

Actividad-1.R

rstudio-user

2021-10-06

```
#####  
#           Actividad 1  
#####  
rm(list=ls()) #limpia el workspace  
  
datos <- read.table("ejercicio1.txt", header = TRUE)  
head(datos)  
  
##   Genero Edad Estatura Nota  Calificacion  
## 1  Mujer   21    1.82    5      Aprobado  
## 2 Hombre   19    1.83    6      Aprobado  
## 3  Mujer   18    1.78    7      Notable  
## 4  Mujer   20    1.79    5      Aprobado  
## 5  Mujer   23    1.80    9 Sobresaliente  
## 6 Hombre   22    1.90    7      Notable  
  
#1)  
tail(datos, n = 10)  
  
##   Genero Edad Estatura Nota  Calificacion  
## 11 Mujer   22    1.73    5      Aprobado  
## 12 Hombre  20    1.79    8      Notable  
## 13 Hombre  22    1.80    7      Notable  
## 14 Mujer   19    1.77    6      Aprobado  
## 15 Mujer   19    1.69    3      Suspenso  
## 16 Hombre  21    1.75    4      Suspenso  
## 17 Mujer   20    1.66    5      Aprobado  
## 18 Mujer   21     NA    6      Aprobado  
## 19 Mujer   22    1.79    2      Suspenso  
## 20 Mujer   23    1.80    8      Notable  
  
#tail(datos, 10)  
  
#2)  
str(datos)  
  
## 'data.frame':   20 obs. of  5 variables:  
## $ Genero      : chr  "Mujer" "Hombre" "Mujer" "Mujer" ...  
## $ Edad        : int   21 19 18 20 23 22 22 20 21 21 ...  
## $ Estatura    : num   1.82 1.83 1.78 1.79 1.8 1.9 1.79 1.83 NA 1.65 ...  
## $ Nota        : int    5 6 7 5 9 7 8 3 9 2 ...  
## $ Calificacion: chr   "Aprobado" "Aprobado" "Notable" "Aprobado" ...
```

```
summary(datos)
```

```
##      Genero      Edad      Estatura      Nota
## Length:20      Min.    :18.0      Min.    :1.650      Min.    :2.00
## Class :character 1st Qu.:20.0      1st Qu.:1.755      1st Qu.:4.75
## Mode  :character Median :21.0      Median :1.790      Median :6.00
##              Mean  :20.8      Mean  :1.776      Mean  :5.75
##              3rd Qu.:22.0      3rd Qu.:1.800      3rd Qu.:7.25
##              Max.   :23.0      Max.   :1.900      Max.   :9.00
##              NA's    :2
## Calificacion
## Length:20
## Class :character
## Mode  :character
##
##
##
##
```

```
dim(datos)
```

```
## [1] 20  5
```

```
class(datos$Genero)
```

```
## [1] "character"
```

```
#3)
```

```
datos$Estatura <- as.numeric(datos$Estatura)
colMeans(datos[,2:4], na.rm = TRUE)
```

```
##      Edad  Estatura      Nota
## 20.800000  1.776111  5.750000
```

```
#4)
```

```
suspensos <- datos[datos$Calificacion == "Suspendido",]
suspensos <- datos[which(datos$Calificacion == "Suspendido"),]
suspensos <- subset(datos, Calificacion == "Suspendido")
suspensos <- datos[datos$Nota < 5,]
```

```
dim(suspensos)
```

```
## [1] 5 5
```

```
write.table(suspensos, file = "suspensos.txt")
write.csv(suspensos, file = "suspensos.csv")
write.csv2(suspensos, file = "suspensos2.csv")
```

```
#5)
```

```
library(modeest)
mfv(datos$Edad)
```

```

## [1] 21 22
mfv(datos$Estatura, na_rm = TRUE)
## [1] 1.79
mfv(datos$Nota)
## [1] 5
table(datos$Edad)
##
## 18 19 20 21 22 23
##  1  3  4  5  5  2

moda_Edad <-
as.numeric(names(which(table(datos$Edad)==max(table(datos$Edad)))))
moda_Edad
## [1] 21 22
table(datos$Estatura)
##
## 1.65 1.66 1.69 1.73 1.75 1.77 1.78 1.79  1.8 1.82 1.83  1.9
##    1    1    1    1    1    1    1    4    3    1    2    1

moda_Estatura <-
as.numeric(names(which(table(datos$Estatura)==max(table(datos$Estatura)))))
moda_Estatura
## [1] 1.79
table(datos$Nota)
##
## 2 3 4 5 6 7 8 9
## 2 2 1 4 3 3 3 2

moda_Nota <-
as.numeric(names(which(table(datos$Nota)==max(table(datos$Nota)))))
moda_Nota
## [1] 5

#6)
datos_nota <- datos[order(datos$Nota),]
#datos_nota <- datos[with(datos, order(datos$Nota)),]
head(datos_nota)

##      Genero Edad Estatura Nota Calificacion
## 10  Mujer   21    1.65    2      Suspenso
## 19  Mujer   22    1.79    2      Suspenso

