Actividad 2

Sergio Rodero Casado

22/1/2022

Lectura/Escritura de ficheros de datos

En primer lugar se definen los directorios de trabajo.

```
workingDir <- "C:/Users/sergi/Desktop/Code/Neurocomputacion/Actividad_2"
setwd(workingDir)
```

Se realiza la lectura del fichero.

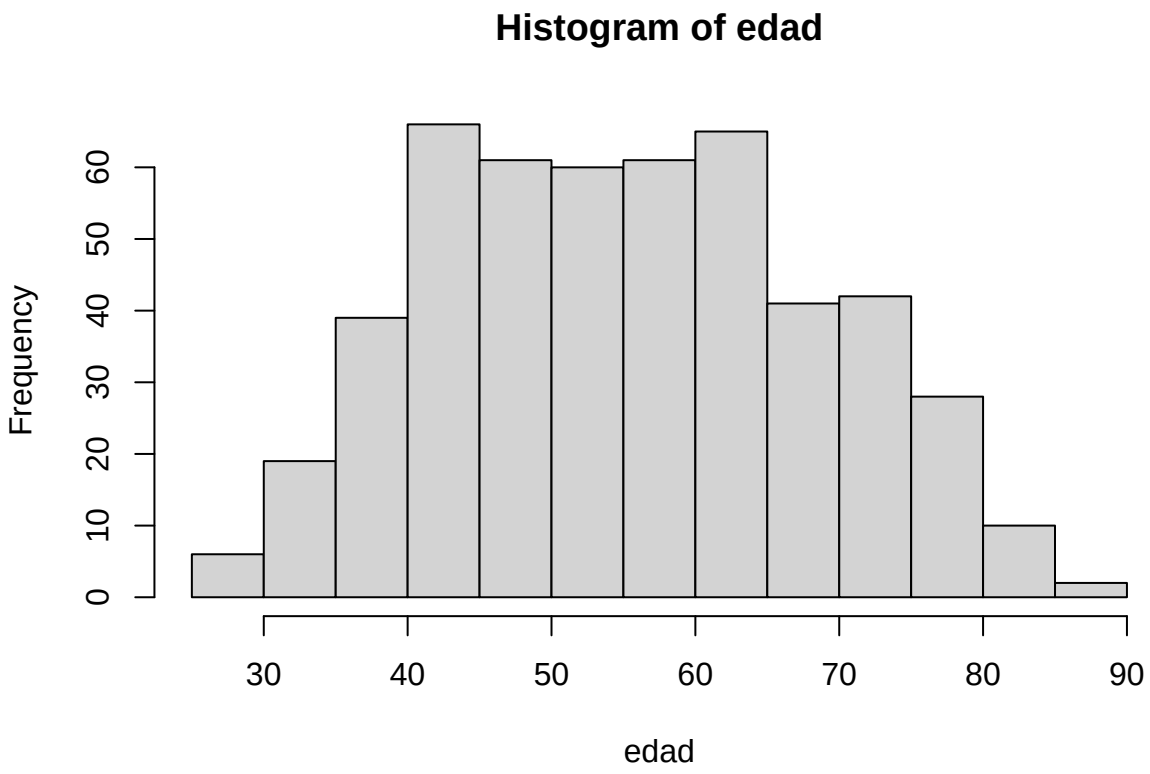
```
datos <- read.table(file=paste(workingDir,"datos_icb.txt",sep="/"), header=T, sep=" ", dec=".")


edad <- datos$edad
tam <- datos$tam
grado <- as.factor(datos$grado)
gang <- datos$gang
feno <- as.factor(datos$feno)
quim <- as.factor(datos$quim)
horm <- as.factor(datos$horm)
recid <- datos$recid
```

Estadística descriptiva

```
# Edad
summary(edad)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   25.00   45.00   55.00   55.76   65.00   88.00
```

