IdiPAZ
Instituto de Investigación
Hospital Universitario La Paz

SaludMadrid
Hospital Universitario La Paz
Hospital de Cantoblanco
Hospital Carlos III
Comunidad de Madrid

ANÁLISIS DE DATOS EN INVESTIGACIÓN BIOMÉDICA

MÉTODOS ESTADÍSTICOS

Dra. Mariana Díaz-Almirón

Investigación Biomédica

1. Recordar definición de variables
2. Depuración de los datos
3. Toma de contacto con la consola de R
4. Estadística descriptiva

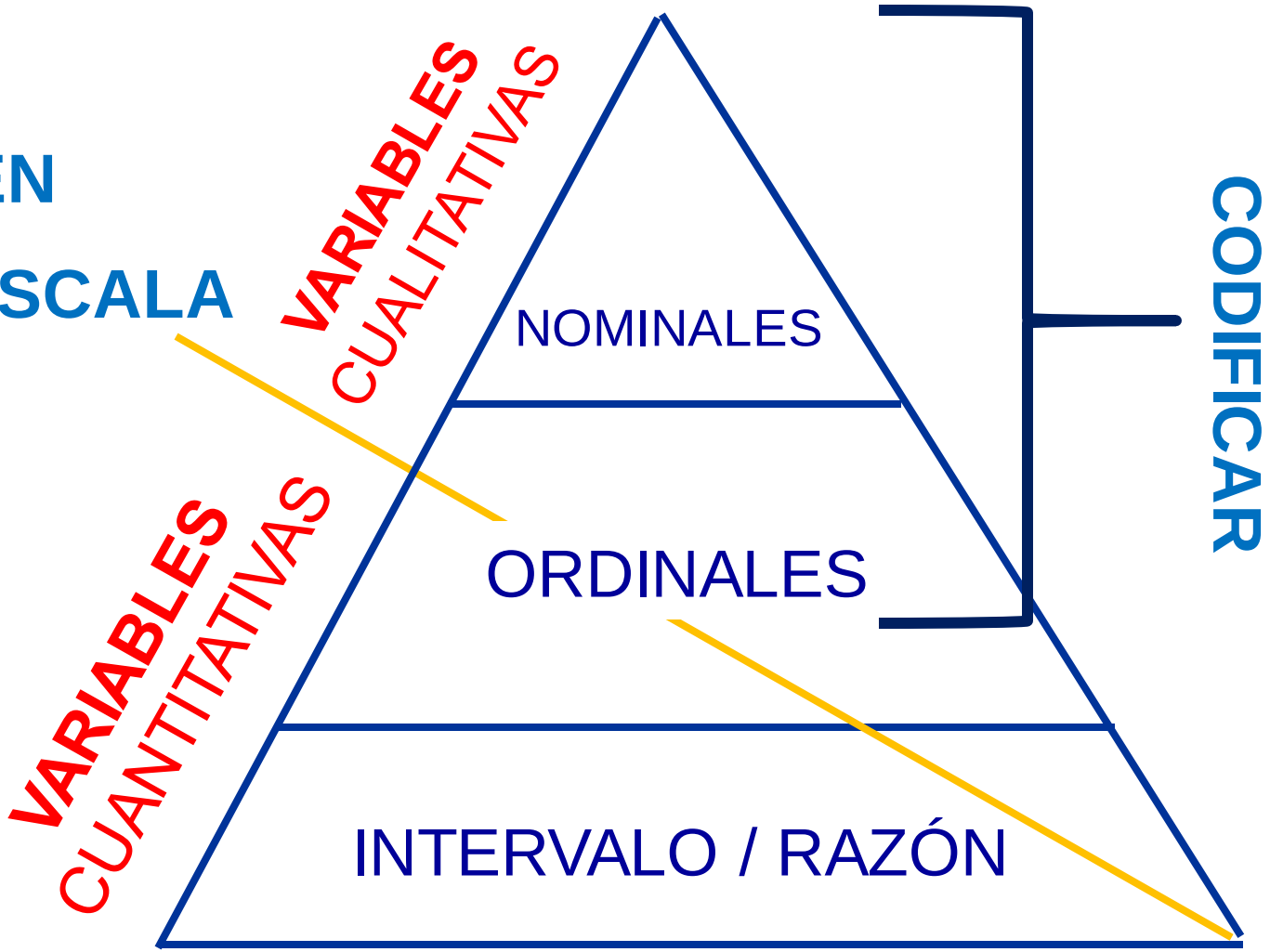
Investigación Biomédica

1. Recordar definición de variables
2. Depuración de los datos
3. Toma de contacto con la consola de R
4. Estadística descriptiva

Recordar definición de variables

RESUMEN

TIPO ~ ESCALA



Investigación Biomédica

1. Recordar definición de variables
2. Depuración de los datos
   - Detección de errores de trascripción
   - Detección de “outliers”,…
3. Toma de contacto con la consola de R
4. Estadística descriptiva

Investigación Biomédica

1. Recordar definición de variables
2. Depuración de los datos
3. Toma de contacto con la consola de R
4. Estadística descriptiva

- CREAR UN CONJUNTO DE DATOS
- DEFINIR LAS VARIABLES
- CREAR (CALCULAR) UNA VARIABLE
- CREAR UNA VARIABLE A PARTIR DE OTRA
- SELECCIONAR CASOS

VARIABLES

Variable	Significado	Códigos	Escala
DATOS PERSONALES			
ID	Nº de Identificación		
SEXO	Sexo	0=Mujer 1=Hombre	Nominal
EDAD	Edad		Continua
TRATAMIENTO	Tratamiento aleatorizado	0=Referencia 1=Nuevo	Nominal
TALLA	Talla basal		Continua
ACTIVIDAD	actividad física	1 'leve' 2 'moderada' 3 'severa'	Ordinal
MOTIVACION	Score de Motivación (1-12, siendo 12 lo mejor)		Ordinal
ANTECEDENTES PERSONALES			
HTA	Hipertensión	0=No 1=Sí	Nominal
DISLIPEMIA	Dislipemia	0=No 1=Sí	Nominal
GLUCEMIA	Alteracion de la Glucemia	0=No 1=Sí	Nominal
DE	Disfuncion Endotelial	0=No 1=Sí	Nominal
LDLox	Altos niveles de LDL Oxidada	0=No 1=Sí	Nominal
FUMA	Fuma	0=No 1=Sí	Nominal
CCI	Índice Cintura-Cadera Elevado	0=No 1=Sí	Nominal
DATOS BASALES			
P0	Peso basal		Continua
TG0	Triglicéridos basal		Continua
ALBUMINA0	Albúmina basal		Continua
MGRASA0	Masa Grasa basal (kg)		Continua
GLUCOSA0	Glucosa basal		Continua
INSULINA0	Insulina basal		Continua
AcURICO0	Ácido Úrico basal		Continua