```

```
## 8 Hombre 20 1.83 3 Suspenso
## 15 Mujer 19 1.69 3 Suspenso
## 16 Hombre 21 1.75 4 Suspenso
## 1 Mujer 21 1.82 5 Aprobado
```

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
datos_nota2 <- arrange(datos, datos$Edad)
```

```
datos_nota2
```

```
## Genero Edad Estatura Nota Calificacion
## 1 Mujer 18 1.78 7 Notable
## 2 Hombre 19 1.83 6 Aprobado
## 3 Mujer 19 1.77 6 Aprobado
## 4 Mujer 19 1.69 3 Suspenso
## 5 Mujer 20 1.79 5 Aprobado
## 6 Hombre 20 1.83 3 Suspenso
## 7 Hombre 20 1.79 8 Notable
## 8 Mujer 20 1.66 5 Aprobado
## 9 Mujer 21 1.82 5 Aprobado
## 10 Hombre 21 NA 9 Sobresaliente
## 11 Mujer 21 1.65 2 Suspenso
## 12 Hombre 21 1.75 4 Suspenso
## 13 Mujer 21 NA 6 Aprobado
## 14 Hombre 22 1.90 7 Notable
## 15 Mujer 22 1.79 8 Notable
## 16 Mujer 22 1.73 5 Aprobado
## 17 Hombre 22 1.80 7 Notable
## 18 Mujer 22 1.79 2 Suspenso
## 19 Mujer 23 1.80 9 Sobresaliente
## 20 Mujer 23 1.80 8 Notable
```

```
#7)
```

```
absolutas <- table(datos$Calificacion)
```

```
absolutas
```

```
##
```

```
## Aprobado Notable Sobresaliente Suspenso
```

```
## 7 6 2 5
```

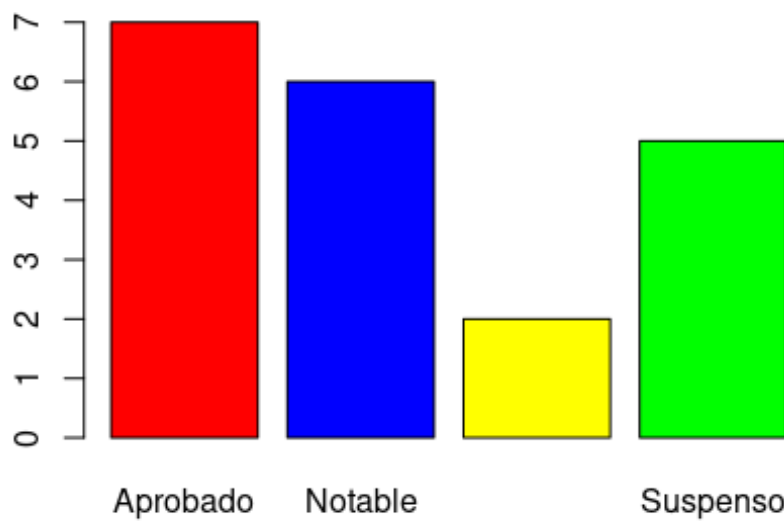
```
relativas <- prop.table(absolutas)
relativas
```

```
##
##      Aprobado      Notable Sobresaliente      Suspenso
##      0.35      0.30      0.10      0.25
```

```
#8)
```

```
barplot(absolutas, col = c("red", "blue", "yellow", "green"), main =
"Diagrama de barras para la variable Calificación")
```

Diagrama de barras para la variable Calificación



