

# Regresion lineal

Diego Alberto Baños Lopez  
A01275100

29-08-2023

```
M <- read.csv("Estatura-peso_HyM.csv")
```

```
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")
```

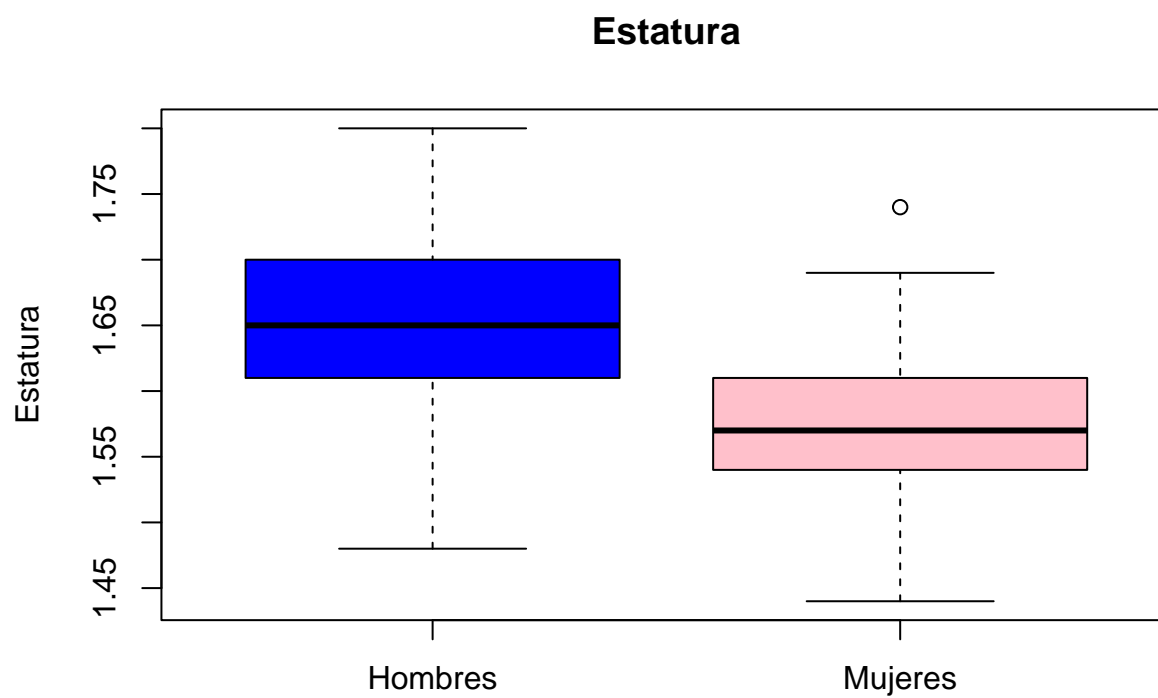
```
cor(M1)
```

```
##              MH.Estatura    MH.Peso  MM.Estatura    MM.Peso
## MH.Estatura 1.0000000000 0.846834792 0.0005540612 0.04724872
## MH.Peso      0.8468347920 1.000000000 0.0035132246 0.02154907
## MM.Estatura 0.0005540612 0.003513225 1.0000000000 0.52449621
## MM.Peso      0.0472487231 0.021549075 0.5244962115 1.00000000
```