DATOS EN SUCESIVAS CONSULTAS			
P1	Peso en la segunda consulta		Continua
P2	Peso final		Continua
SEGUIMIENTO	Tiempo desde P0 a P2		Continua
TCAMBIO	Tiempo hasta perder al menos 10K		Continua
TG2	Triglicéridos final		Continua
CCI2	Índice Cintura-Cadera Elevado final	0=No 1=Sí	Nominal
ALBUMINA2	Albúmina final		Continua
MGRASA2	Masa Grasa final (kg)		Continua
GLUCOSA2	Glucosa final		Continua
INSULINA2	Insulina final		Continua
AcURICO2	Ácido Úrico final		Continua
SATISFFINAL	Satisfacción final		Ordinal

Usamos R (R Gui / Rstudio / TinnR) o un compilador de R que ejecute el código:

https://rdrr.io/snippets/

Para más información del software, un manual de R se puede descargar aquí:

- https://www.dropbox.com/s/293cw299i0lb32x/R%20para%20principiantes.pdf?dl=0
- https://javieralvarezliebana.es/curso-intro-R/

- CREAR UN CONJUNTO DE DATOS
- CREAR (CALCULAR) UNA VARIABLE
- DEFINIR LAS VARIABLES
- CREAR UNA VARIABLE A PARTIR DE OTRA
- SELECCIONAR CASOS

DATOS

ESTRUCTURA

```
datos = data.frame(
        ID=c(1, 2, 3, 4, …),
        F_INI=c(42917, 42989, …),
        F_FIN=c(43049, 43116, …),
        F_10K=c(NA, 43116, NA…),
        SEXO=c(0, 1, 0, …),
        EDAD=c(49, 48, 22, …),
        TALLA=c(165.83, 190.94, …),
        ACTIVIDAD=c(1, 1, 1…),
        TRATAMIENTO=c(1, 0, 0, …),
        MOTIVACION=c(4, 4, 4, 9…),
        HTA=c(1, 0, 0…),
        DISLIPEMIA=c(1, 1, …),
        GLUCEMIA=c(0, 1, 1…),
        DE=c(0, 1, 1…0),
        LDLox=c(1, 1, 1, …),
        FUMA=c(1, 1, 1,…),
        CCI=c(1, 1, 1, …),
        TG0=c(82.61, 76.41, 77.81, …),
        TG2=c(76.52, 72.59, 72.63, …),
        ALBUMINA0=c(4.06, 4.75, 4.74,…),
        ALBUMINA2=c(2.06, 4.17, 4.3, …),
        MGRASA0=c(38.31, 36.52, 27.96, …),
        MGRASA2=c(26.84, 34.69, 25.25, …),
        GLUCOSA0=c(54.91, 95.52, 98.97,…),
        GLUCOSA2=c(45.93, 93.06, 95.9, …),
        INSULINA0=c(33.33, 7.53, 20.66, …),
        INSULINA2=c(22.71, 3.82, 19.54,…),
        AcURICO0=c(14.86, 3.68, …),
        AcURICO2=c(1.89, 1.02, 5.63, …),
        P0=c(97.53, 103.81, 103.91, …),
        P1=c(94.905, 89.31,…),
        P2=c(92.28, …),
        SATISFFINAL=c(5, 8, 5…)
)
head(datos)
```