```
pie(absolutas, col = c("red", "blue", "yellow", "green"), main = "Diagrama de
barras para la variable Calificación")
```

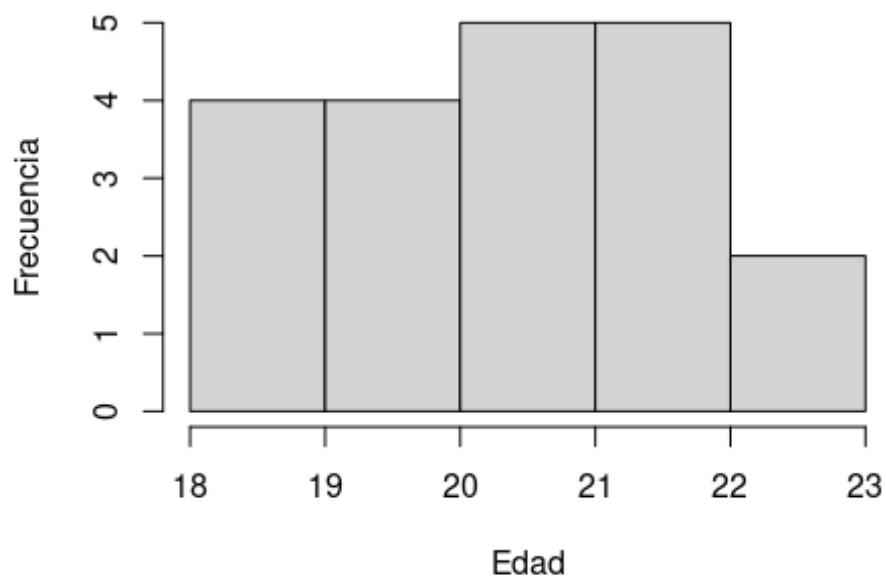
Diagrama de barras para la variable Calificación



#9)

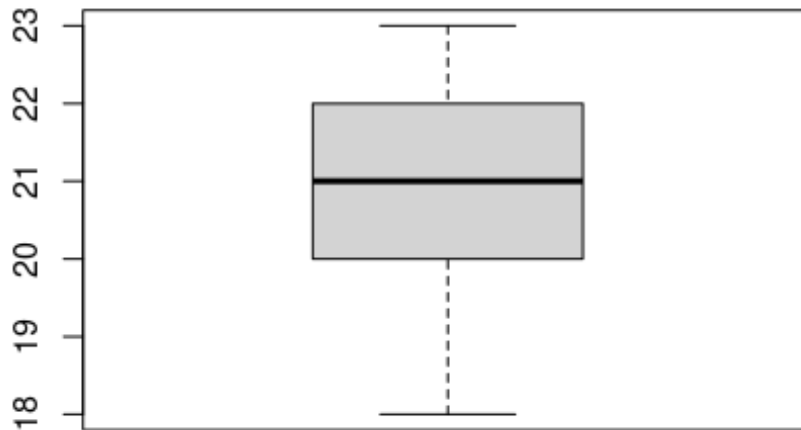
```
hist(datos$Edad, main = "Histograma para la variable Edad", xlab = "Edad",  
ylab = "Frecuencia")
```

Histograma para la variable Edad



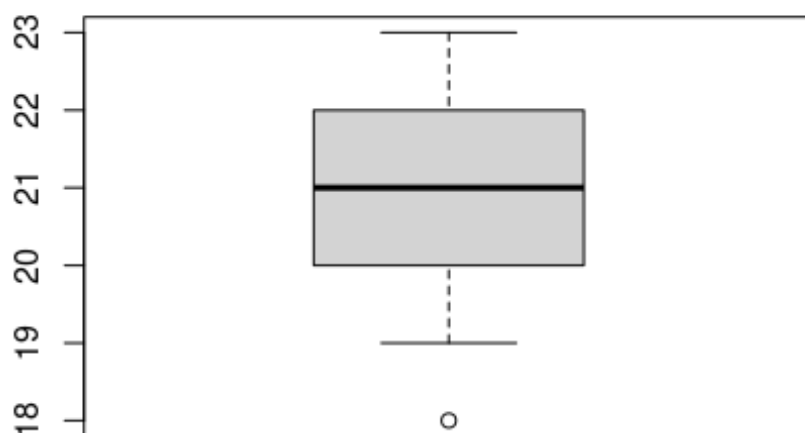
```
boxplot(datos$Edad, main = "Cajas y bigotes para la variable Edad (range = 1.5)")
```

Cajas y bigotes para la variable Edad (range = 1.5)



```
b <- boxplot(datos$Edad, range = 0.5, main = "Cajas y bigotes para la variable Edad (range = 0.5)")
```

Cajas y bigotes para la variable Edad (range = 0.5)