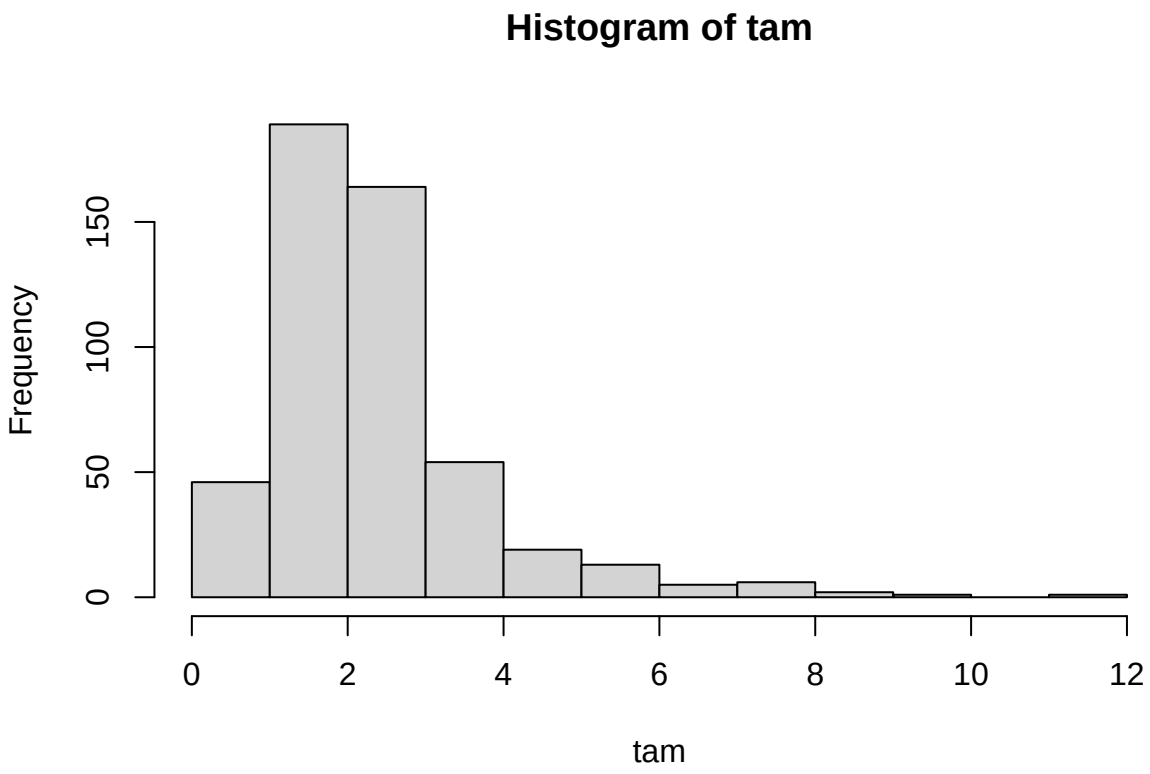
```
hist(edad)
```




```
# Tamaño
summary(tam)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.200   1.500   2.250   2.524   3.000  12.000
```

```
hist(tam)
```



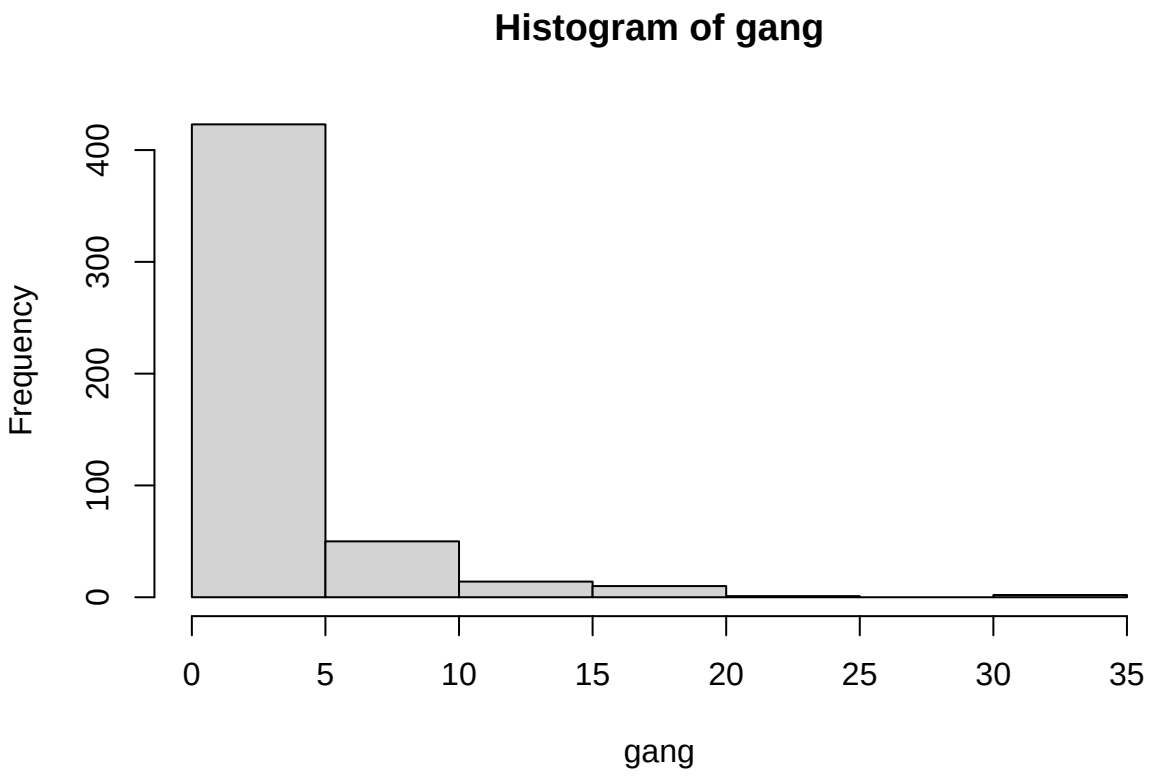

```
# Grado
summary(grado)
```

```
##  G1  G2  G3
##  76 279 145
```

```
# Ganglio
summary(gang)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   1.000   2.478   3.000  34.000
```

```
hist(gang)
```




```
# Feno
summary(feno)
```

```
##    Basal like HER2 enriched  Luminal-HER2     Luminal A     Luminal B
##            64             31             21           240           120
##   TN no-basal
##             24
```

```
# Quimioterapia
summary(quim)
```

```
##  No Yes
## 176 324
```

```
# Horm
summary(horm)
```

```
##  No Yes
## 154 346
```

```
# Recidiva
summary(recid)
```

```
##    Length     Class      Mode
##       500 character character
```