DATOS

Sintaxis (copiar el texto entero y pegar en el compilador):

```
datos = data.frame(ID=c(1, 2, 3,4, 5, 6,7, 8, 9,10, 11, 12,13, 14, 15,16, 17, 18,19, 20, 21,22, 23, 24,25, 26, 27,28, 29, 30,31, 32, 33,34, 35, 36,37, 38, 39,40), F_INI=c(42917, 43045,
42862,42740, 42959, 42898,42998, 43059, 42785,42807, 43028, 42963,42984, 42946, 42761,42893, 42800, 42835,42999, 43052, 42956,42786, 42870, 42910,42940, 43003, 42772,
42907,43074, 42883, 43055,43055, 42853, 43010,43052, 42866, 43089,42932, 42751, 43028), F_FIN=c(43049, 43159,42973, 42864, 43073,43017, 43106, 43170,42908, 42928, 43144,
43084,43107, 43068,42893,43002, 42925, 42937,43116, 43176, 43080,42905, 42984, 43034,43060, 43125, 42895,43033, 43193, 42986,43166, 43179,42998,43114, 43167, 42979,
43218,43067, 42882, 43146), F_10K=c(NA, 43159,42941, NA, 43073, 43017,43106, NA, 42874, NA, 43144,43046, 43067, 43031, NA, 42971, 42925, NA, NA, 43176,43080, 42869, NA,
42996,43027, 43087, 42859,43033, 43157, NA, 43126,43179, 42998, 43114,43129, 42979, 43218,43067, 42850, NA), TRATAMIENTO=c(1, 0, 1,1, 0, 0,0, 0, 1,1, 0, 1,1, 1, 0,1, 0, 0,1, 0, 0,
1, 0, 1, 1,1, 1, 0,1, 1,1, 0,0, 0, 1,0, 0, 0,1, 0), P0=c(97.53, 109.32,111.09, 96.2,105.79, 105.4, 108.86, 116.02,117.34, 102.29, 103.21,113.95, 115.29, 115.71,99.64,110.92, 103.63, 103.67,
97.38, 115.36, 119.29,108.31, 87.49, 106.92, 105.1, 108.61,107.12, 100.2, 107.38, 108.12,115.68, 97.04, 94.78, 107.77, 106.14,107.61, 107.48, 109.04,110.43, 108.32), P1=c(94.91, 100.51,
95.64, 93.42, 99.76, 98.98, 102.88, 111.79,101.06, 97.95, 97.67, 93.9,99.62,95.37,96.08,89.5, 95.08, 99.22, 94.11, 108.6, 110.74,92.95,83.19,89.62,78.71,86.68,90.55,93.1, 92.21,
104.54, 97.81, 91.7, 87.59, 100.03, 90.45, 101.64,99.11,100.96, 91.65, 104.96), P2=c(92.28, 91.7, 80.18, 90.64, 93.72, 92.5 5, 96.91, 107.56, 84.78, 93.6, 92.13, 73.84, 83.94, 75.02, 92.51,
68.07, 86.53, 94.76, 90.84, 101.85, 102.19,77.58, 78.88,72.31,52.31,64.74,73.97,85.99,77.03,100.96, 79.94, 86.36, 80.4,92.28,74.76,95.67,90.73,92.88,72.86,101.59),
SATISFFINAL=c(5, 5, 10,2, 8, 8,8, 0, 10, 1, 7,7, 9, 7,0, 8, 5,3, 4, 8,8, 10,2, 10,9, 8, 9,7, 7, 4,8, 6, 8,8, 10,6 , 6,5,10, 1), SEXO=c(0,0, 1, 0,0, 0, 0,0, 0, 0,0, 1, 1,1, 1, 0,0, 1, 0,0, 1, 1,0, 0, 0,
0, 1, 1, 1, 0,0, 1, 1,0, 1, 0,0, 1, 0,1), EDAD=c(49, 35,51, 43, 18,20, 40, 19,30, 38, 38,50, 54, 36,45, 19, 37,18, 56, 41,61, 64, 21,30, 62, 45,33, 46, 64,19, 38, 46,52, 22, 25,29, 40, 48,61,
19), TALLA=c(165.83,199.32, 203.77, 165.16,192.65, 186.74, 182.15,188.41, 195.35, 176.2, 194.69,185.42, 193.11, 202.6, 184.93,204.87, 187.9, 196.47, 182.47,204.15, 183.36, 193.73,
166.79, 177.4, 193.01,175.82, 178.32, 186.89,183.69, 195.64, 175.93,185.74, 185.76, 175.33,196.01, 177.52, 180.92,204.52, 193.79, 190.6), ACTIVIDAD=c(1, 1,3, 3, 3,3, 1, 1,3, 2, 3,3, 2, 3,
2, 3, 2, 2,2, 2, 3,3, 3, 3,3, 3, 1,3, 3, 3,1, 2, 2,2, 2, 3,3, 3, 1), MOTIVACION=c(4, 9,12, 7,6, 12,11, 12, 10,6 , 7,11, 9, 9,8, 9, 6,8, 3, 6,6, 6, 2,9, 9, 12, 9,4,8, 3, 11,4, 2,12, 10, 12,7, 8, 10,
2), HTA=c(1, 0, 0, 1,1, 1, 1,0, 0, 1,0, 0, 0,0, 0, 0,0, 0, 1,1, 0, 0,1, 0, 0,0, 0, 1,0, 1, 0,1, 0, 1,0, 0, 0,0, 0, 0), DISLIPEMIA=c(1, 1, 1,0, 1, 1,1, 1, 0,0, 1, 0,0, 1, 0,0, 1, 0,0, 1, 0,0, 1, 1,1, 1, 1,0,
0, 1, 1, 1,1, 1, 1,1, 0, 1,0, 1), GLUCEMIA=c(0, 0,0, 1, 1,1, 1, 1,0, 0,0, 1, 0,0, 0,0, 1,0, 0, 1,0, 0, 1,0, 0,0, 0, 0,0, 0,0, 0, 1,1, 1, 0,1, 0, 1,0, 1), DE=c(0, 0, 1,1, 1, 1,1, 0, 1,0, 0,0, 0, 1,1, 0,1, 0,
0, 0,0, 0, 1,1, 1, 1,0, 1,0, 0, 1,1, 1, 1,1, 1,0, 1,1, 1,1, 0,1, 1,1), LDLox=c(1,1, 1, 1,1, 1, 0,0, 0, 0,1, 0, 0,0, 0, 1,1, 0, 1,0, 0, 0,1, 1, 1,0, 0, 1,0, 1, 0,1, 1, 0,0, 1, 0,1, 0, 0), FUMA=c(1, 1, 1,1, 1, 1,0, 1,
1, 0, 1,1, 1,1, 0, 1,1, 0,0,0, 0, 1,1, 1, 1,1, 1, 1,1, 1, 1,1, 1, 1,1, 1, 1,1, 1, 0,1, 1, 1), CCI=c(1, 0,0, 0, 1,1, 0, 1,0, 1,0, 1, 0,0, 1,0, 1,0, 1, 1,0, 0, 1,0, 0,0, 0, 0,1, 1, 0,1, 1, 1,1, 1, 0,0, 1, 0,0, 0),
TG0=c(82.61, 83.86, 84.47, 85.72, 84.47, 77.28, 77.3,81.16,86.53,87.6, 81.98, 83.43, 82.27, 89.07, 82.69, 89.13, 76.08, 73.82, 83.31, 76.57, 74.93, 92.85, 76.37, 80.42, 89.47, 84.95, 83.19,
78.78, 88.65, 89.23, 91.09, 77.95, 82.66, 77.41, 78.91, 79.15, 80.35, 72.67, 82.44, 77.45), ALBUMINA0=c(4.06, 4.48, 4.08,5.18, 4.77, 5.22,5.51, 4.67, 4.7, 4.05, 5.45,4.65, 4.31, 4.9, 4.38, 4.82,
5.22, 5.84, 5.06, 4.7, 5.79, 5.02,4.9, 4.59,5.53, 5.9,4.15, 5.82, 4.4, 4.52, 4.48,4.92, 5.59, 5.08,4.39, 4.07, 4.46,4.9, 4.65,4.81), MGRASA0=c(38.31, 21.29,33.99,30.46,32.34,30.36,45.65,
16.27, 31.3,21.4, 37.41, 34.21, 36.99, 20.03, 47.45, 33.09, 31.19, 19.15, 19.2, 36.6,35.4, 30.72, 19, 41.77, 52.59, 32.12, 20.66, 23.58, 24.31, 51.66, 33.24, 33.4,19.53,33.54,35.81,39.2,
33.49, 57.91, 29.35, 19.87), GLUCOSA0=c(54.91, 94.9,71.88,66.34,92.21,84.44,91.05,90.41,65.26,87.41,93.04,71.4, 73.47, 83.15, 91.79, 108.19,80.04,92.76,64.66,89.25,91.06,80.32,
100.12, 69.28, 55.53, 94.27, 66.6, 78.77, 89.13, 54.45, 78.12, 91.75, 78.78, 91.43, 96.28, 92.13, 81.39, 92.77, 76.43, 96.85), INSULINA0=c(33.33, 11.94,20.04,14.46,29.87,13.21,8.62, 19.25,
17.58, 27.21, 8.16,35.51,20.65, 32.54,23.33,22.06,28.6, 24.33, 16.6, 13.45, 15.9,15.73,6.86, 22.73, 26.8, 15.1,10.27, 33.6, 19.54, 23.36, 31, 17.84, 20.81, 11.03, 14.88, 21.63, 24.91, 7.99,
17.15, 27.45), AcURICO0=c(14.86, 9.13,11.68,11.15,9.87, 6.88, 8.81,9.35, 17.46, 17.61, 8.26, 15.74, 14.91, 12.55, 10.06, 13.16, 5.97,9.85, 13.64, 11.97, 8.13, 14.51, 11.16, 14, 11.56, 15.95,
14.08, 9.71, 17.75,16.47,13.7, 6.04, 8.44,8.58, 14.17, 11.31, 6.03, 5.77,10.17,9.59), TG2=c(76.52, 79.49,72.22,73.13, 80.6, 75.15, 73.06, 75.31, 76.33, 78.05, 80.65, 74.25, 74.39, 76.84,
77.69, 76.32, 73.61, 72.43, 76.46, 73.19, 72.25, 80.49, 73.89, 71.51, 76.68, 77.21, 72.05, 74.12, 76.56, 79.46, 79.37, 76.16, 80.01, 72.21, 72.18, 73.69, 74.89, 71.27, 72.65, 74.3), CCI2=c(1, 1,0,
0, 1, 1, 0, 1,0, 0, 1,0, 0, 0,0, 0, 0,1, 1, 0, 1,1, 0, 0,1, 0, 1,1, 1, 0,0, 0, 0,0, 0), ALBUMINA2=c(2.06, 3.82, 3.08, 3.18,3.91, 4.28, 4.62,4.12, 3.7,2.05, 5.38, 3.65,2.31, 3.9,3.79,
3.82, 5.21, 5.11, 3.06,4.16, 5.43, 3.02, 3.9, 3.59,4.53, 3.9,3.15, 5.63, 3.4, 2.52, 2.48,4.1, 5.42,4.5, 2.39,3.21, 3.66, 4.84,2.65, 4.52), MGRASA2=c(26.84, 17.87, 24.4, 23.42, 29.18, 27.51,
43.89, 10.46, 23.58, 12.89, 35.38, 26.8,26.39,7.38, 43.21, 26.52, 27.13, 13.62, 10.5, 31.7, 29.99, 22.42, 17.89, 33.82, 40. 64, 21.21, 14.41, 21.93, 13.74, 44.41, 24.05, 30.11, 15.1, 32.25, 28.27,
37.24, 32.41, 53.41, 20.33, 17.7), GLUCOSA2=c(45.93, 90, 62.15, 55.44, 89.74, 80.91, 89.35, 85.62, 56.65, 79.53, 91.9, 60.78, 62.22, 75.95, 88.29, 96.18, 78.66, 87.23, 55.96, 83.41, 88.2, 73.82,
94.86, 61.7, 45.64, 82.05, 57.24, 77.03, 79.56, 46.76, 67.31, 88.28, 77.06, 89.02, 88.52, 86.33, 76.77, 89.5, 69.37, 95.07), INSULINA2=c(22.71, 8.22, 12.03, 3.03, 25.71, 10.53, 5.24, 14.25, 6.08,
17.68, 4.61, 23.58, 10.15, 23.81, 21.02, 10.69, 26.28, 20.61, 5.9, 8, 11.11, 7.39, 5.39, 12.5, 15.55, 2.45, 1.2, 27.89, 11.97, 14.79, 19.9, 16.17, 15.06, 6.4, 3.86, 17.9, 20.28, 5.86, 10.37, 24.43),
AcURICO2=c(1.89, 5.85, 3.12, 3.17, 7.99, 4.3, 5.68, 5.04, 5.49, 4.78, 4.33, 3.17, 5.52, 3.92, 4.81, 4.89, 4.83, 6.51, 7.21, 6 .61, 5.72, 3.35, 8.46, 4.37, 5.03, 3.92, 6.09, 4.01, 5.94, 5.82, 5.53, 4.54,
4.58, 3.74, 5.14, 5.57, 4.37, 4.48, 3.85, 5.37))
head(datos)
```