```
atipico <- b$out
atipico

## [1] 18

#10)
summary(datos$Nota)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      2.00   4.75   6.00   5.75   7.25   9.00

min(datos$Nota)

## [1] 2

max(datos$Nota)

## [1] 9

quantile(datos$Nota, probs = c(0.25, 0.75))

## 25% 75%
## 4.75 7.25

mean(datos$Nota)

## [1] 5.75

median(datos$Nota)
```



```
## [1] 6

#11)
mean(datos$Estatura, na.rm = TRUE)

## [1] 1.776111

#Como medidas de dispersión, se van a calcular la varianza y el recorrido intercuartílico.
var(datos$Estatura, na.rm = TRUE)

## [1] 0.003836928

IQR(datos$Estatura, na.rm = TRUE)

## [1] 0.045

#12)
#Para determinar la homogeneidad de una variable (o, lo que es lo mismo, la representatividad de su media), calculamos el Coeficiente de Variación de Pearson para cada una de ellas, el cual se define como el cociente entre la desviación típica y la media de la variable.
CV_Edad <- sd(datos$Edad)/mean(datos$Edad) #coef. de variación para la edad
CV_Edad

## [1] 0.06727153

CV_Estatura <- sd(datos$Estatura, na.rm = TRUE)/mean(datos$Estatura, na.rm = TRUE)
CV_Estatura

## [1] 0.0348756

#La variable más homogénea es la Estatura, ya que presenta un coeficiente de variación más próximo a 0.

#13)
library(moments)

