



UNIVERSITAT^{DE}
BARCELONA

Treball en R. Estadística Descriptiva

Noah Márquez
Alejandro Guzman

21 març 2022

ÍNDEX

1	Introducció	3
2	Estudi descriptiu univariant	4
1	Variable \$SEX	4
2	Variable \$ESTUDIS	4
3	Variable \$TABAC	5
4	Variable \$EDAT	6
5	Variable \$PES	7
6	Variable \$TALLA	7
7	Estudi descriptiu	9
3	Estudi descriptiu bivariant	11
1	Variables \$SEX & \$ESTUDIS	11
1.1	Interpretació d'alguns valors pels homes	11
1.2	Interpretació d'alguns valors per les dones	11
2	Variables \$SEX & \$TABAC	12
2.1	Interpretació d'alguns valors pels homes	13
2.2	Interpretació d'alguns valors per les dones	13
3	Variables \$PES & \$TALLA	14
4	Variables \$PES & \$IMC	15
4	Conclusions	16
5	Script en R	17
	Bibliografia	24

1 INTRODUCCIÓ

Per a realitzar l'estudi estadístic hem considerat com a població un grup de ciutadans galecs que es van sotmetre a un estudi el 2018 per part de la Xunta de Galícia englobat en el SICRI (Sistema d'Informació sobre Conductes de Risc) amb l'objectiu d'identificar la presència que tenen en la població galega diversos factors de risc per a la salut.

Els resultats obtinguts en l'estudi del SICRI estan a la seva web (indicada al final d'aquesta pàgina). Nosaltres no hem fet servir totes les diferents variables que van tenir en compte, sinó només unes quantes que són les que detallem a continuació:

- **SEX:** Sexe (1-Home, 2-Dona).
- **EDAT:** Edat.
- **ESTUDIS:** Màxim nivell d'estudis completats (1-Sense estudis, 2-Nivell bàsic, 3-Nivell mitjà, 4-Nivell superior).
- **E-SALUT:** Estat de salut autopercebuda (1-Molt bo, 2-Bé, 3-Regular, 4-Dolent).
- **TABAC:** Relació amb el tabac (1-Diari, 2-Ocasional, 3-Exfumador, 4-Mai ha fumat).
- **PES:** Pes en kg.
- **TALLA:** Talla en cm.
- **IMC:** Índex de massa corporal en Kg/m^2 .

Hi ha variables qualitatives (SEX & ESTUDIS & E-SALUT & TABAC), variables quantitatives discretes (EDAT & PES & TALLA & IMC), entre d'altres.

El present estudi conté un total de 7853 mostres, hem cregut convenient aprofitar-les totes per tal de fer l'anàlisi estadístic univariant i bivariant.

Tota la informació relacionada amb l'estudi es pot trobar a la bibliografia segons:

- **Metodologia** [1]
- **Qüestionari** [2]
- **Resultats** [3]
- **Microdades** [4]

2 ESTUDI DESCRIPTIU UNIVARIANT

1 Variable \$SEX

El primer que hem fet ha estat la taula de freqüències de la variable SEX. No hem afegit les freqüències acumulades ja que no tindrien gaire sentit.

SEX	Freqüència absoluta	Freqüència relativa
1	3925	0,4998
2	3928	0,5002
Totals	7853	1

Taula 2.1: Freqüències de la variable \$SEX

2 Variable \$ESTUDIS

Per la variable ESTUDIS hem realitzat una taula completa amb totes les seves freqüències, per tal de veure el nivell educatiu de la població enquestada. Es detalla a continuació:

ESTUDIS	Freq. abs	Freq. abs. acum.	Freq. rel	Freq. rel. acum.
1	328	328	0,04177	0,04177
2	3340	3668	0,42532	0,46708
3	2692	6360	0,34280	0,80988
4	1493	7853	0,19012	1

Taula 2.2: Freqüències de la variable \$ESTUDIS

També hem realitzat un diagrama de barres i un diagrama de sectors per tal de visualitzar d'una forma molt més gràfica les dades:

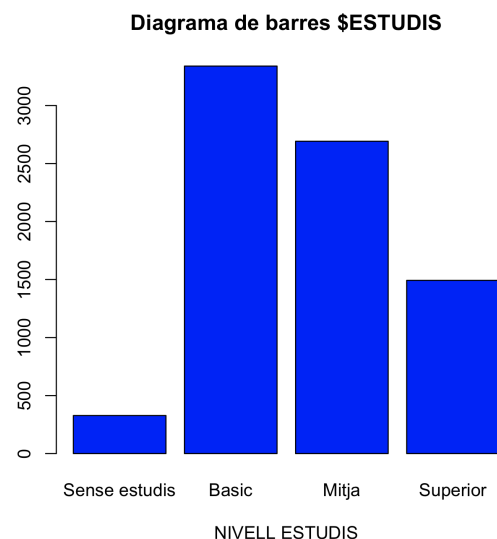


Figura 2.1: Diagrama de barres \$ESTUDIS

Diagrama de sectors \$ESTUDIS

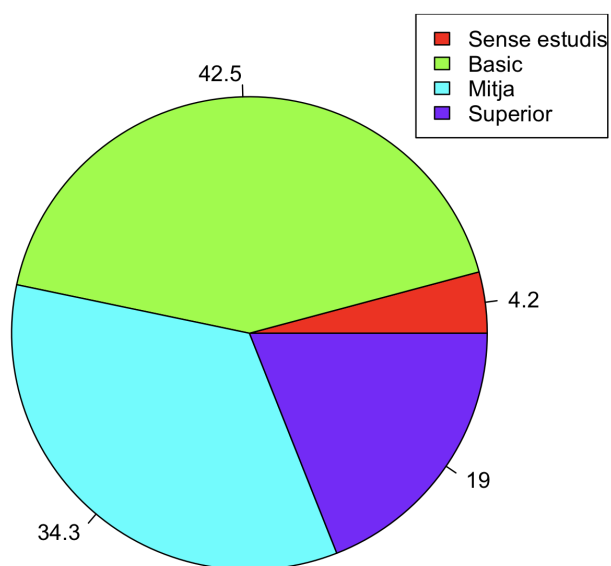


Figura 2.2: Diagrama de sectors \$ESTUDIS

3 Variable \$TABAC

Per la variable TABAC també hem realitzat la seva taula de freqüències completa, per tal de comprovar els hàbits en relació amb el tabac de la població de Galícia.