- CREAR UN CONJUNTO DE DATOS
- CREAR (CALCULAR) UNA VARIABLE
- CREAR UNA VARIABLE A PARTIR DE OTRA
- DEFINIR LAS VARIABLES
- SELECCIONAR CASOS

CREAR (CALCULAR) UNA VARIABLE

nº de antecedentes personales totales (NAT)

NAT= {HTA + DISLIPEMIA + GLUCEMIA + DE +
LDLox + FUMA + CCI}

DATOS

Sintaxis (copiar el texto entero y pegar en el compilador):

Copiamos datos

```
# Creamos: NAT= {HTA + DISLIPEMIA + GLUCEMIA + DE + LDLox + FUMA + CCI}

datos$NAT= datos$HTA + datos$DISLIPEMIA + datos$GLUCEMIA +
datos$DE + datos$LDLox + datos$FUMA + datos$CCI

head(datos)
```

- CREAR UN CONJUNTO DE DATOS
- CREAR (CALCULAR) UNA VARIABLE
- **CREAR UNA VARIABLE A PARTIR DE OTRA**
- DEFINIR LAS VARIABLES
- SELECCIONAR CASOS

CREAR UNA VARIABLE A PARTIR DE OTRA

Codifica la variable NAT en otra variable que sea NAT2c, de tal manera que:

NAT2c = {1='0-2‘ y 2='=>3'}

DATOS

Sintaxis (copiar el texto entero y pegar en el compilador):

Copiamos datos

```
# Creamos: NAT2c  = {1='0-2', 2='=>3'}

datos$NAT2c= ifelse( datos$NAT<=2,  1,  2)

head(datos)
```

- CREAR UN CONJUNTO DE DATOS
- CREAR (CALCULAR) UNA VARIABLE
- CREAR UNA VARIABLE A PARTIR DE OTRA
- DEFINIR LAS VARIABLES
- SELECCIONAR CASOS

DATOS

Con la ayuda del archivo que contiene la definición de variables, vamos a definir nuestra BD

Sintaxis (copiar el texto entero y pegar en el compilador):

```
Copiamos datos
# Convertimos en formato fecha los valores numéricos obtenidos del excel
  datos$F_INI   = as.Date(datos$F_INI, origin = "1899-12-30")

# Convertimos las variables nominales y añadimos las etiquetas
  datos$SEXO =   factor(datos$SEXO, levels = c(0,1), labels = c('Mujer', 'Hombre'),
  ordered = FALSE)
# Fíjate bien que los valores de las etiquetas llevan comillas

# Las variables continuas no hay que definirlas
```

- CREAR UN CONJUNTO DE DATOS
- CREAR (CALCULAR) UNA VARIABLE
- CREAR UNA VARIABLE A PARTIR DE OTRA
- DEFINIR LAS VARIABLES
- SELECCIONAR CASOS

SELECCIONAR CASOS

Seleccionamos los pacientes tratados con el nuevo tratamiento.

- Si no hemos puesto las etiquetas, la codificación es 1 (valor numérico), en otro caso, la etiqueta es ‘Referencia’ (texto).

El resultado será la nueva base solo con esos pacientes.