##
## Attaching package: 'moments'

## The following object is masked from 'package:modeest':
##
##      skewness

asimetría_Edad <- skewness(datos$Edad, na.rm = TRUE)
asimetría_Edad

## [1] -0.2270665
```

```

asimetría_Estatura <- skewness(datos$Estatura, na.rm = TRUE)
asimetría_Estatura

## [1] -0.5029718

asimetría_Nota <- skewness(datos$Nota, na.rm = TRUE)
asimetría_Nota

## [1] -0.2386602

curtosis_Edad <- kurtosis(datos$Edad, na.rm = TRUE)
curtosis_Edad

## [1] 2.194242

curtosis_Estatura <- kurtosis(datos$Estatura, na.rm = TRUE)
curtosis_Estatura

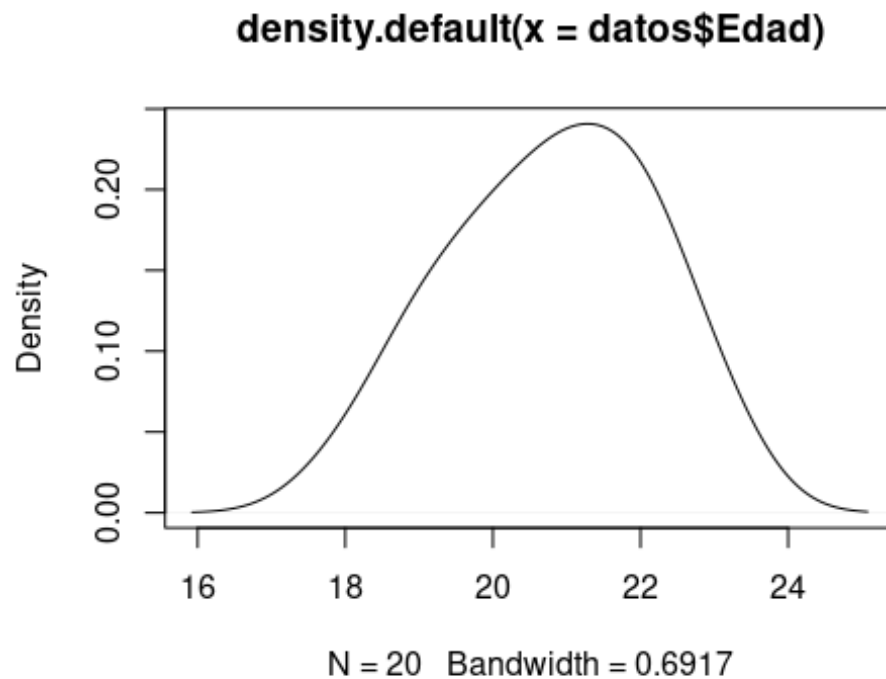
## [1] 3.18122

curtosis_Nota <- kurtosis(datos$Nota, na.rm = TRUE)
curtosis_Nota

## [1] 2.055578

plot(density(datos$Edad))

```

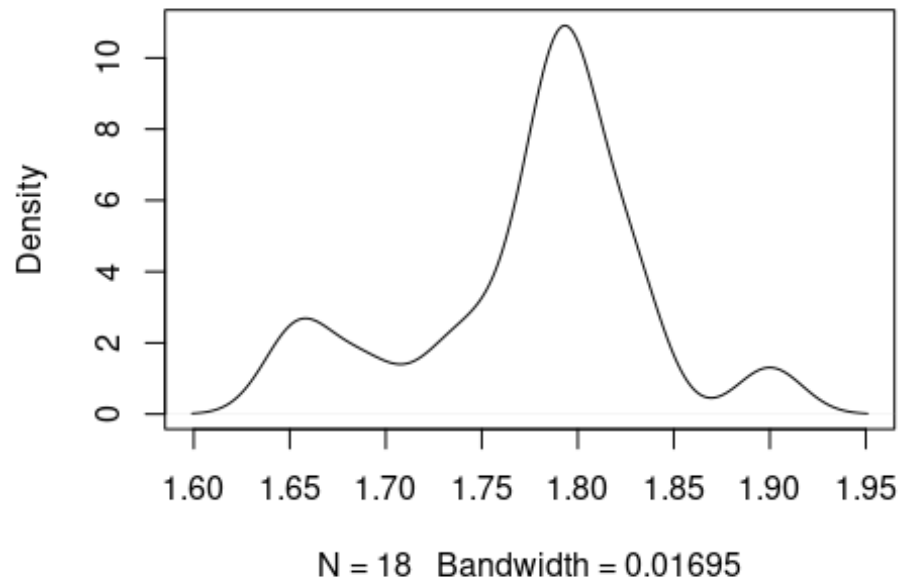


```

plot(density(datos$Estatura, na.rm = TRUE))

```

density.default(x = datos\$Estatura, na.rm = TRUE



```
plot(density(datos$Nota, na.rm = TRUE))
```

```
library(nortest)
```

```
shapiro.test(datos$Edad)
```

```
##
```

```
##  Shapiro-Wilk normality test
```

```
##
```

```
## data:  datos$Edad
```

```
## W = 0.9435, p-value = 0.279
```

```
shapiro.test(datos$Estatura)
```

```
##
```

```
##  Shapiro-Wilk normality test
```

```
##
```

```
## data:  datos$Estatura
```

```
## W = 0.91723, p-value = 0.1154
```

```
shapiro.test(datos$Nota)
```

```
##
```

```
##  Shapiro-Wilk normality test
```

```
##
```

```
## data:  datos$Nota
```

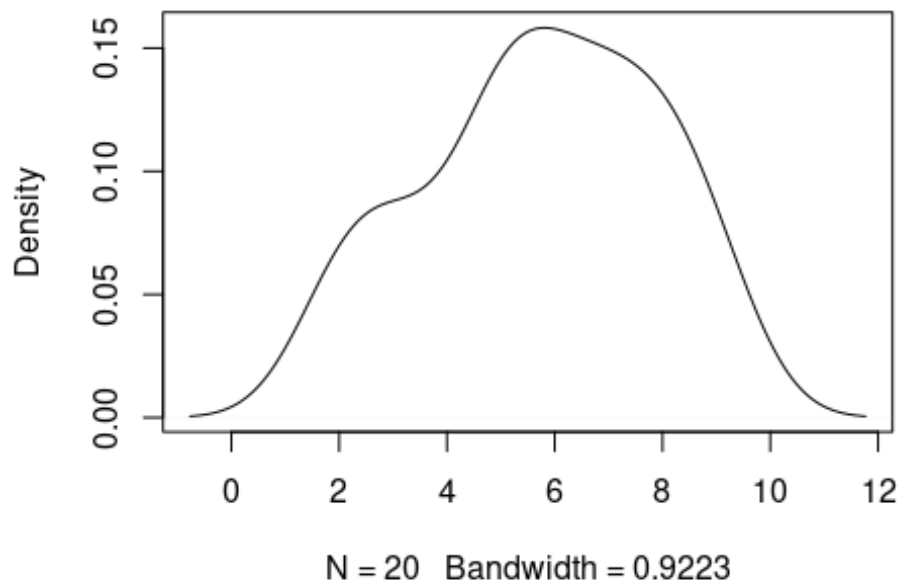
```
## W = 0.94617, p-value = 0.3127
```

Si $p < 0.05$ se rechaza la hipótesis de normalidad

