7. Regresión lineal

Francisco Castorena Salazar, A00827756

2023-08-29

La recta de mejor ajuste

```
library(readr)
M = read_csv("Estatura-peso_HyM.csv")
```

```
## Rows: 440 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Sexo
## dbl (2): Estatura, Peso
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
M = data.frame(M)
head(M)
```

```
##   Estatura  Peso Sexo
## 1     1.61 72.21    H
## 2     1.61 65.71    H
## 3     1.70 75.08    H
## 4     1.65 68.55    H
## 5     1.72 70.77    H
## 6     1.63 77.18    H
```

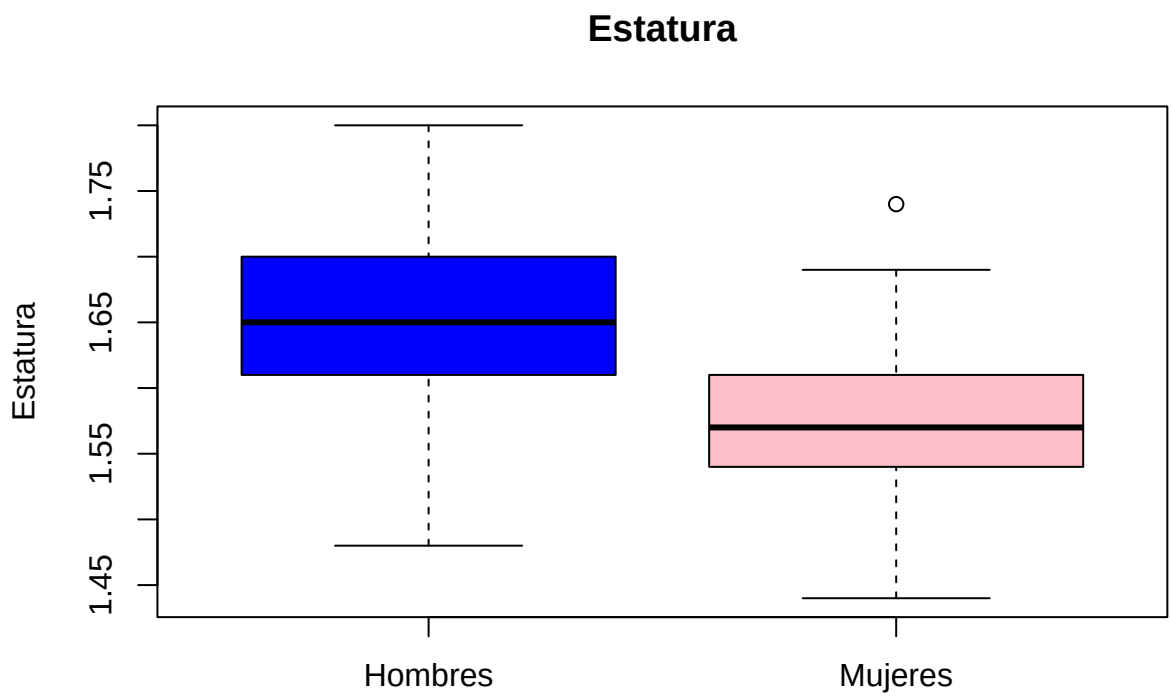
1. Obtén la matriz de correlación de los datos que se te proporcionan. Interpreta.

```
MM = subset(M,M$Sexo=="M")
MH = subset(M,M$Sexo=="H")
M1=data.frame(MH$Estatura,MH$Peso,MM$Estatura,MM$Peso)

n=4 #número de variables
d=matrix(NA,ncol=7,nrow=n)
for(i in 1:n){
  d[i,]<-c(as.numeric(summary(M1[,i])),sd(M1[ ,i]))
```

