de peso vs colesterol
plot(diabetes$weight, diabetes$chol,
      col = "darkorange",
      pch = 19,
      main = "Relación entre peso y colesterol",
      xlab = "Peso (lbs)",
      ylab = "Colesterol (mg/dl)")
```

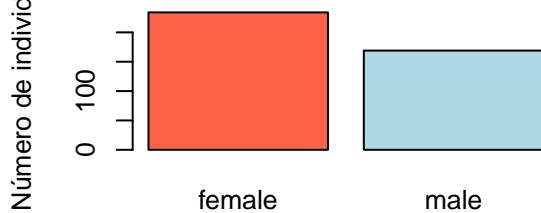
Distribución del colesterol total



Boxplot de HDL



Frecuencia inclusión en el estudio según



Relación entre peso y colesterol