```
library(MVN)

## Registered S3 method overwritten by 'psych':
##   method      from
## plot.residuals rmutil
```

density.default(x = datos\$Nota, na.rm = TRUE)



```
matriz <- datos[,2:4]
mvn(matriz)

## $multivariateNormality
##           Test      HZ    p value MVN
## 1 Henze-Zirkler 0.5141128 0.5972213 YES
##
## $univariateNormality
##           Test Variable Statistic    p value Normality
## 1 Anderson-Darling   Edad      0.5114    0.1696      YES
## 2 Anderson-Darling Estatura    0.7724    0.0362      NO
## 3 Anderson-Darling   Nota      0.3625    0.4032      YES
##
## $Descriptives
##           n      Mean   Std.Dev Median   Min  Max   25th 75th      Skew
## Edad      18 20.777778 1.47750010  21.00 18.00 23.0 20.000 22.0 -0.1560831
## Estatura  18  1.776111 0.06194294   1.79  1.65  1.9  1.755  1.8 -0.4616451
## Nota      18  5.555556 2.14811058   5.50  2.00  9.0  4.250  7.0 -0.1895934
##           Kurtosis
```

```

## Edad      -1.2407723
## Estatura  -0.1624299
## Nota      -1.2461496

#mvn(matriz, univariatePlot = "histogram")
#mvn(matriz, multivariatePlot = "qq")
#n=18!!!!

#14)
cor(datos[,2:4])

##              Edad  Estatura      Nota
## Edad      1.0000000      NA 0.2422921
## Estatura      NA      1      NA
## Nota      0.2422921      NA 1.0000000

datos_corr <- na.omit(datos)
cor(datos_corr[,2:4])

##              Edad  Estatura      Nota
## Edad      1.0000000 0.2085317 0.2450594
## Estatura 0.2085317 1.0000000 0.4548530
## Nota      0.2450594 0.4548530 1.0000000

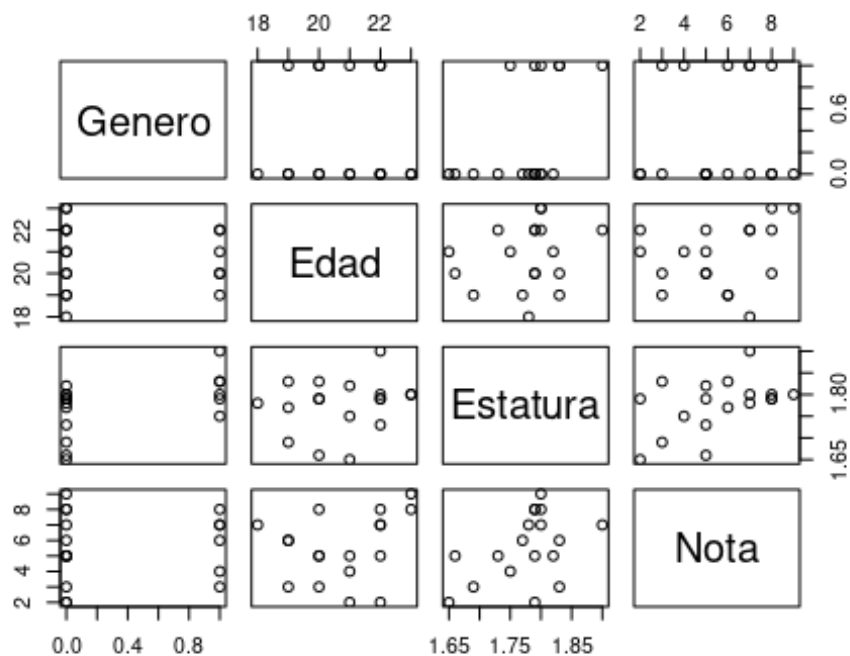
datos_corr$Genero <- ifelse(datos_corr$Genero == "Mujer", 0, 1)

cor(datos_corr[,1:4])

##              Genero      Edad  Estatura      Nota
## Genero      1.00000000 -0.05471757 0.4763820 0.09408874
## Edad      -0.05471757 1.00000000 0.2085317 0.24505941
## Estatura  0.47638197 0.20853169 1.00000000 0.45485303
## Nota      0.09408874 0.24505941 0.4548530 1.00000000

pairs(datos_corr[,1:4])

```



#15)