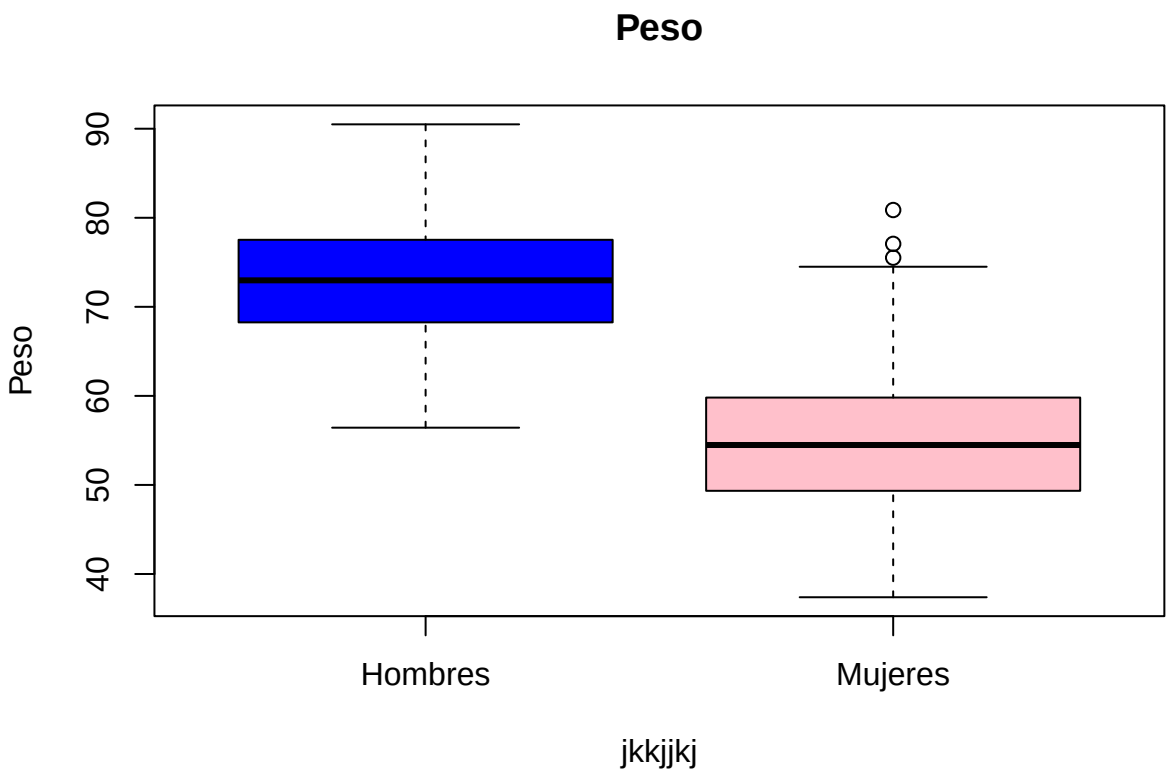
```
}
m=as.data.frame(d)

row.names(m)=c("H-Estatura","H-Peso","M-Estatura","M-Peso")
names(m)=c("Minimo","Q1","Mediana","Media","Q3","Máximo","Desv Est")
m
```

```
##            Minimo      Q1 Mediana     Media      Q3 Máximo   Desv Est
## H-Estatura   1.48  1.6100   1.650  1.653727  1.7000   1.80 0.06173088
## H-Peso      56.43 68.2575  72.975 72.857682 77.5225  90.49 6.90035408
## M-Estatura   1.44  1.5400   1.570  1.572955  1.6100   1.74 0.05036758
## M-Peso      37.39 49.3550  54.485 55.083409 59.7950  80.87 7.79278074
```

```
boxplot(M$Estatura~M$Sexo, ylab="Estatura", xlab="", col=c("blue","pink"), names=c("Hombres", "Mujeres"
```




```
boxplot(M$Peso~M$Sexo, ylab="Peso",xlab="jkkjjkj", names=c("Hombres", "Mujeres"), col=c('blue','pink'),
```



Regresion Lineal

Preparacion de los datos

Verificar como esta definida la variable sexo: Caracter (variable categorica)

El modelo con sexo

```
A = lm(M$Peso~M$Estatura+M$Sexo)
A
```

```
##
## Call:
## lm(formula = M$Peso ~ M$Estatura + M$Sexo)
##
## Coefficients:
## (Intercept)   M$Estatura      M$SexoM
##      -74.75        89.26       -10.56
```

```
b0 = A$coefficients[1]
b1 = A$coefficients[2]
b2 = A$coefficients[3]

cat("Peso=",b0,"+","Estatura",b2,"sexoM")
```

```
## Peso= -74.7546 + Estatura -10.56447 sexoM
```