DATOS

Sintaxis (copiar el texto entero y pegar en el compilador):

Copiamos datos con las etiquetas:

```
# Creamos la sub-base datos2 cuyos pacientes solo recibieron el tratamiento de referencia.
datos2= datos[datos$TRATAMIENTO=='Referencia', ]
head(datos2)

#datos2= datos[which(datos$TRATAMIENTO=='Referencia'), ]
#head(datos2)
# fíjate bien que Referencia lleva comillas porque es una etiqueta, no un número.
```

Copiamos datos sin las etiquetas:

```
# Creamos la sub-base datos2 cuyos pacientes solo recibieron el tratamiento de referencia.
datos2= datos[datos$TRATAMIENTO==1,]
head(datos2)

#datos2= datos[which(datos$TRATAMIENTO==1),]
#head(datos2)
```

Investigación Biomédica

1. Recordar definición de variables
2. Depuración de los datos
3. Toma de contacto con la consola de R
4. Estadística descriptiva

Investigación Biomédica

Estadística descriptiva

Métodos estadísticos **DESCRIPTIVOS**:

Sirven para describir y resaltar numéricamente aquello que es esencial en los resultados de un estudio usando los métodos apropiados

Métodos estadísticos **INFERENCIALES**:

Conjunto de métodos que permiten obtener conclusiones, de una manera objetiva, sobre los datos que se están investigando.

Investigación Biomédica

Recordatorio. Medidas descriptivas

TIPO DE VARIABLE	GRÁFICA	MEDIDAS CENTRALES	MEDIDAS DE DISPERSIÓN
NOMINAL	•DIAG. BARRAS •DIAG. SECTORES	MODA (%)	
ORDINAL	•DIAG. BARRAS •BOX-PLOT	MEDIANA	RANGO P_{75}-P_{25}
INTERVALO /RAZON	•BOX-PLOT •HISTOGRAMAS	MEDIA	DESV. TÍPICA

Investigación Biomédica

Estadística descriptiva

Variable categórica

- ✓ *Frecuencias absolutas*
- ✓ *Frecuencias relativas (Porcentajes)*

Variable continua

- ✓ *Medidas de centralización y orden*
- ✓ *Medidas de dispersión*

Investigación Biomédica

Estadística descriptiva

Variable categórica

- ✓ *Frecuencias absolutas*
- ✓ *Frecuencias relativas (Porcentajes)*

Variable continua

- ✓ *Medidas de centralización y orden*
- ✓ *Medidas de dispersión*

Investigación Biomédica

Estadística descriptiva

Descripción numérica - cualitativas

- Frecuencias absolutas

Contabilizan el número de individuos de cada atributo

- Frecuencias relativas (%)

Ídem, pero se divide por el total

Investigación Biomédica

Estadística descriptiva

Descripción gráfica - cualitativas

- Diagramas de barras

Alturas proporcionales a las frecuencias (absolutas o relativas).

Se pueden aplicar también a variables discretas.

- Diagramas de sectores (tartas, polares)

No usarlo con variables ordinales.

El área de cada sector es proporcional a su frecuencia (abs. o rel.).

DESCRIPCIÓN NUMÉRICA

Describir la variable ÉXITO (pacientes que perdieron > 10 Kg, PCAMBIO).

DATOS

Sintaxis (copiar el texto entero y pegar en el compilador):

Copiamos datos

```
# Creamos las frecuencias absolutas y relativas en %
frec = table(datos$PCAMBIO)
frec
frecRel = 100*table(datos$PCAMBIO)/length(datos$PCAMBIO)
frecRel
```

No	Si
10	30

→ Hace referencia a la frecuencia absoluta

No	Si
25	75

→ Hace referencia al %

DESCRIPCIÓN GRÁFICA

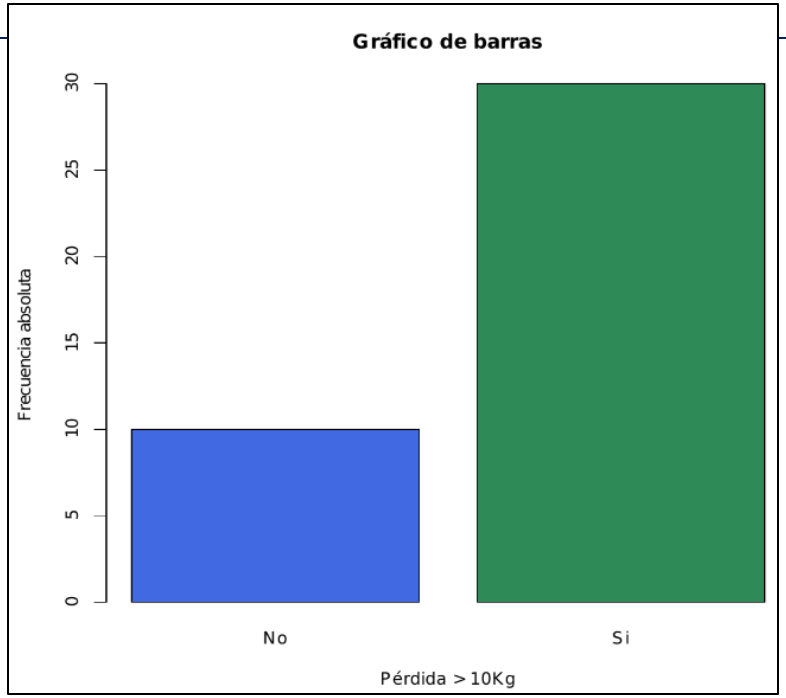
Describir la variable ÉXITO (pacientes que perdieron > 10 Kg, PCAMBIO).

DATOS

Sintaxis (copiar el texto entero y pegar en el compilador):

Copiamos datos

```
# Creamos el gráfico de barras
plot(x = datos$PCAMBIO, main = "Gráfico de barras",
    xlab = "Pérdida > 10Kg", ylab = "Frecuencia absoluta",
    col = c("royalblue", "seagreen"))
```



DATOS

Sintaxis (copiar el texto entero y pegar en el compilador):

Copiamos datos

```
# Creamos el gráfico de sectores
frecRel = 100*table(datos$PCAMBIO)/length(datos$PCAMBIO)
pie(x = frecRel, main = "Gráfico de sectores",
    xlab = "Pérdida >10Kg", ylab = "%",
    col = c("royalblue","seagreen"))
```

Investigación Biomédica

Estadística descriptiva

Variable categórica

- ✓ *Frecuencias absolutas*
- ✓ *Frecuencias relativas (Porcentajes)*

Variable continua

- ✓ *Medidas de centralización y orden*
- ✓ *Medidas de dispersión*

Descripción - cuantitativas

Orden

Dividen un conjunto ordenado de datos en grupos con la misma cantidad de individuos.

Cuantiles, Percentiles, Cuartiles, Deciles,...

Centralización

Indican valores con respecto a los que los datos parecen agruparse.