TABAC	Freq. abs	Freq. abs. acum.	Freq. rel	Freq. rel. acum.
1	1248	1248	0,15892	0,15892
2	310	1558	0,03948	0,19840
3	2120	3678	0,26996	0,46836
4	4175	7853	0,53164	1

Taula 2.3: Freqüències de la variable \$TABAC

A la següent pàgina s'adjunten dues imatges que consisteixen en els diagrames de barres i sectors per la variable TABAC, per tal de poder apreciar de forma molt més fàcil la distribució dels resultats.

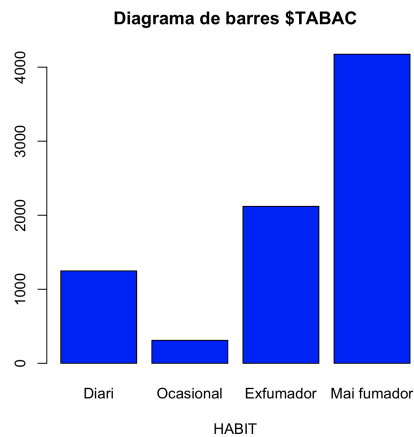


Figura 2.3: Diagrama de barres \$TABAC

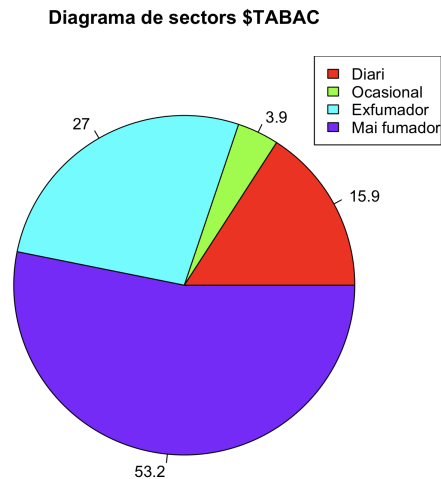


Figura 2.4: Diagrama de sectors \$TABAC

4 Variable \$EDAT

Per la variable EDAT hem fet els seus histogrames. Un histograma és una representació gràfica en forma de barres que simbolitza la distribució d'un conjunt de dades.

Primerament mostrem dos histogrames (freqüència absoluta), un més detallat que l'altre.

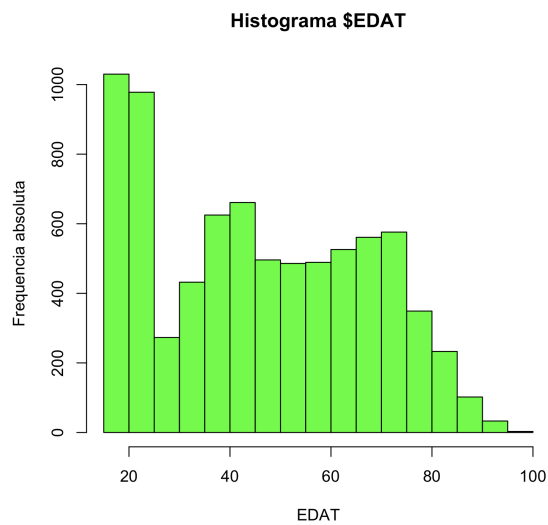


Figura 2.5: Histograma \$EDAT

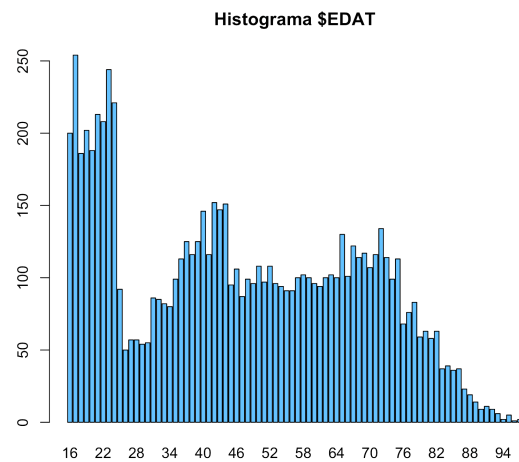


Figura 2.6: Histograma detallat \$EDAT

Podem veure com la distribució de les dades en l'histograma no és simètrica, degut sobretot al pic de freqüència provocat per la quantitat de persones de entre 16 i 22 anys.

Ara mostrarem els histogrames de la variable EDAT però per la freqüència absoluta acumulada i la freqüència relativa:

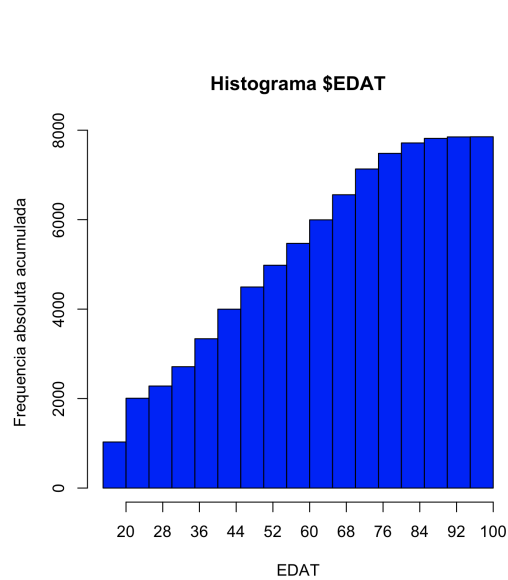


Figura 2.7: Histograma f. abs. acum. \$EDAT

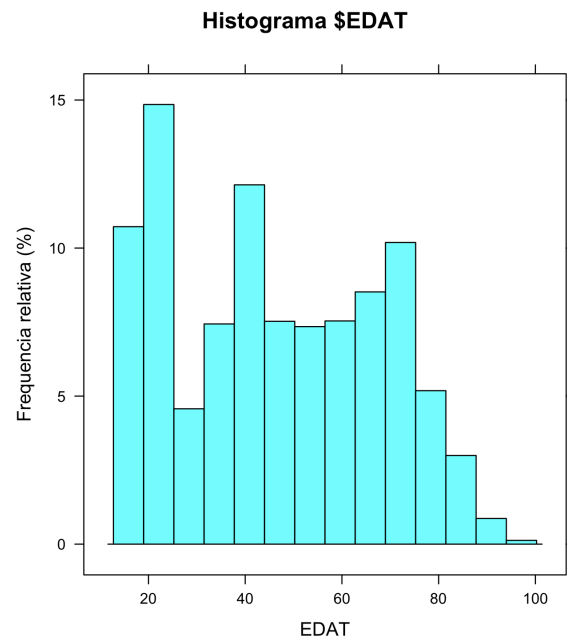


Figura 2.8: Histograma f. relativa \$EDAT

5 Variable \$PES

Per la variable PES també hem fet el seu histograma (de la freqüència absoluta).