Verificacion del Modelo

- Significancia global (statistic F)
- Significancia individual (t value)
- Porcentaje de variacion explicada por el modelo (Adjusted R-squared)

```
summary(A)
```

```
##
## Call:
## lm(formula = M$Peso ~ M$Estatura + M$Sexo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.9505  -3.2491   0.0489   3.2880  17.1243
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -74.7546     7.5555  -9.894   <2e-16 ***
## M$Estatura   89.2604     4.5635  19.560   <2e-16 ***
## M$SexoM     -10.5645     0.6317 -16.724   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.381 on 437 degrees of freedom
## Multiple R-squared:  0.7837, Adjusted R-squared:  0.7827
## F-statistic: 791.5 on 2 and 437 DF,  p-value: < 2.2e-16
```

Ecuacion del modelo

```
#Para mujeres (SexoM =0)
cat("Para mujeres","/n")
```

```
## Para mujeres /n
```

```
cat("Peso =", b0+b2,"+",b1,"Estatura")
```

```
## Peso = -85.31907 + 89.26035 Estatura
```

```
#Para hombres (SexoM =0)
cat("Para hombres","/n")
```

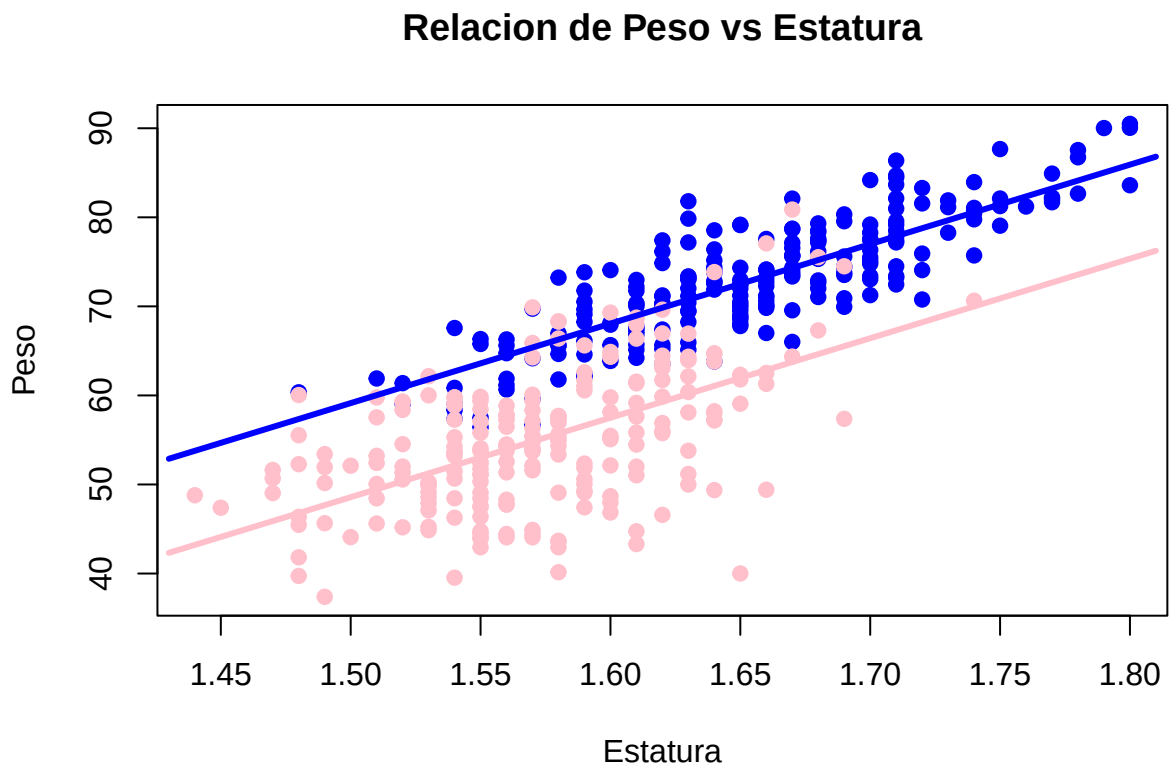
```
## Para hombres /n
```

```r
cat("Peso =", b0,"+",b1,"Estatura")
```

```
## Peso = -74.7546 + 89.26035 Estatura
```

Grafica

```r
Ym = function(x){b0+b2+b1*x}
Yh = function(x){b0+b1*x}

colores = c('blue','pink')
plot(M$Estatura,M$Peso,col=colores[factor(M$Sexo)],pch=19,ylab='Peso',xlab='Estatura',
     main='Relacion de Peso vs Estatura')
x = seq(1.43,1.81,0.01)
lines(x,Ym(x),col='pink',lwd=3)
lines(x,Yh(x),col='blue',lwd=3)
```