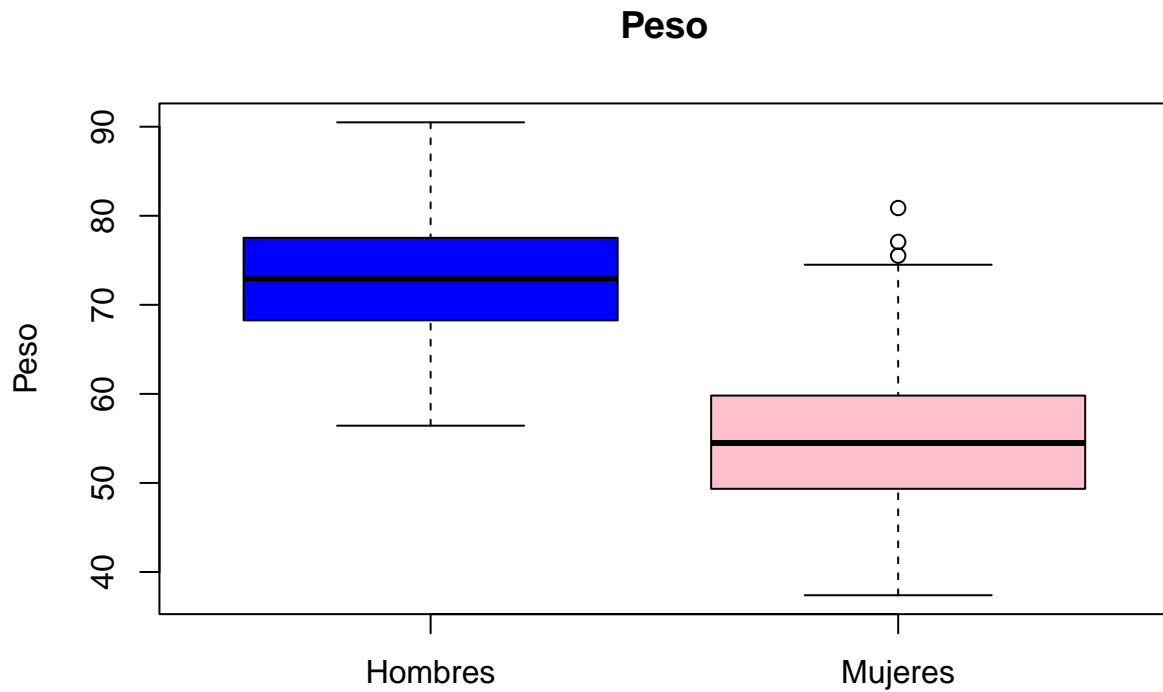
```
cor(M$Estatura, M$Peso)
```

```
## [1] 0.8032449
```

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



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



```
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, "+", b1, "Estatura", b2, "SexoM")
```

```
## Peso = -74.7546 + 89.26035 Estatura -10.56447 SexoM
```

## Verificación del modelo

- significancia global (stadistic f)

- significancia individual (t-value)
- porcentaje de variación explicada por el modelo

```
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

```
cat("Para mujeres: \nPeso =", b0 - b2, "+", b1, "Estatura\n")
```

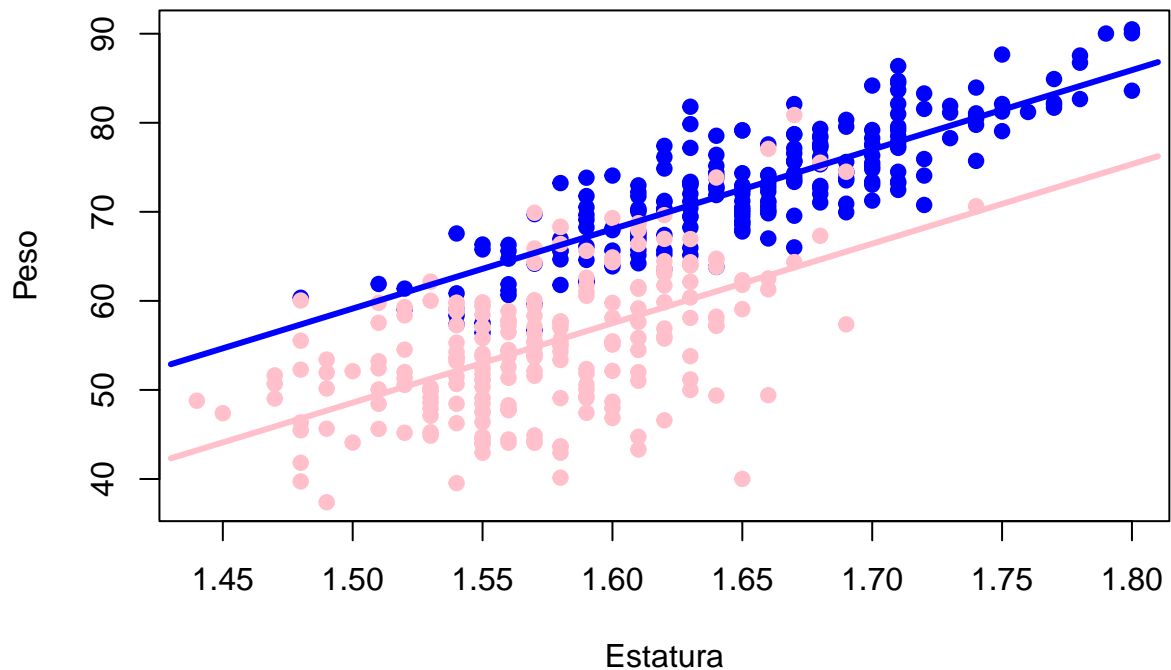
```
## Para mujeres:
## Peso = -64.19013 + 89.26035 Estatura
```