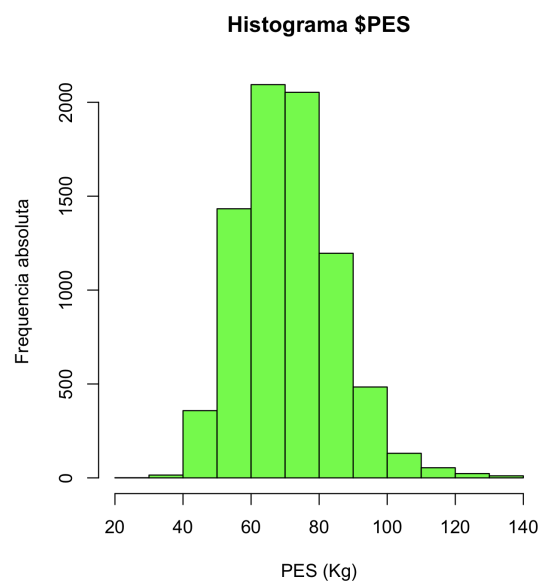


Figura 2.9: Histograma \$PES

A diferència de la variable EDAT, podem veure la simetria de les dades proporcionades per l'histograma de la variable PES.

6 Variable \$TALLA

A continuació mostrem també l'histograma de la variable TALLA.

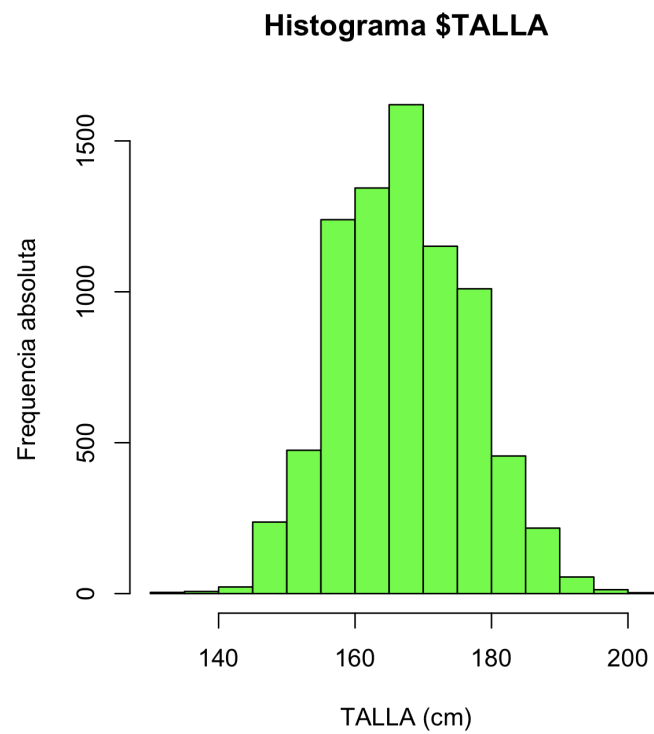


Figura 2.10: Histograma \$TALLA

La variable TALLA també ens ofereix un histograma amb una simetria bastant clara de les dades.

7 Estudi descriptiu

Hem dut a terme també un estudi descriptiu de la variable \$PES -en Kg- distingint segons la variable \$SEX -home i dona-. Recollim els estadístics numèrics a continuació:

	Tots	Homes	Dones
Mitjana	72,26573	79,16195	65,37478
Mediana	71	78	64
Moda	70	80	60
Desviació	14,3663	12,87477	12,32932
Variància	206,3906	165,7597	152,0121
Coefficient de variació	19,87983	16,26384	18,85944
Mínim	26,9431	44	26,9431
Q1	62	70	57
Q3	80	86	72
Màxim	140	140	140
Rang	113,0569	96	113,0569
Rang interquartílic	18	16	15
Skewness	0,6278376	0,8228888	0,8976143
Curtosis	0,9230276	1,680371	1,695012

Taula 2.4: Estudi descriptiu \$PES segons \$SEX

I els gràfics:

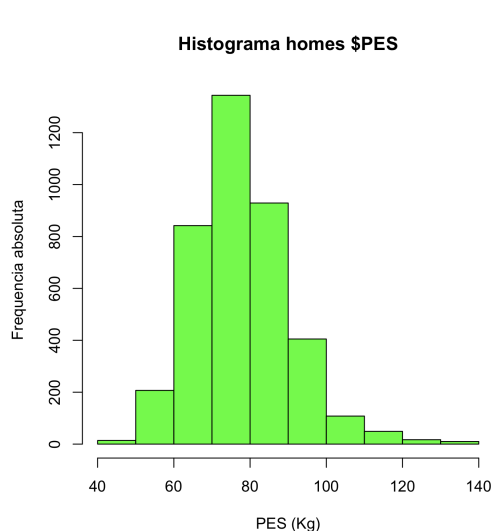


Figura 2.11: Histograma homes segons \$PES

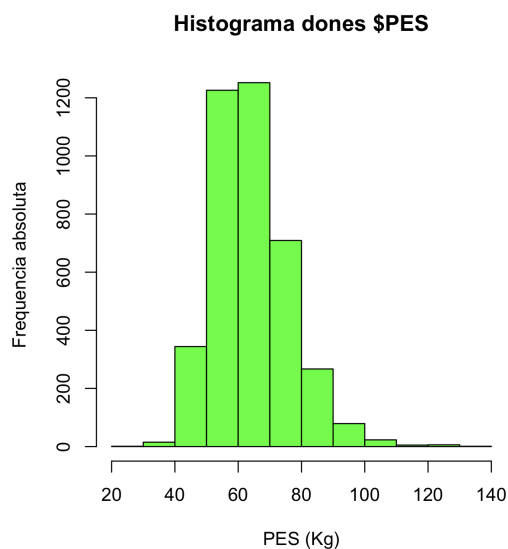


Figura 2.12: Histograma dones segons \$PES

També hem representat les dades mitjançant un diagrama de caixes. Un diagrama de caixa (també, *boxplot*) és un mètode estandarditzat per representar gràficament una sèrie de dades numèriques a través dels seus quartils. D'aquesta manera, es mostren a simple vista la mitjana i els quartils de les dades, i també se'n poden representar els valors atípics.