Media, Mediana y Moda

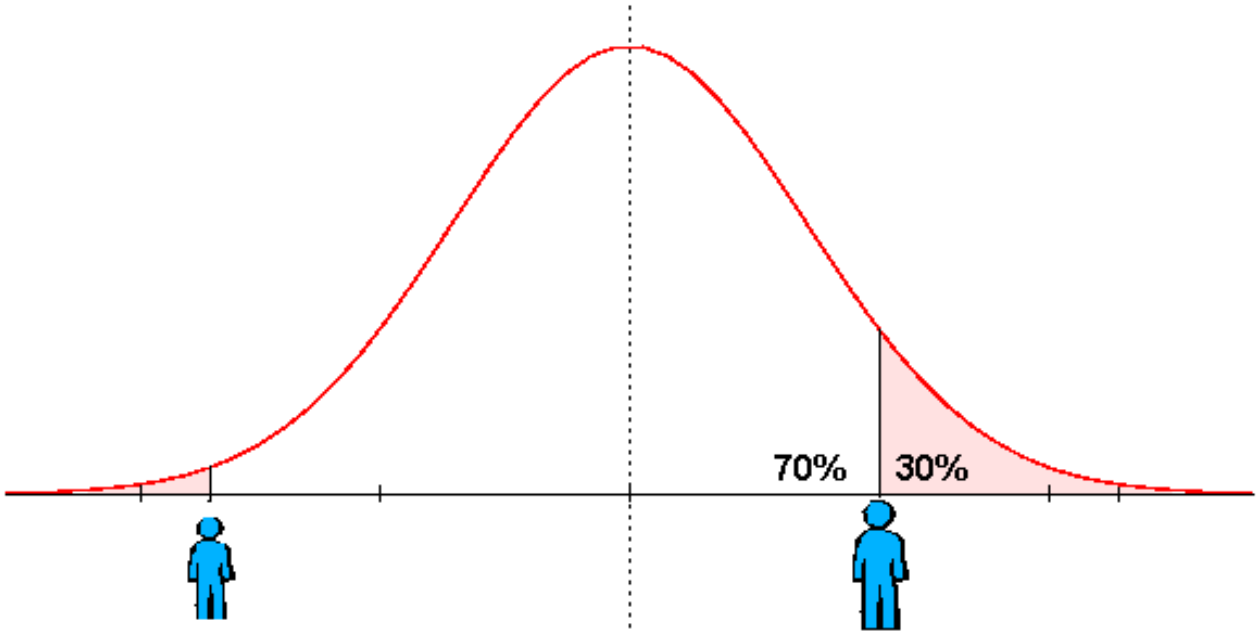
Dispersión

Indican la mayor o menor concentración de los datos con respecto a las medidas de centralización.

Varianza, Desviación típica (st), Coeficiente de Variación (CV), Rango, Rango Intercuartílico (RI)

Orden - Cuantil

Se define el cuantil de orden α como un valor de la variable por debajo del cual se encuentra una frecuencia acumulada α



38

Orden - Cuantil

- ✓ **Cuartiles,** Q_k con k=1, 2 ó 3

  Dividen a la muestra en 4 grupos con frecuencias similares.

$$Q_K : \frac{K*(N+1)}{4}, K = 1,2,3$$

  - Q_1 = Cuantil 0,25 (Percentil 25, P_{25})
  - Q_2 = Cuantil 0,5 = mediana (Percentil 50, P_{50})
  - Q_3 = Cuantil 0,75 (Percentil 75, P_{75})

39

Descripción - cuantitativas

Centralización

Son medidas que buscan posiciones (valores) con respecto a los cuales los datos muestran tendencia a agruparse.

Media

Promedio de los valores de una variable. Suma de los valores dividido por el tamaño muestral.

Conveniente cuando los datos se concentran simétricamente con respecto a ese valor.

Muy sensible a valores extremos.

$$\bar{x} = \frac{\sum_i x_i}{n}$$

Mediana

Valor que divide a las observaciones en dos grupos con el mismo número de individuos (cuartil 2). Si el número de datos es par, se elige la media de los dos datos centrales.

Es conveniente cuando los datos son asimétricos.

No es sensible a valores extremos.

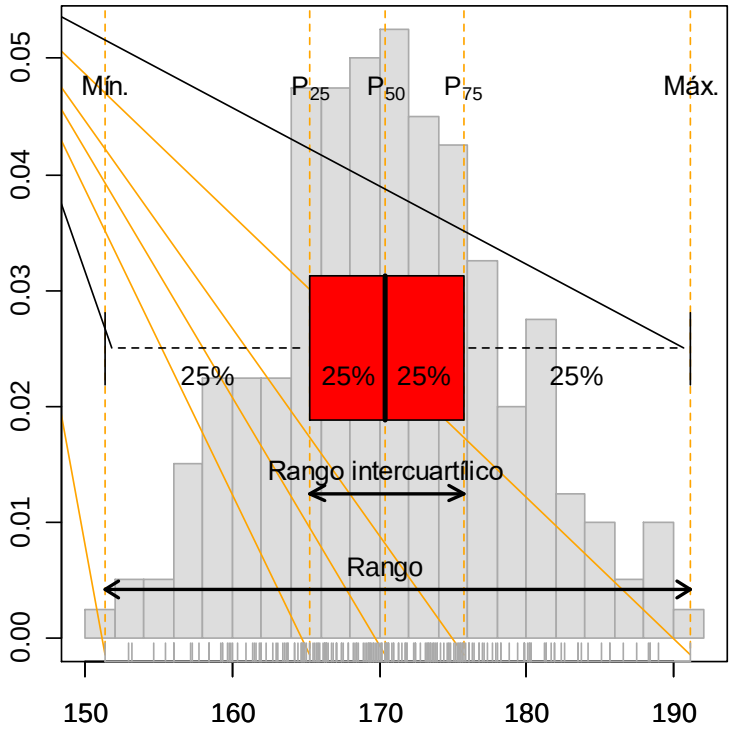
Moda Valor/es donde la distribución de frecuencia alcanza un máximo.

40

Descripción - cuantitativas

Dispersión

Miden el grado de dispersión (variabilidad) de los datos, independientemente de su causa.



Amplitud o Rango

Máximo - Mínimo

Es muy sensible a los valores extremos.

Rango intercuartílico

$Q_3 - Q_1$

No es tan sensible a valores extremos.