```
aprobadas <- subset(datos, Genero=="Mujer" & Nota>=5)
aprobados <- subset(datos, Genero=="Hombre" & Nota>=5)
aprobadas
```

##	Genero	Edad	Estatura	Nota	Calificacion
## 1	Mujer	21	1.82	5	Aprobado
## 3	Mujer	18	1.78	7	Notable
## 4	Mujer	20	1.79	5	Aprobado
## 5	Mujer	23	1.80	9	Sobresaliente
## 7	Mujer	22	1.79	8	Notable
## 11	Mujer	22	1.73	5	Aprobado
## 14	Mujer	19	1.77	6	Aprobado
## 17	Mujer	20	1.66	5	Aprobado
## 18	Mujer	21	NA	6	Aprobado
## 20	Mujer	23	1.80	8	Notable

aprobados

##	Genero	Edad	Estatura	Nota	Calificacion
## 2	Hombre	19	1.83	6	Aprobado
## 6	Hombre	22	1.90	7	Notable
## 9	Hombre	21	NA	9	Sobresaliente
## 12	Hombre	20	1.79	8	Notable
## 13	Hombre	22	1.80	7	Notable

#16)

```
tapply(datos$Nota, datos$Genero, mean)
```

```
##   Hombre   Mujer
```

```
## 6.285714 5.461538
```

#17)

```
d1 <- na.omit(matriz)
```

```
rownames(d1) <- NULL
```

```
d1
```

```
##      Edad Estatura Nota
```

```
## 1      21      1.82    5
```

```
## 2      19      1.83    6
```

```
## 3      18      1.78    7
```

```
## 4      20      1.79    5
```

```
## 5      23      1.80    9
```

```
## 6      22      1.90    7
```

```
## 7      22      1.79    8
```

```
## 8      20      1.83    3
```

```
## 9      21      1.65    2
```

```
## 10     22      1.73    5
```

```
## 11     20      1.79    8
```

```
## 12     22      1.80    7
```

```
## 13     19      1.77    6
```

```
## 14     19      1.69    3
```

```
## 15     21      1.75    4
```

```
## 16     20      1.66    5
```

```
## 17     22      1.79    2
```

```
## 18     23      1.80    8
```

#library(MVN)

```
atipicos <- mvoutlier::uni.plot(d1)
```

```
atipicos #6,9,14,16
```

```
## $outliers
```

```
## [1] FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE  
FALSE
```

```
## [13] FALSE  TRUE FALSE  TRUE FALSE FALSE
```

```
##
```

```
## $md
```

```
## [1] 0.8385508 1.3760481 1.4222888 0.4664501 1.2919839 3.3144618 0.8051042
```

```
## [8] 1.4589187 4.8124716 2.2014582 0.8307237 0.6071732 1.1024962 3.5024069
```

```
## [15] 1.6159648 4.2552103 1.5854735 1.1070432
```

```
atipicos$outliers
```

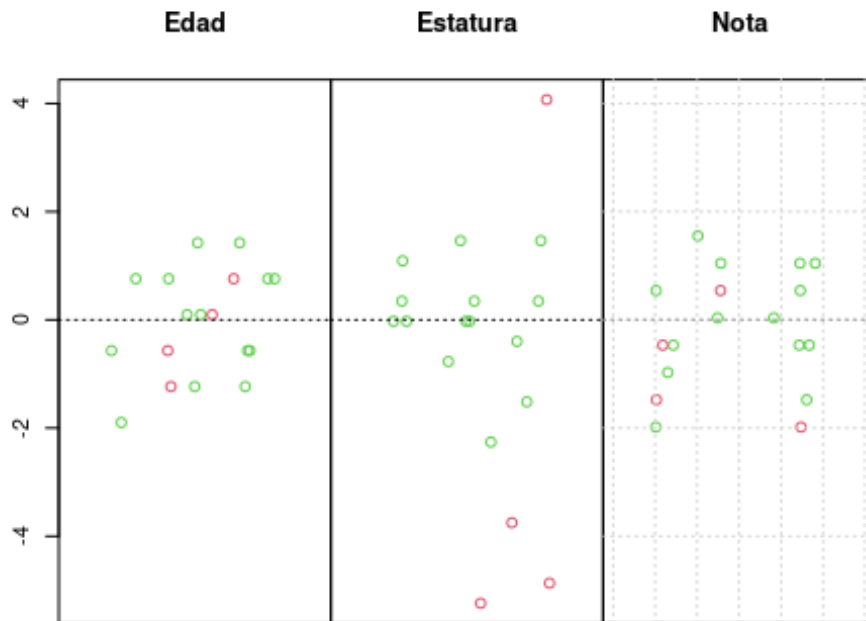
```
## [1] FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE  
FALSE
```