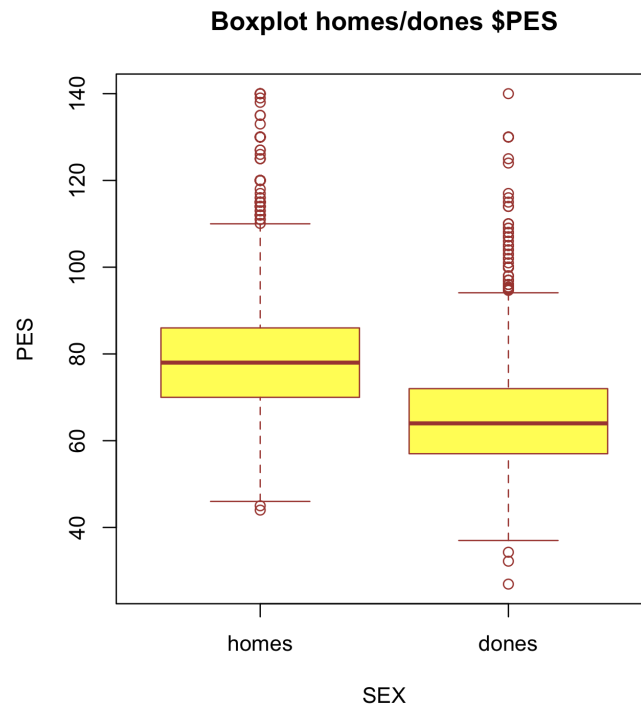


Figura 2.13: Boxplot homes/dones segons \$PES

Podem veure que tant en el cas dels homes com en el de les dones hi ha molts *outliers*, individus amb pes superior a la mitjana.

No hem realitzat cap diagrama de tija i fulles degut a que amb la quantitat de dades que hem treballat no es podia apreciar gaire bé.

De la taula 2.4 podem veure com el valor de *Skeewness* o coeficient d'asimetria és positiu, indicant que les dades tendeixen a l'esquerra de la mitjana (més proporció de valors petits), es pot comprovar també amb els boxplots i els histogrames.

També podem veure com la mesura d'apuntament o *Curtosi* és positiva, indicant que la corba és més apuntada que una campana de Gauss.

3 ESTUDI DESCRIPTIU BIVARIANT

L'anàlisi bivariant s'utilitza per tal de descobrir si existeix una certa relació entre les variables estudiades. És per això que els següents apartats estaran orientats a descobrir si existeix alguna relació entre les estudiades en el present document.

1 Variables \$SEX & \$ESTUDIS

Les taules de contingència es fan servir per registrar i analitzar l'associació entre dues o més variables, habitualment de naturalesa qualitativa (nominals o ordinals).

A continuació mostrem la taula de contingència entre les variables SEX i ESTUDIS i la seva posterior interpretació.

	Sense estudis	Bàsic	Mitjà	Superior	Totals
Home	139	1704	1412	670	3925
Dona	189	1636	1280	823	3928
Totals	328	3340	2692	1493	7853

Taula 3.1: Taula contingència \$SEX & \$ESTUDIS

1.1 Interpretació d'alguns valors pels homes

- La freqüència absoluta ("Home", "Sense estudis") és 139.
- La freqüència absoluta de "Home" és 3925.
- La freqüència relativa de "Home" és $\frac{3925}{7853} = 0,4998$.
- La freqüència relativa de ("Home", "Sense estudis") és $\frac{139}{7853} = 0,0177$.
- La freqüència relativa de "Sense estudis" dins dels homes és de $\frac{139}{3925} = 0,0354$.
- La freqüència relativa de "Home" dins dels que són "Sense estudis" és de $\frac{139}{328} = 0,4238$.

1.2 Interpretació d'alguns valors per les dones

- La freqüència absoluta ("Dona", "Superior") és 823.
- La freqüència absoluta de "Dona" és 3928.
- La freqüència relativa de "Dona" és $\frac{3928}{7853} = 0,5002$.
- La freqüència relativa de ("Dona", "Superior") és $\frac{823}{7853} = 0,1048$.
- La freqüència relativa de "Superior" dins de les dones és de $\frac{823}{3928} = 0,2095$.
- La freqüència relativa de "Dona" dins dels que són "Superior" és de $\frac{139}{1493} = 0,4238$.

Hi ha alguns gràfics -tipus diagrames de barres- que volen representar aquestes taules de contingència:

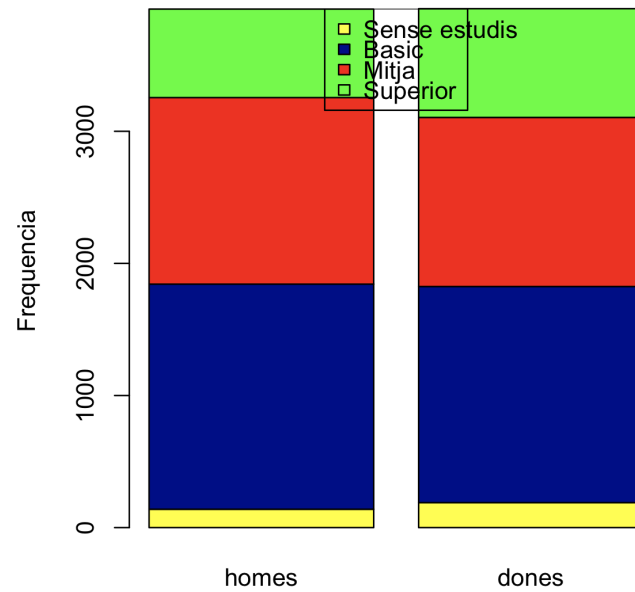


Figura 3.1: Diagrama de barres I

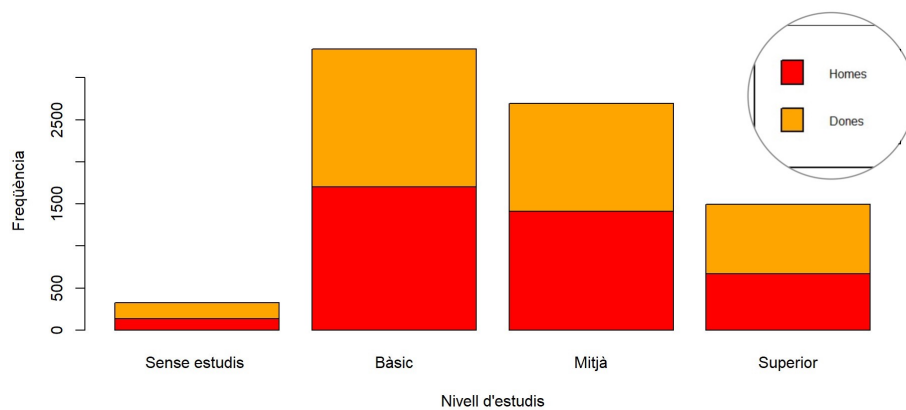


Figura 3.2: Diagrama de barres II

Es pot comprovar que en relació amb els estudis no varia gaire tant si ho mirem pels homes com per les dones; veiem més o menys la mateixa proporció pels dos gèneres.