Modelo completo

En primer lugar se realiza la estimación del modelo de regresión logística usando la función *glm*

```
rlog <- glm(as.factor(datos$recid) ~ edad + tam + grado + gang + feno + quim + horm,data = datos,family
```

Una vez obtenido el modelo, se calcula la precisión de este mediante el cómputo del ACC aparente.

```
train <- datos[1:7]
prediction <- predict(rlog,train,type = "response")
```

Se realiza la regla de decisión.

```
prediction[prediction<0.5] <- 0
prediction[prediction>=0.5] <- 1
recid[recid=="NO"] <- 0
recid[recid=="SI"] <- 1
```

Comparación de lo valores reales con los de la predicción para obtener la precisión

```
rate <- (sum(recid == prediction)/length(recid))*100

print(rate)
```

```
## [1] 85.4
```

Modelos alternativos

Primer modelo

```
rlog1 <- glm(as.factor(datos$recid) ~ edad + tam,data = datos,family = binomial("logit"))

train1 <- data.frame(edad,tam)
prediction1 <- predict(rlog1,train1,type = "response")

prediction1[prediction1<0.5] <- 0
prediction1[prediction1>=0.5] <- 1

rate1 <- (sum(recid == prediction1)/length(recid))*100

print(rate1)
```

```
## [1] 84.2
```

El resultado obtenido es menor que el modelo completo.

Segundo modelo.

```
rlog2 <- glm(as.factor(datos$recid) ~ tam + quim + feno,data = datos,family = binomial("logit"))

train2 <- data.frame(tam,quim,feno)
prediction2 <- predict(rlog2,train2,type = "response")

prediction2[prediction2<0.5] <- 0
prediction2[prediction2>=0.5] <- 1

rate2 <- (sum(recid == prediction2)/length(recid))*100

print(rate2)
```

```
## [1] 85
```

El resultado obtenido es ligeramente menor que el modelo completo.

Tercer modelo

```
rlog3 <- glm(as.factor(datos$recid) ~ edad + grado + quim,data = datos,family = binomial("logit"))

train3 <- data.frame(edad,grado,quim)
prediction3 <- predict(rlog3,train3,type = "response")

prediction3[prediction3<0.5] <- 0
prediction3[prediction3>=0.5] <- 1

rate3 <- (sum(recid == prediction3)/length(recid))*100

print(rate3)
```

```
## [1] 84.8
```

El resultado obtenido es menor que el modelo completo.

Cuarto modelo

```
rlog4 <- glm(as.factor(datos$recid) ~ gang + grado + feno,data = datos,family = binomial("logit"))

train4 <- data.frame(edad,tam)
prediction4 <- predict(rlog4,train4,type = "response")

prediction4[prediction4<0.5] <- 0
prediction4[prediction4>=0.5] <- 1

rate4 <- (sum(recid == prediction4)/length(recid))*100

print(rate4)
```

```
## [1] 85.8
```

El resultado obtenido es mayor que el modelo completo.

Quinto modelo

```
rlog5 <- glm(as.factor(datos$recid) ~ edad + gang + feno + quim,data = datos,family = binomial("logit"))

train5 <- data.frame(edad,gang,feno,quim)
prediction5 <- predict(rlog5,train5,type = "response")

prediction5[prediction5<0.5] <- 0
prediction5[prediction5>=0.5] <- 1

rate5 <- (sum(recid == prediction5)/length(recid))*100

print(rate5)
```

```
## [1] 85.6
```

El resultado obtenido es ligeramente mayor que el modelo completo.