Dispersión

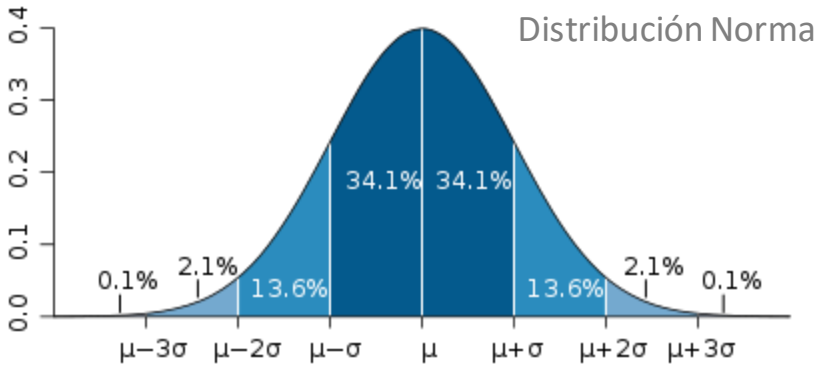
Varianza muestral S^2

Mide el promedio de las desviaciones (al cuadrado) de las observaciones con respecto a la media.
Es sensible a valores extremos (alejados de la media).

$$S^2 = \frac{1}{n}\sum_i (x_i - \bar{x})^2$$

Desviación típica S

$$S = \sqrt{S^2}$$



Centrado en la media (μ) y a 1 dt (σ) de distancia hay aproximadamente el 68% de las observaciones.
A dos desviaciones típicas tenemos el 95% (aprox.)

42

Dispersión

Coeficiente de Variación CV

Es la razón entre la desviación típica y la media.

- Es frecuente mostrarla en porcentajes.

  Si la media es 80 y la desviación típica 20 entonces CV=20/80=0,25=25% (variabilidad relativa).

- Interesante para comparar la variabilidad de diferentes variables.

  Si el peso tiene CV=30% y la altura tiene CV=10%, los individuos presentan más dispersión en peso que en altura.

- No debe usarse cuando la variable presenta valores negativos.

$$CV = \frac{S}{\bar{x}}$$

43

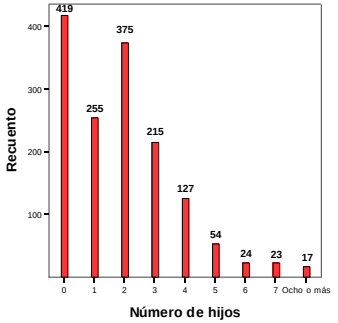
Descripción gráfica - cuantitativas

- ## Histograma

Alturas proporcionales a las frecuencias (abs. o rel.)

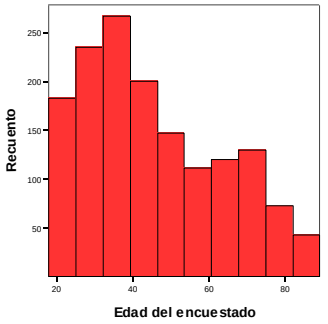
Diagramas barras para v. discretas

Se deja un hueco entre barras para indicar los valores que no son posibles.



Histogramas para v. continuas

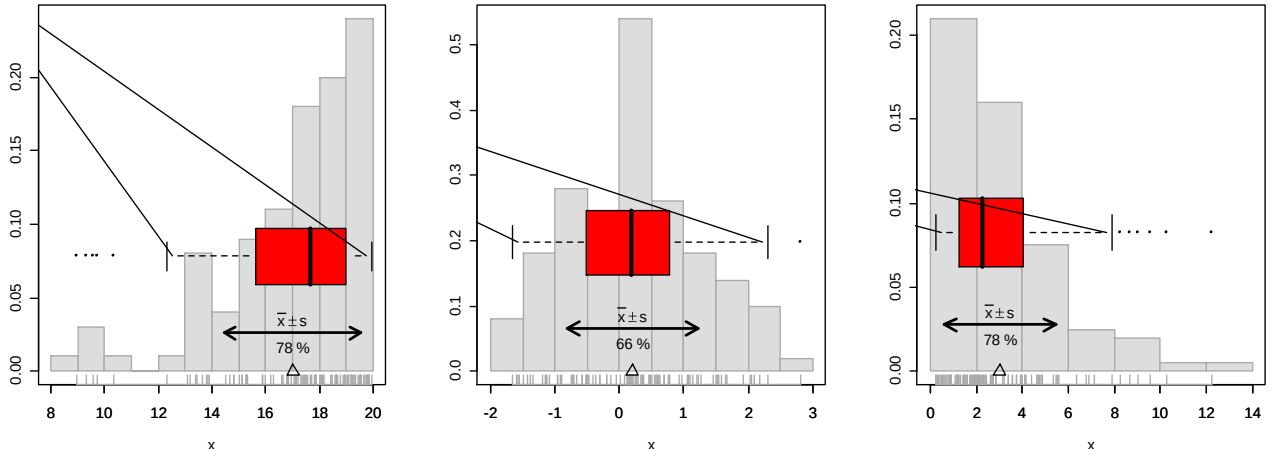
El área que hay bajo el histograma entre dos puntos cualesquiera indica la cantidad (% o frecuencia) de individuos en el intervalo.



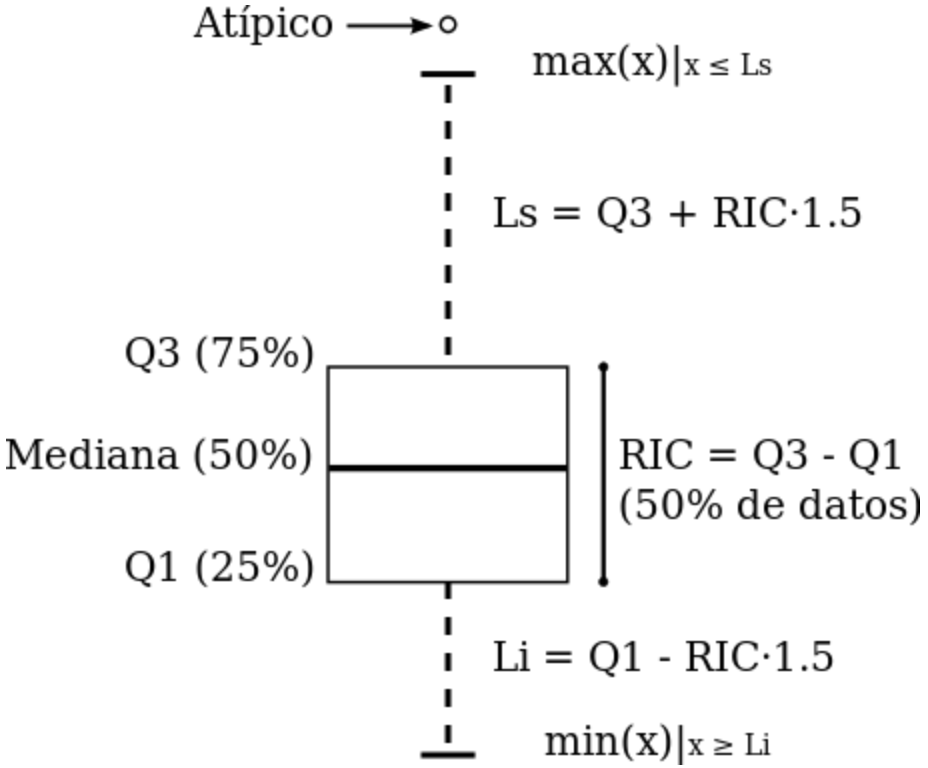
Descripción gráfica - cuantitativas

Box-Plot (diagrama de cajas y bigotes)