2 Variables \$SEX & \$TABAC

Realitzem un segon estudi bivariant entre dues variables qualitatives: SEX i TABAC.

A continuació mostrem la taula de contingència entre aquestes dues variables i la seva posterior interpretació.

	Diari	Ocasional	Exfumador	Mai ha fumat	Totals
Home	722	171	1374	1658	3925
Dona	526	139	746	2517	3928
Totals	1248	310	2120	4185	7853

Taula 3.2: Taula contingència \$SEX & \$TABAC

2.1 Interpretació d'alguns valors pels homes

- La freqüència absoluta (“Home”, “Diari”) és 722.
- La freqüència absoluta de “Home” és 3925.
- La freqüència relativa de “Home” és $\frac{3925}{7853} = 0,4998$.
- La freqüència relativa de (“Home”, “Diari”) és $\frac{722}{7853} = 0,0919$.
- La freqüència relativa de “Diari” dins dels homes és de $\frac{722}{3925} = 0,1839$.
- La freqüència relativa de “Home” dins dels que són “Sense estudis” és de $\frac{722}{1248} = 0,5785$.

2.2 Interpretació d'alguns valors per les dones

- La freqüència absoluta (“Dona”, “Exfumador”) és 746.
- La freqüència absoluta de “Dona” és 3928.
- La freqüència relativa de “Dona” és $\frac{3928}{7853} = 0,5002$.
- La freqüència relativa de (“Dona”, “Exfumador”) és $\frac{746}{7853} = 0,095$.
- La freqüència relativa de “Exfumador” dins de les dones és de $\frac{746}{3928} = 0,1899$.
- La freqüència relativa de “Dona” dins dels que són “Exfumador” és de $\frac{746}{2120} = 0,3519$.

Representem aquesta taula de contingència amb dos gràfics de barres:

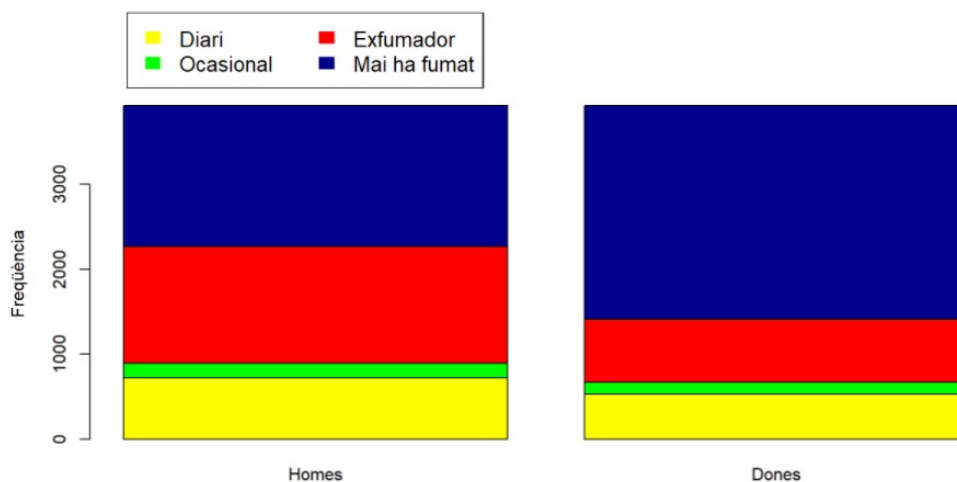


Figura 3.3: Diagrama de barres III

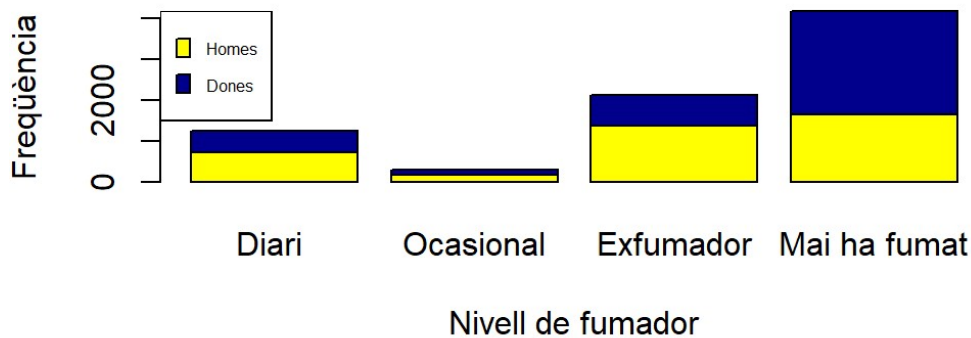


Figura 3.4: Diagrama de barres IV

Es pot comprovar visualment que aproximadament el mateix nombre de dones que homes tenen els mateixos hàbits fumadors.

3 Variables \$PES & \$TALLA

A continuació comprovem si existeix una relació entre el pes i la talla en les mostres recollides en l'estudi. Primer de tot mostrem el diagrama de dispersió de la talla segons el pes acompanyat de la recta de regressió.

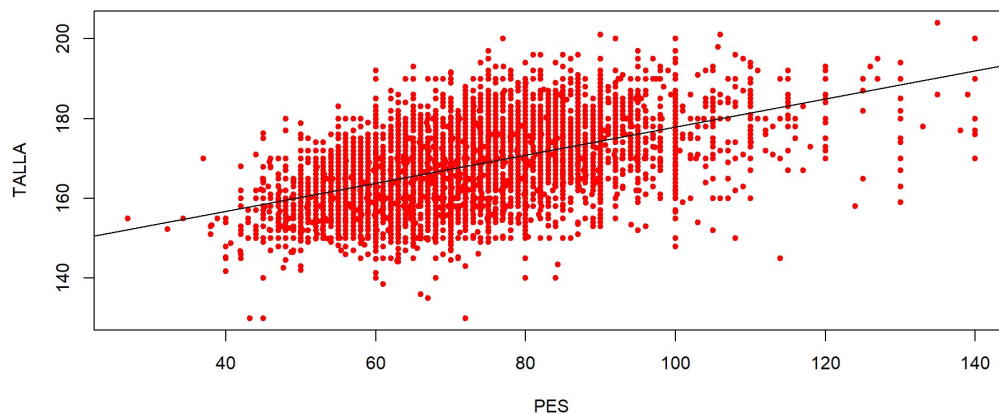


Figura 3.5: Diagrama de dispersió de \$TALLA segons \$PES

Tal i com es pot veure, les dues variables guarden certa relació de dependència lineal, això es veu notablement en el pendent de la recta de regressió. També podem afirmar la dependència que guarden dues variables amb el coeficient de correlació de Pearson, aquest és un índex que mesura la relació lineal entre dues variables quantitatives. A diferència de la covariància, que també la podríem utilitzar per mesurar la relació entre les dues variables, la correlació de Pearson és independent de l'escala de mesura de les variables. El coeficient es calcula en R amb la comanda `cor(x, y)`.