```
cat("Para hombres: \nPeso =", b0, "+", b1, "Estatura\n")
```

```
## Para hombres:
## Peso = -74.7546 + 89.26035 Estatura
```

```
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)
```

## Relacion de peso vs estatura



```
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)
```

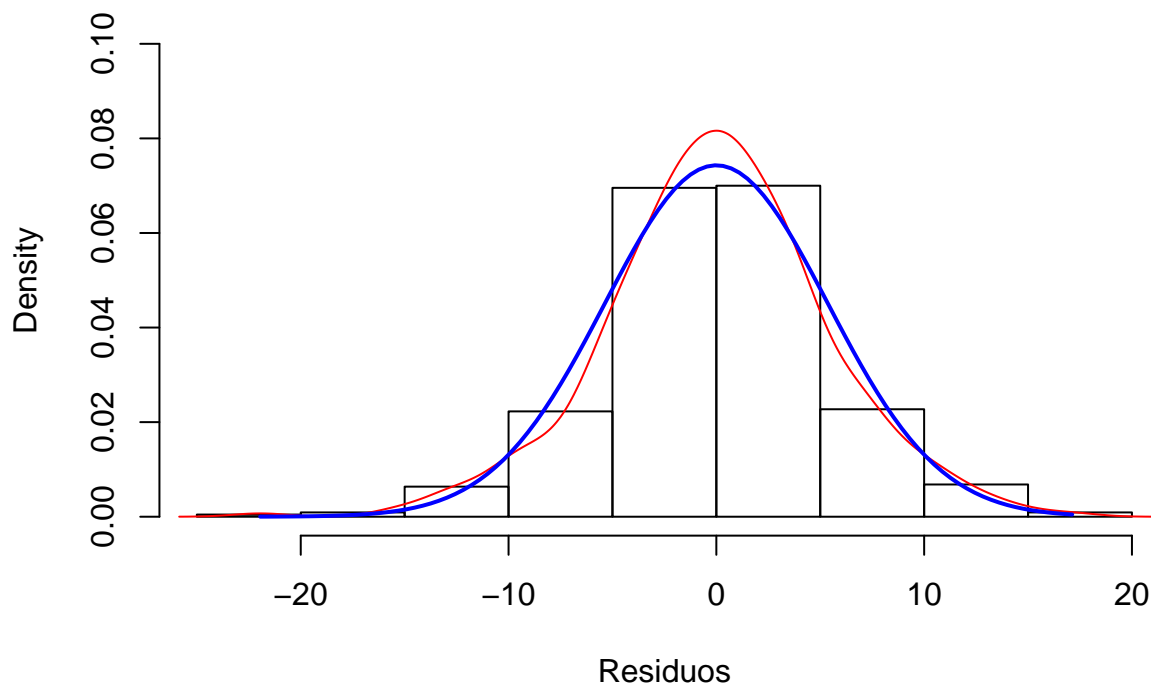
```
##
## Call:
## lm(formula = M$Peso ~ M$Estatura * M$Sexo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.3256  -3.1107   0.0204   3.2691  17.9114
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -83.685     9.735  -8.597  <2e-16 ***
## M$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
```

## Normalidad

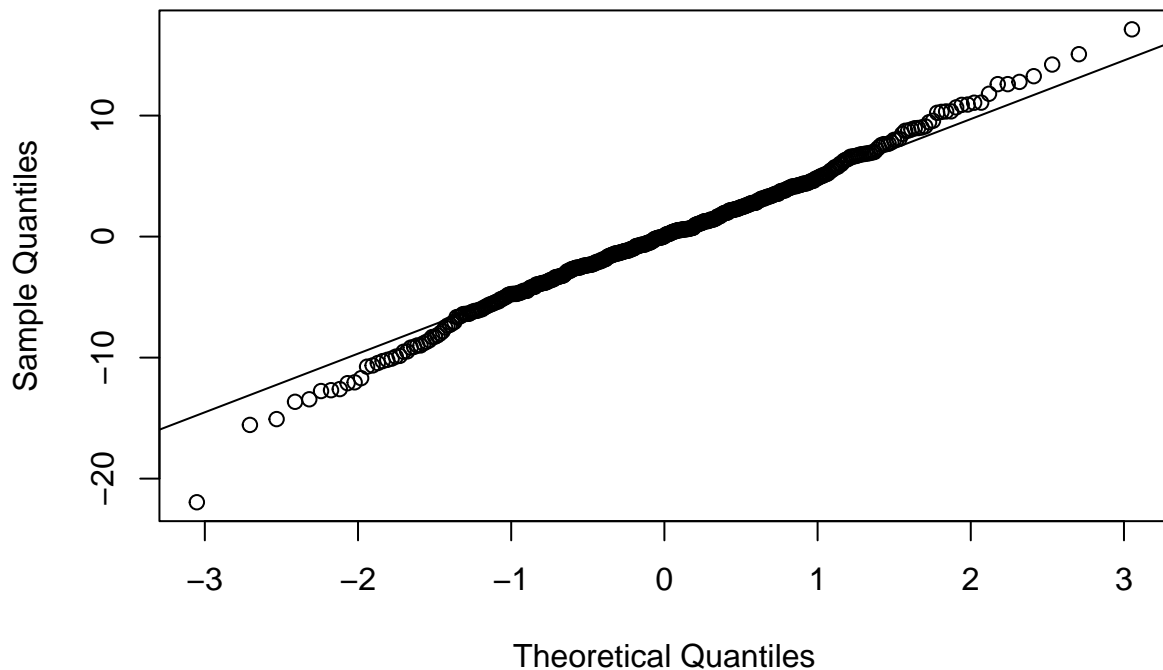
```
hist(A$residuals,
     freq = FALSE, ylim = c(0, 0.1),
     xlab = "Residuos", col = 0, main = "Histograma de residuos"
)
lines(density(A$residuals), col = "red")
curve(dnorm(x, mean = mean(A$residuals), sd = sd(A$residuals)),
     from = min(A$residuals), to = max(A$residuals), add = TRUE,
     col = "blue", lwd = 2)
```

### Histograma de residuos



```
qqnorm(A$residuals, main = "QQ-Plot de los residuos")
qqline(A$residuals)
```

## QQ-Plot de los residuos



```
library(nortest)

ad_test <- ad.test(A$residuals)
ad_test

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

En base a los resultados podemos concluir lo siguiente:

- La correlación entre la estatura y el peso de los hombres es de 0.8468, lo que indica una correlación positiva
- La correlación entre la estatura y el peso de las mujeres es de 0.5245, lo que indica una correlación positiva
- El R-cuadrado es 0.7837, lo que significa que el modelo explica el 78.37% de la variabilidad en el peso.
- El F-statistic es 791.5 con un p-valor muy pequeño, lo que indica que el modelo es significativo.

Esto puede interpretarse en otras palabras que hay una relacion noteoria y lean entre peso y estatura para ambos sexos, asi como que el modelo de regresión es significativo y puede explicar como se comportan los valores de peso.