Modelo con interaccion

```r
B = lm(M$Peso~M$Estatura*M$Sexo)
B
```

```
##
## Call:
## lm(formula = M$Peso ~ M$Estatura * M$Sexo)
##
```

```
## Coefficients:
##        (Intercept)          M$Estatura             M$SexoM  M$Estatura:M$SexoM
##             -83.68               94.66               11.12              -13.51
```

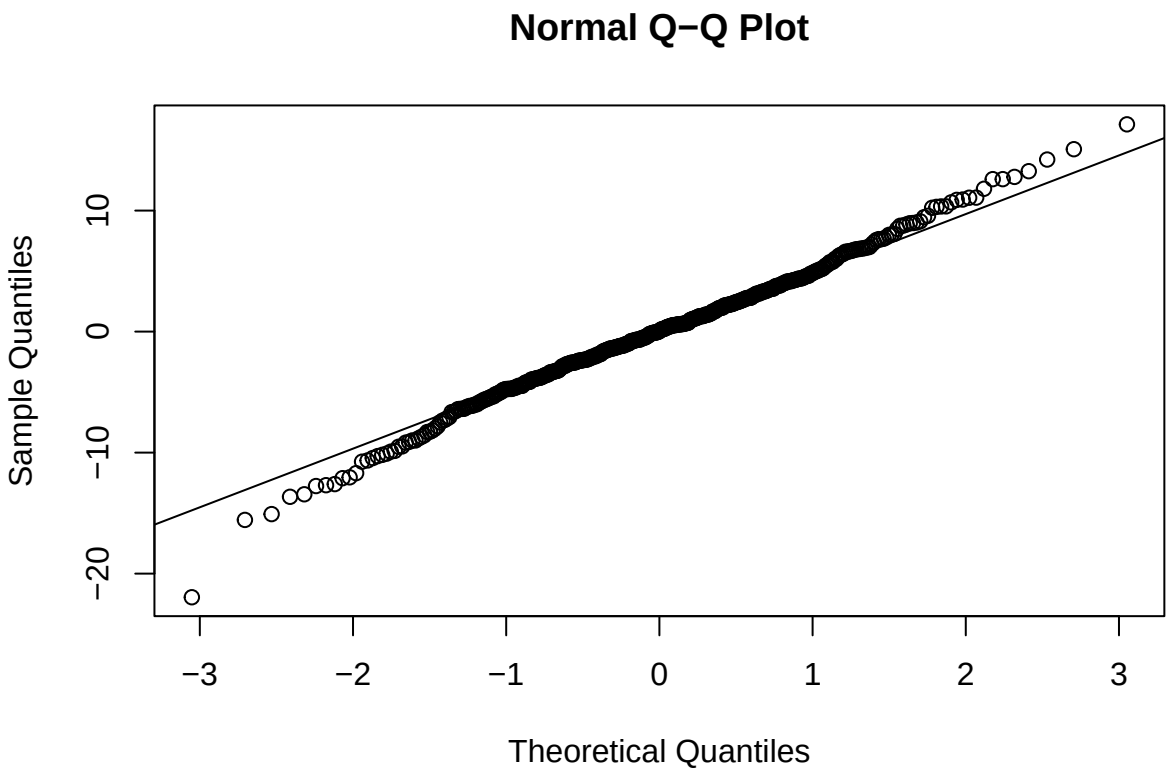
```
summary(B)
```

```
##
## Call:
## lm(formula = M$Peso ~ M$Estatura * M$Sexo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.3256  -3.1107   0.0204   3.2691  17.9114
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -83.685      9.735  -8.597   <2e-16 ***
## M$Estatura            94.660      5.882  16.092   <2e-16 ***
## M$SexoM               11.124     14.950   0.744    0.457
## M$Estatura:M$SexoM  -13.511      9.305  -1.452    0.147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.374 on 436 degrees of freedom
## Multiple R-squared:  0.7847, Adjusted R-squared:  0.7832
## F-statistic: 529.7 on 3 and 436 DF,  p-value: < 2.2e-16
```