Si executem `cor(excel$PES, excel$TALLA)` el output del programa és `[1] 0.5218075`. La interpretació d'aquest resultat la obtenim de la pròpia definició del coeficient de correlació i de les seves característiques, contra més a prop estigui d'1, més codependents seràn les variables. Llavors és evident que existeix una dependència entre `$TALLA` i `$PES`, però la gran magnitud de mostres recollides augmenta el nombre de marginals i disminueix la correlació global.

4 Variables \$PES & \$IMC

Una altra qüestió que pot sorgir alhora d'estudiar correlacions entre variables d'aquest estudi és si l'IMC augmenta quan augmenta el pes. Segons la definició d'IMC això hauria de ser cert perquè el pes està estrictament relacionat amb el càlcul de l'IMC.

En primer lloc mostrem el diagrama de dispersió juntament amb la recta de regressió respectiva per comprovar visualment la correlació entre variables.

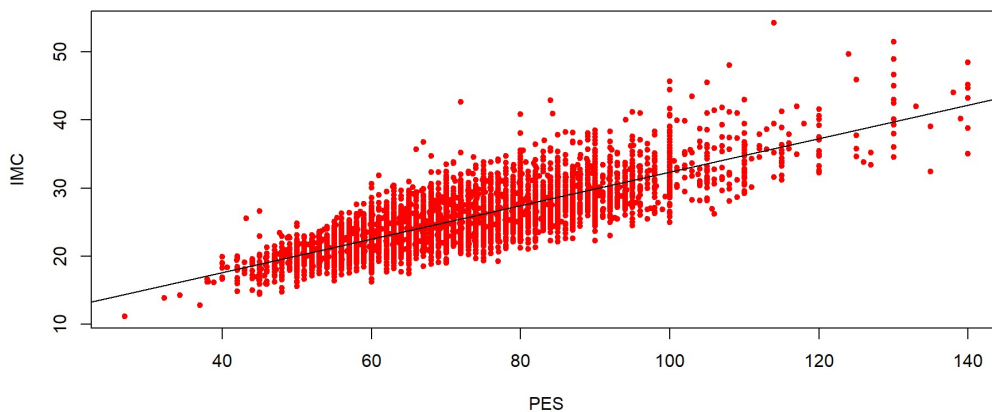


Figura 3.6: Diagrama de dispersió de \$IMC segons \$PES

Es pot veure notablement la dependència lineal que guarden les dues variables. També es veu reflectit en el coeficient de correlació de Pearson que, executant la comanda `cor(excel$PES, excel$IMC)`, l'output del programa és `[1] 0.8052559`. El coeficient està molt més pròxim a 1 que el vist en 3.5, això té sentit ja que les observacions més llunyanes estan situades a menor distància en mitjana de la recta de regressió.

4 CONCLUSIONS

En aquest treball hem pogut treballar en l'àmbit de l'estadística descriptiva, fent tant un estudi univariant com un de bivariant. Ens ha ajudat a treballar d'una forma molt més visual i descriptiva i alhora tenir més en contacte amb el funcionament del llenguatge **R**.

Ens va ser difícil començar ja que no trobàvem cap base de dades que complís totes les condicions de l'enunciat i vam estar buscant bastant temps. Un cop trobada una base de dades prou interessant i amb totes les condicions necessàries per dur a terme l'exercici vam haver de preparar les dades per tal d'estudiar-les correctament amb l'*RStudio*.

Vam tenir un petit dubte amb les dades proporcionades per l'estudi ja que en el document on s'explicaven les variables, no especificaven què volien dir els diferents valors de la variable *TABAC*, i vam haver de posar-nos en contacte amb l'organisme que va dur a terme l'estudi.

En general la realització d'aquest treball ens ha ajudat a tenir un bon contacte i entendre molt millor els conceptes de teoria i el potencial que té el llenguatge **R** en l'àmbit de l'estadística.