```
## [13] FALSE  TRUE FALSE  TRUE FALSE FALSE
```

```
atipicos$md
```

```
## [1] 0.8385508 1.3760481 1.4222888 0.4664501 1.2919839 3.3144618 0.8051042
## [8] 1.4589187 4.8124716 2.2014582 0.8307237 0.6071732 1.1024962 3.5024069
## [15] 1.6159648 4.2552103 1.5854735 1.1070432
```

```
library(scatterplot3d)
colores <- c("blue", "red")
colores <- colores[as.numeric(as.factor(atipicos$outliers))]
grid()
```



```
scatterplot3d(d1, pch = 16, type="h", color=colores)
```

```
#####
```

```
datos_limpios <- na.omit(datos)
head(datos_limpios)
```

```
##   Genero Edad Estatura Nota  Calificacion
## 1  Mujer   21    1.82    5    Aprobado
## 2 Hombre   19    1.83    6    Aprobado
## 3  Mujer   18    1.78    7    Notable
## 4  Mujer   20    1.79    5    Aprobado
## 5  Mujer   23    1.80    9 Sobresaliente
## 6 Hombre   22    1.90    7    Notable
```

```
rownames(datos_limpios) <- NULL
```

```
matriz <- datos_limpios[,2:4]
```



```

S <- cov(matriz)
c <- colMeans(matriz)
dMH <- mahalanobis(matriz,c,S)
dMH

## [1] 0.9411830 2.7623462 4.9019647 0.4888947 4.1529117 4.2943585 1.7712790
## [8] 4.1049654 5.3632008 1.5704284 2.0412746 0.9191843 1.7207366 3.0273319
## [15] 0.6590339 4.0248387 5.2788671 2.9772004

datos_limpios <- cbind.data.frame(datos_limpios,dMH)
datos_limpios <- datos_limpios[order(datos_limpios$dMH, decreasing = TRUE),]
datos_limpios

##      Genero Edad Estatura Nota  Calificacion      dMH
## 9      Mujer  21      1.65   2      Suspenso 5.3632008
## 17     Mujer  22      1.79   2      Suspenso 5.2788671
## 3      Mujer  18      1.78   7      Notable 4.9019647
## 6     Hombre  22      1.90   7      Notable 4.2943585
## 5      Mujer  23      1.80   9 Sobresaliente 4.1529117
## 8     Hombre  20      1.83   3      Suspenso 4.1049654
## 16     Mujer  20      1.66   5      Aprobado 4.0248387
## 14     Mujer  19      1.69   3      Suspenso 3.0273319
## 18     Mujer  23      1.80   8      Notable 2.9772004
## 2     Hombre  19      1.83   6      Aprobado 2.7623462
## 11     Hombre  20      1.79   8      Notable 2.0412746
## 7      Mujer  22      1.79   8      Notable 1.7712790
## 13     Mujer  19      1.77   6      Aprobado 1.7207366
## 10     Mujer  22      1.73   5      Aprobado 1.5704284
## 1      Mujer  21      1.82   5      Aprobado 0.9411830
## 12     Hombre  22      1.80   7      Notable 0.9191843
## 15     Hombre  21      1.75   4      Suspenso 0.6590339
## 4      Mujer  20      1.79   5      Aprobado 0.4888947

#####

d1 <- na.omit(matriz)
rownames(d1) <- NULL
head(d1)

##      Edad Estatura Nota
## 1      21      1.82   5
## 2      19      1.83   6
## 3      18      1.78   7
## 4      20      1.79   5
## 5      23      1.80   9
## 6      22      1.90   7

library(robustbase)
a1 <- covMcd(d1) #covPlot(d)
at1 <- a1$mcd.wt
at1

```

```
## [1] 1 1 1 1 1 0 1 1 0 1 1 1 1 0 1 0 1 1

library(rrcov)

## Scalable Robust Estimators with High Breakdown Point (version 1.6-0)

a2 <- CovMve(d1)
#6,9,14,16
at2 <- a2$wt
at2

## [1] 1 1 1 1 1 0 1 1 0 1 1 1 1 0 1 0 1 1

#library(DetMCD)
#a3 <- DetMCD(d1)
#plot.DetMCD(a3)

#library(MVN)
#at <- mvoutlier::mvoutlier.CoDa(d)
```