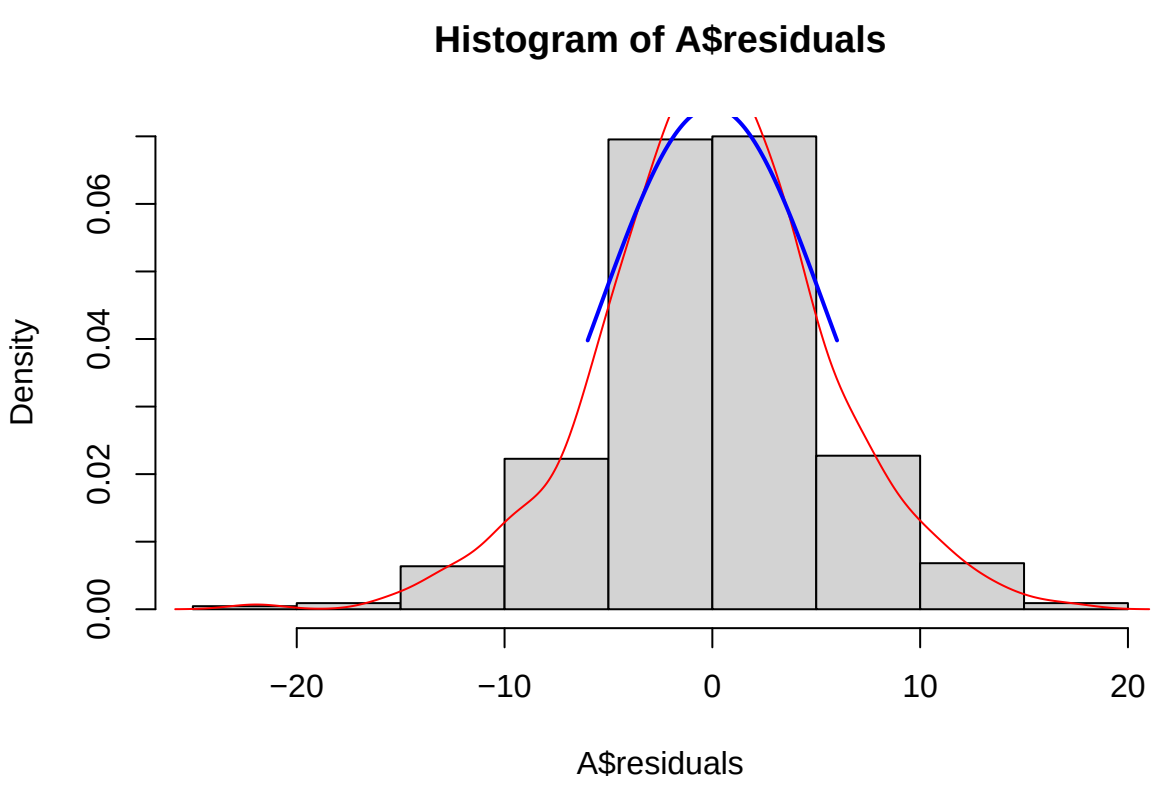
```
library(nortest)
ad.test(A$residuals)
```

```
##
##  Anderson-Darling normality test
##
## data:  A$residuals
## A = 0.79651, p-value = 0.03879
```

```
qqnorm(A$residuals)
qqline(A$residuals)
```




```r
hist(A$residuals,freq=FALSE)
lines(density(A$residual),col='red')
curve(dnorm(x,mean=mean(A$residuals),sd=sd(A$residuals)),from=-6,to=6,add=TRUE,col='blue',lwd=2)
```




```
# plot(M$Peso,A$Ressiduals,pch=19,ylab='Peso',xlab='Estatura',
#     main='Relacion de Peso vs Estatura')
```

```
t.test(A$residuals)
```

```
##
##  One Sample t-test
##
## data:  A$residuals
## t = 6.8754e-16, df = 439, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.5029859  0.5029859
## sample estimates:
##    mean of x
## 1.759562e-16
```

```
plot.new
```

```
## function ()
## {
##     for (fun in getHook("before.plot.new")) {
##         if (is.character(fun))
##             fun <- get(fun)
```

```
##         try(fun())
##     }
##     .External2(C_plot_new)
##     grDevices:::recordPalette()
##     for (fun in getHook("plot.new")) {
##         if (is.character(fun))
##             fun <- get(fun)
##         try(fun())
##     }
##     invisible()
## }
## <bytecode: 0x000001f239f77918>
## <environment: namespace:graphics>
```

```
plot(A$fitted.values,A$residuals)
abline(h=0)
```