5 SCRIPT EN R

```
install.packages("openxlsx") # Comandes amb Excel
library(openxlsx)
install.packages("modeest") # Package per calcular la moda
library(modeest)
install.packages("descr")
library(descr)
install.packages("ggplot2")
library(ggplot2)
install.packages("lattice")
library(lattice)
library("HistogramTools")
install.packages("Hmisc")
library(Hmisc)
install.packages("agricolae")
library(agricolae)

# Passos per carregar el fitxer
# Ens deixarà escollir el fitxer per afegir al nostre data
  frame (my_data)
my_data <- read.xlsx(file.choose(),startRow = 1,detectDates=
  TRUE)
my_data

#### Estudi descriptiu univariant d'aquestes variables####
## Taules de freqüències de les variables ##
# (SEX) #
sex_f_absoluta <- table(my_data$SEX)
sex_f_relativa <- table(my_data$SEX) / length(my_data$SEX)
write.table(sex_f_absoluta, file = "sex_f_absoluta.txt", sep =
  ",", quote = FALSE, row.names = F)
write.table(sex_f_relativa, file = "sex_f_relativa.txt", sep =
  ",", quote = FALSE, row.names = F)

# (EDAT) #
edat_f_absoluta <- table(my_data$EDAT)
edat_f_abs_acumulada <- cumsum(edat_f_absoluta)
edat_f_relativa <- table(my_data$EDAT)/length(my_data$EDAT)
edat_f_rel_acumulada <- cumsum(edat_f_relativa)
write.table(edat_f_absoluta, file = "edat_f_absoluta.txt", sep
  = ",", quote = FALSE, row.names = F)
write.table(edat_f_abs_acumulada, file = "edat_f_abs_acumulada
  .txt", sep = ",", quote = FALSE, row.names = F)
write.table(edat_f_relativa, file = "edat_f_relativa.txt", sep
  = ",", quote = FALSE, row.names = F)
write.table(edat_f_rel_acumulada, file = "edat_f_rel_acumulada
  .txt", sep = ",", quote = FALSE, row.names = F)

# (ESTUDIS) #
```

```

estudis_f_absoluta <- table(my_data$ESTUDIS)
estudis_f_abs_acumulada <- cumsum(estudis_f_absoluta)
estudis_f_relativa <- table(my_data$ESTUDIS)/length(my_data$
  ESTUDIS)
estudis_f_rel_acumulada <- cumsum(estudis_f_relativa)
write.table(estudis_f_absoluta, file = "estudis_f_absoluta.txt",
  sep = ",", quote = FALSE, row.names = F)
write.table(estudis_f_abs_acumulada, file = "estudis_f_abs_
  acumulada.txt", sep = ",", quote = FALSE, row.names = F)
write.table(estudis_f_relativa, file = "estudis_f_relativa.txt",
  sep = ",", quote = FALSE, row.names = F)
write.table(estudis_f_rel_acumulada, file = "estudis_f_rel_
  acumulada.txt", sep = ",", quote = FALSE, row.names = F)

# (TABAC) #
tabac_f_absoluta <- table(my_data$TABAC)
tabac_f_abs_acumulada <- cumsum(tabac_f_absoluta)
tabac_f_relativa <- table(my_data$TABAC)/length(my_data$TABAC)
tabac_f_rel_acumulada <- cumsum(tabac_f_relativa)
write.table(tabac_f_absoluta, file = "tabac_f_absoluta.txt",
  sep = ",", quote = FALSE, row.names = F)
write.table(tabac_f_abs_acumulada, file = "tabac_f_abs_
  acumulada.txt", sep = ",", quote = FALSE, row.names = F)
write.table(tabac_f_relativa, file = "tabac_f_relativa.txt",
  sep = ",", quote = FALSE, row.names = F)
write.table(tabac_f_rel_acumulada, file = "tabac_f_rel_
  acumulada.txt", sep = ",", quote = FALSE, row.names = F)

## Diagrama de barres per les variables (ESTUDIS) i (TABAC) ##
estudis <- c("Sense estudis", "Basic", "Mitja", "Superior")
barplot(estudis_f_absoluta, main = "Diagrama de barres $
  ESTUDIS", xlab = "NIVELL ESTUDIS", names.arg = estudis, col
    = "blue")

tabac <- c("Diari", "Ocasional", "Exfumador", "Mai fumador")
barplot(tabac_f_absoluta, main = "Diagrama de barres $TABAC",
  xlab = "HABIT", names.arg = tabac, col = "blue")

## Diagrama circular per les variables (ESTUDIS) i (TABAC) ##
piepercent_estudis <- round(100*estudis_f_absoluta/sum(estudis
  _f_absoluta),1)
pie(estudis_f_absoluta, labels = piepercent_estudis, main = "
  Diagrama de sectors $ESTUDIS", col = rainbow(length(estudis
    _f_absoluta)))
legend("topright", estudis, fill = rainbow(length(estudis_f_
  absoluta)))

piepercent_tabac <- round(100*tabac_f_absoluta/sum(tabac_f_
  absoluta),1)
pie(tabac_f_absoluta, labels = piepercent_tabac, main = "

```

```

Diagrama de sectors $TABAC", col = rainbow(length(tabac_f_
absoluta)))
legend("topright", tabac, fill = rainbow(length(tabac_f_
absoluta)))

## Histogrames de les variables (EDAT), (PES), (TALLA)##
hist(my_data$EDAT, main = "Histograma $EDAT", xlab = "EDAT",
      ylab = "Frequencia absoluta", col = "green")

my_hist_f_abs_acum <- hist(my_data$EDAT)
my_hist_f_abs_acum$counts <- cumsum(my_hist_f_abs_acum$counts)
plot(my_hist_f_abs_acum, xaxp=c(20, 100, 10), col="blue", xlab
     ="EDAT", ylab="Frequencia absoluta acumulada", main = "
Histograma $EDAT")

histogram(my_data$EDAT, aspect= "fill", main = "Histograma $
EDAT", xlab="EDAT", ylab = "Frequencia relativa (%)")

hist(my_data$PES, main = "Histograma $PES", xlab = "PES (Kg)",
      ylab = "Frequencia absoluta", col = "green")

hist(my_data$TALLA, main = "Histograma $TALLA", xlab = "TALLA
(cm)", ylab = "Frequencia absoluta", col = "green")

## Boxplot de les variables (TABAC) i (E_SALUT) segons la
variable (SEX) ##
df_tabac <- data.frame(my_data$TABAC, my_data$SEX)
tabac_homes <- df_tabac$my_data.TABAC[df_tabac$my_data.SEX==1]
tabac_dones <- df_tabac$my_data.TABAC[df_tabac$my_data.SEX==2]
boxplot(tabac_homes)
boxplot(tabac_dones)

df_talla <- data.frame(my_data$TALLA, my_data$SEX)
talla_homes <- df_talla$my_data.TALLA[df_talla$my_data.SEX==1]
talla_dones <- df_talla$my_data.TALLA[df_talla$my_data.SEX==2]

## Estudi descriptiu de la variable $PES -pes en kg-
distingint segons la variable $SEX -home i dona-. Recollim
els estadístics numerics ##

# Mitjana #
mean(my_data$PES) # Tots
df1 <- data.frame(my_data$SEX, my_data$PES)
aggregate(my_data$PES~my_data$SEX, data = df1, FUN = mean) #
Homes & dones

# Mediana #
median(my_data$PES) # Tots
aggregate(my_data$PES~my_data$SEX, data = df1, FUN = median) #
Homes & dones

