

a01749489

Jorge Javier Sosa Briseño

August 29, 2023

Regresión Lineal

```
1* ---
2  title: "Regresion Lineal"
3  author: "Jorge Javier A01749489"
4  date: "8/29/2023"
5  output: html_document
6* ---
7* ````{r}
8# Cargar los datos
9M <- read.csv("Estatura-peso_HyM.csv")
10M
11* ````
```

Description: df [440 x 3]

Estatura	Peso	Sexo
1.61	72.21	H
1.61	65.71	H
1.70	75.08	H
1.65	68.55	H
1.72	70.77	H
1.63	77.18	H
1.76	81.21	H
1.67	75.71	H
1.67	76.57	H
1.65	68.78	H

1-10 of 440 rows

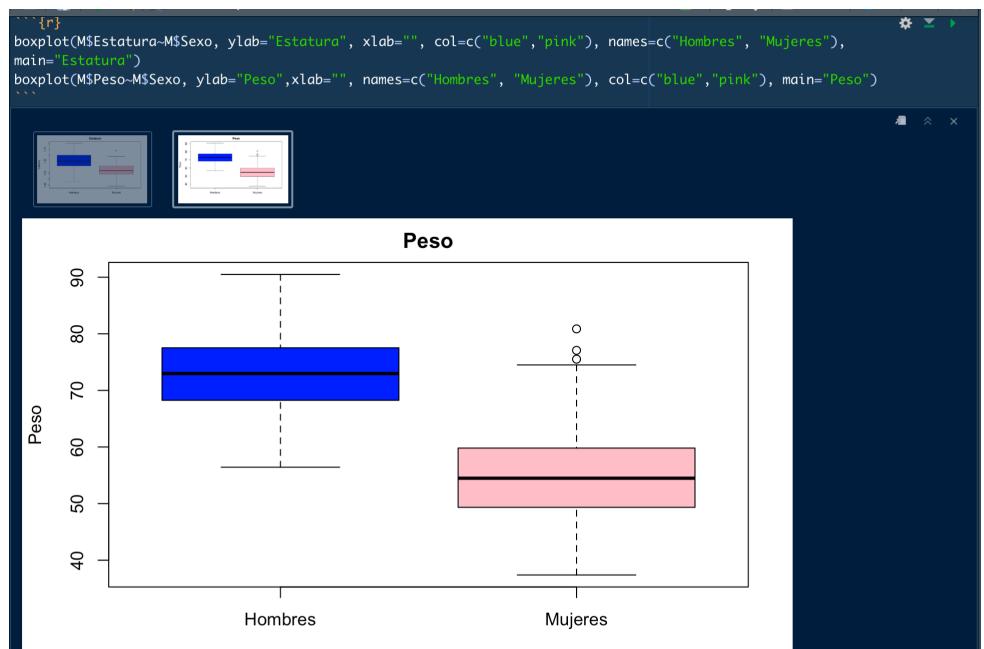
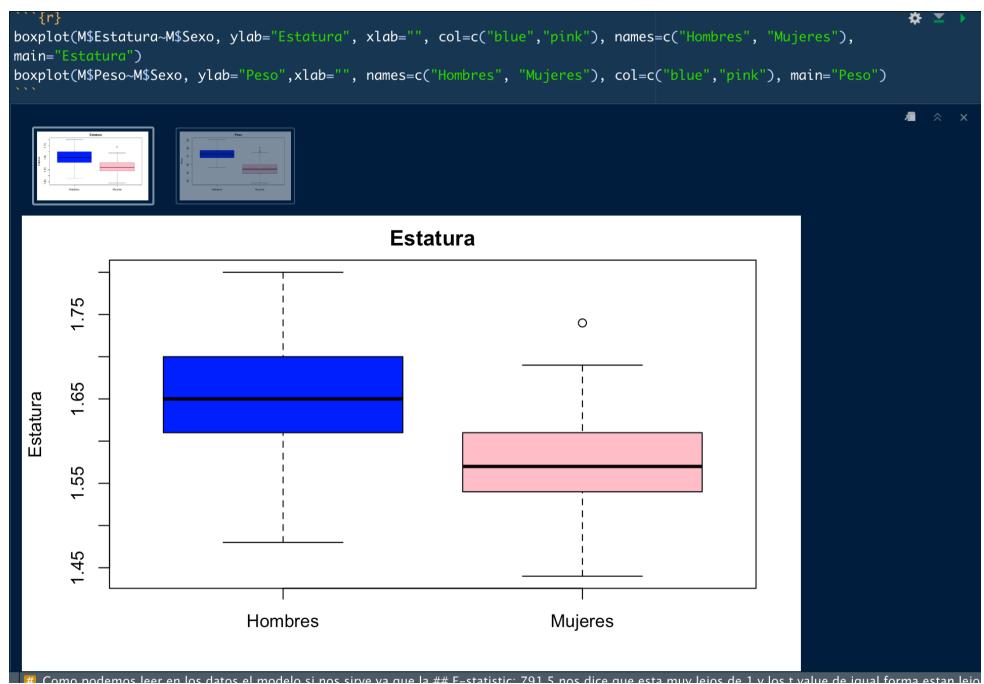
Previous 1 2 3 4 5 6 ... 44 Next

```
12* ##Medidas
13* ````{r}
14 MM = subset(M,MSSexo=="M")
15 MH = subset(M,MSSexo=="H")
16
17 M1=data.frame(MH$Estatura,MH$Peso,MM$Estatura,MM$Peso)
18
19 n=4 #número de variables
20 d=matrix(NA,ncol=7 ,nrow=n)
21 for(i in 1:n){
22 d[i,]=c(as.numeric(summary(M1[,i])),sd(M1[ ,i]))
23 }
24 m=as.data.frame(d)
25
26 row.names(m)=c("H-Estatura","H-Peso","M-Estatura","M-Peso")
27 names(m)=c("Minimo","Q1","Mediana","Media","Q3","Máximo","Desv Est")
28 m
29* ````
```