- Buena aproximación de la distribución.
- La zona central, 'caja', contiene al 50% central de las observaciones. Su tamaño se llama 'rango intercuartílico' (RI)
- Es costumbre que 'los bigotes', no lleguen hasta los extremos, sino hasta las observaciones que se separan de la caja en no más de 1,5 RI.
- Más allá de esa distancia se consideran anómalas.



- Box-Plot (diagrama de cajas y bigotes)



DESCRIPCIÓN NUMÉRICA

Describir la variable P0 (peso basal).

DATOS

Sintaxis (copiar el texto entero y pegar en el compilador):

Copiamos datos

#media
mean(datos$P0)

#desviación típica
sd(datos$P0)

#varianza
var(datos$P0)

#CV
100*sd(datos$P0)/mean(datos$P0)

#mediana
median(datos$P0)

# rango intercuartílico
IQR(datos$P0)

# rango
max(datos$P0)-min(datos$P0)

Media	$\bar{x} = \frac{\sum_i x_i}{n}$	106,78
D.t	$S^2 = \frac{1}{n}\sum_i (x_i - \bar{x})^2$	6,86
Varianza	$S = \sqrt{S^2}$	47,10
CV	$CV = \frac{S}{\bar{x}}$	6,42
Mediana	Q_2	107,54
RI	$Q_3 - Q_1$	7,02
Rango	Max – Min	31,80

DESCRIPCIÓN GRÁFICA

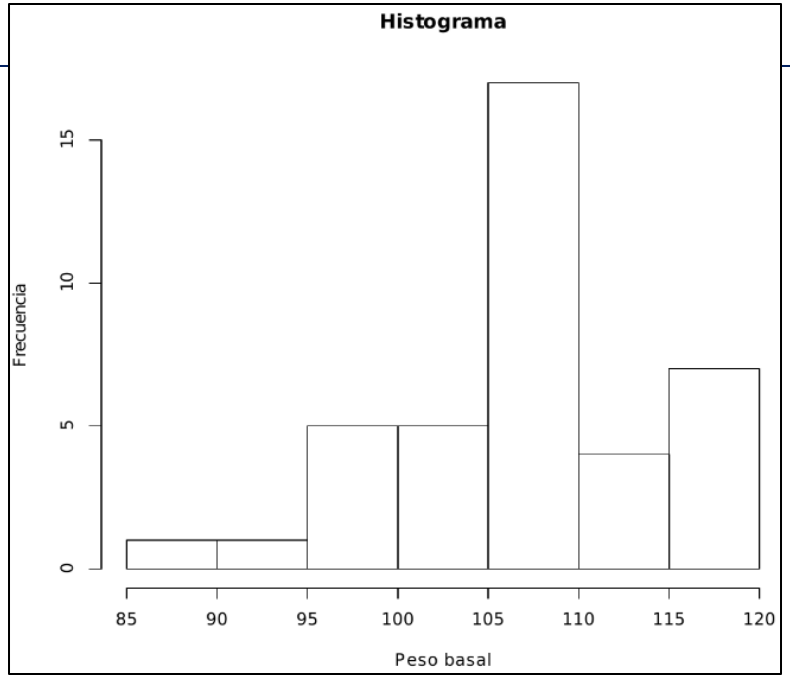
Describir la variable P0 (peso basal).

DATOS

Sintaxis (copiar el texto entero y pegar en el compilador):

Copiamos datos

```
# Creamos el histograma
hist(x = datos$P0, main = "Histograma",
    xlab = "Peso basal", ylab = "Frecuencia")
```

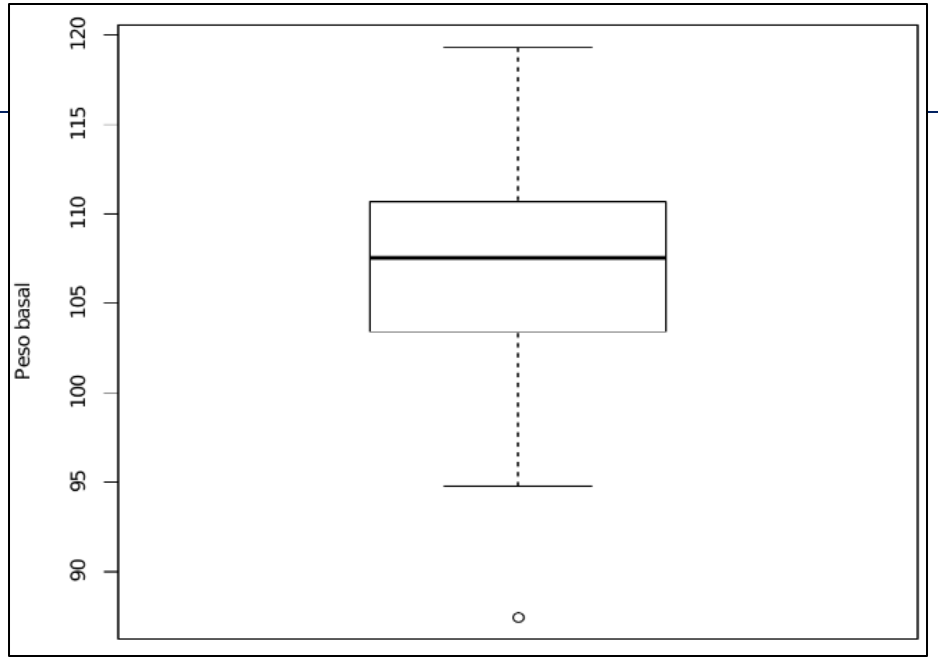


DATOS

Sintaxis (copiar el texto entero y pegar en el compilador):

Copiamos datos

```
# Creamos el diagrama de cajas - boxplot
boxplot(datos$P0,  ylab = "Peso basal")
```



IdiPAZ
Instituto de Investigación
Hospital Universitario La Paz

SaludMadrid
Hospital Universitario La Paz
Hospital de Cantoblanco
Hospital Carlos III
Comunidad de Madrid

THANK YOU !

mariana.diaz@salud.madrid.org

BIOESTADÍSTICA