```

```
# Moda #
mfv(my_data$PES) # Tots
aggregate(my_data$PES~my_data$SEX, data = df1, FUN = mfv) #
  Homes & dones

# Desviació #
sd(my_data$PES) # Tots
aggregate(my_data$PES~my_data$SEX, data = df1, FUN = sd) #
  Homes & dones

# Variancia #
var(my_data$PES) # Tots
aggregate(my_data$PES~my_data$SEX, data = df1, FUN = var) #
  Homes & dones

# Coeficient de variació #
coef_var <- sd(my_data$PES)/mean(my_data$PES) * 100 # Tots
coef_var_homes <- 12.87477/79.16195 * 100
coef_var_dones <- 12.32932/65.37478 * 100
coef_var
coef_var_homes
coef_var_dones

# Q1 & Q3#
quantile(my_data$PES) # Tots
aggregate(my_data$PES~my_data$SEX, data = df1, FUN = quantile)
  # Homes & dones

# Mínim #
# Quartil 0 és el mínim

# Màxim #
# Quartil 4 és igual al màxim

# Rang #
# Rang = max(x1, ..., xn) - min(x1, ..., xn)
rang_tots <- quantile(my_data$PES, 1) - 26.9431 # Tots (rang)
rang_homes <- 140 - 44
rang_dones <- 140 - 26.9431
rang_tots
rang_homes
rang_dones

# Rang interquartílic #
# IQR = Q3 - Q1
Q1 <- quantile(my_data$PES, 0.25)
Q3 <- quantile(my_data$PES, 0.75)
IQR <- Q3 - Q1
IQR # Tots
```

```

Q1_homes <- 70
Q3_homes <- 86
Q1_dones <- 57
Q3_dones <- 72

IQR_homes <- Q3_homes - Q1_homes
IQR_dones <- Q3_dones - Q1_dones
IQR_homes
IQR_dones

# Skewness #
skewness(my_data$PES) # Tots
aggregate(my_data$PES~my_data$SEX, data = df1, FUN = skewness)
# Homes & dones

# Curtosis #
kurtosis(my_data$PES) # Tots
aggregate(my_data$PES~my_data$SEX, data = df1, FUN = kurtosis)
# Homes & dones
hist(my_data$PES, freq=FALSE)
lines(density(my_data$PES), col="red", lwd=3)
abline(v=c(mean(my_data$PES), median(my_data$PES)), col=c("
green", "blue"), lty=c(2,2), lwd=c(3,3))

# Coeficient de correlació de Pearson #
cor(my_data$PES, my_data$TALLA)
cor(my_data$PES, my_data$IMC)

# Grafics #
df_pes <- data.frame(my_data$PES, my_data$SEX) # Data frame
# per guardar el pes i el sexe
pes_homes <- df_pes$my_data.PES[df_pes$my_data.SEX==1]
pes_dones <- df_pes$my_data.PES[df_pes$my_data.SEX==2]

hist(pes_homes, main = "Histograma homes $PES", xlab = "PES (
Kg)", ylab = "Frequencia absoluta", col = "green")

hist(pes_dones, main = "Histograma dones $PES", xlab = "PES (
Kg)", ylab = "Frequencia absoluta", col = "green")

mynames <- c("homes", "dones")
boxplot(pes_homes, pes_dones, main="Boxplot homes/dones $PES",
xlab="SEX", ylab="PES", col="yellow", border="brown",
names=mynames)

#### Estudi bivariant, utilitzant el model lineal, per les
observacions de dues variables del fitxer ####
## Taula de contingència de les variables $SEX i $ESTUDIS ##
df_estudis <- data.frame(my_data$ESTUDIS, my_data$SEX)

```

```

estudis_homes <- df_estudis$my_data.ESTUDIS[df_estudis$my_data
  .SEX==1]
estudis_dones <- df_estudis$my_data.ESTUDIS[df_estudis$my_data
  .SEX==2]

table(unlist(estudis_homes))
table(unlist(estudis_dones))

cole = c("yellow", "darkblue", "red", "green")
x_text = c("Sense estudis", "Basic", "Mitja", "Superior")
matrix1 <- matrix(c(139,1704,1412,670,189,1636,1280,823), nrow
  =4, dimnames= list(rownames, colnames))
rownames = c("Sense estudis", "Basic", "Mitja", "Superior")
colnames = c("homes", "dones")
legend("top", x_text, fill = cole)
barplot(matrix1, ylab="Frequencia", col = cole, beside=FALSE)

contin1 <- table(excel$ESTUDIS,excel$SEX) # Taula de
  contingencia

barplot(contin1, col = c("yellow","green", "red", "darkblue"),
  bg = "red", names.arg = c("Homes", "Dones"), ylab = "
  Frequencia")

# Llegenda pel grafic anterior #
legend(legend = c("Sense estudis","Basic", "Mitja", "Superior"
  ), fill = c("yellow","darkblue", "red", "green"), x = "
  center", bg="transparent", cex = 0.8, ncol = 1)

barplot(contin1, col = c("red","orange"), bg = "red", names.
  arg = c("Sense estudis", "Basic", "Mitja", "Superior"),
  ylab = "Frequencia")

# Llegenda pel grafic anterior #
legend(legend = c("Homes", "Dones"), fill = c("red","orange"),
  x = "topright", bg="transparent", ncol = 1)

contin2 <- table(excel$TABAC,excel$SEX) # Taula de
  contingencia

barplot(ta, col = c("yellow","darkblue"), bg = "red", names.
  arg = c("Diari ", "Ocasional", "Exfumador", "Mai ha fumat")
  , ylab = "Frequencia", xlab = "Nivell de fumador")

# Llegenda pel grafic anterior #
legend(legend = c("Homes", "Dones"), fill = c("yellow","
  darkblue"), x = "topleft", bg="transparent", ncol = 1)

barplot(contin2, col = c("yellow","green", "red", "darkblue"),
  bg = "red", names.arg = c("Homes", "Dones"), ylab = "

```

```
    Freqüencia")

# Llegendes pel gràfic anterior #
legend(legend = c("Diari","Ocasional", "Exfumador", "Mai ha
    fumat"), fill = c("yellow","green", "red", "darkblue"), x =
    "topleft", bg="transparent", cex = 0.8, ncol = 2)

# Diagrames de dispersió #
plot(my_data$PES,my_data$TALLA, xlab="PES", ylab="TALLA", col=
    "red", pch=20)
x<-lm(formula = excel$TALLA~excel$PES) # Recta de regressió
abline(x)

plot(my_data$PES,my_data$IMC, xlab="PES", ylab="IMC", col="red
    ", pch=20)
x<-lm(formula = excel$TALLA~excel$IMC) # Recta de regressió
abline(x)
```

BIBLIOGRAFIA

- [1] Servizo Galego de Saúde. *SICRI: Sistema de Información sobre Conductas de Riesgo*. [Online; Recuperat el 5 de març de 2022]. 2018. URL: <https://www.sergas.es/Saude-publica/SICRI-Sistema-Informaci%C3%B3n-sobre-Conducitas-de-Risco?idioma=es>.
- [2] Servizo Galego de Saúde. *Ficha metodoloxica SICRI-2018*. [Online; Recuperat el 5 de març de 2022]. 2018. URL: https://www.sergas.es/Saude-publica/Documents/6437/Metodoloxia_SICRI-2018.pdf.
- [3] Servizo Galego de Saúde. *SICRI-2018; Resultados*. [Online; Recuperat el 5 de març de 2022]. 2018. URL: https://www.sergas.es/Saude-publica/Documents/6439/NunhaOllada_SICRI-2018.pdf.
- [4] Servizo Galego de Saúde. *SICRI-2018; Microdatos*. [Online; Recuperat el 5 de març de 2022]. 2018. URL: <https://www.sergas.es/Saude-publica/SICRI-2018-Microdatos>.