Description: df [4 x 7]

	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

4 rows



```

35 ## Regresion Lineal
36
37 ## Preparación de los datos
38
39 Verificar cómo está definida la variable Sexo: (variable categórica)
40
41 ## El modelo con Sexo
42 ````{r}
43 A = lm(M$Peso~M$Estatura+M$Sexo)
44 A
45 ````
```

```

Call:
lm(formula = M$Peso ~ M$Estatura + M$Sexo)

Coefficients:
(Intercept)  M$Estatura      M$SexoM
-74.75       89.26        -10.56
```

```

46 ````{r}
47 b0 = A$coefficients[1]
48 b1 = A$coefficients[2]
49 b2 = A$coefficients[3]
50
51 cat ("PesoM=",b0,"+",b1,"Estatura",b2,"SexoM")
52 cat ("PesoH=",b0,"+",b1,"Estatura")
53 ````
```

```
PesoM= -74.7546 + 89.26035 Estatura -10.56447 SexoMPesoH= -74.7546 + 89.26035 Estatura
```

```

46 ````{r}
47 b0 = A$coefficients[1]
48 b1 = A$coefficients[2]
49 b2 = A$coefficients[3]
50
51 cat ("PesoM=",b0,"+",b1,"Estatura",b2,"SexoM")
52 cat ("PesoH=",b0,"+",b1,"Estatura")
53 ````
```

```
PesoM= -74.7546 + 89.26035 Estatura -10.56447 SexoMPesoH= -74.7546 + 89.26035 Estatura
```

```

54 ## verificación del modelo
55 *Significancia Global
56 * Significancia Individual
57 *Porcentaje de variación explicada por el modelo
58 ````{r}
59 summary(A)
60 ````
```

```

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
```

```
````{r}
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
```

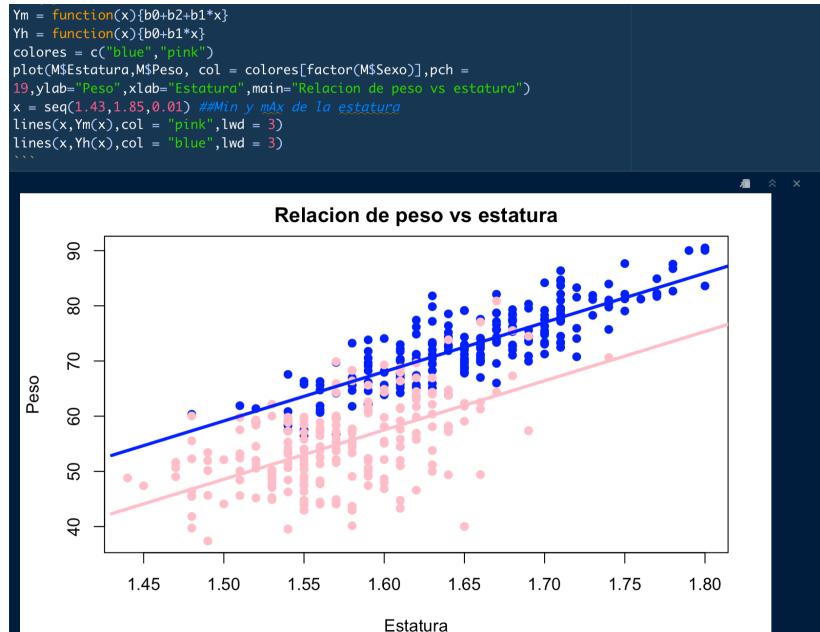
```

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
```

```

Como podemos leer en los datos el modelo si nos sirve ya que la ## F-statistic: 791.5 nos dice que esta muy lejos de 1 y los t value de igual forma estan lejos del valor hipotético. Por ultimo el 0.7827 nos habla de un buen ajuste en el número de las variables y el tamaño de la muestra.
```

```
Ecuacion del modelo
```{r}
#Para mujeres SexoM = 1
cat("Para mujeres","\n")
cat("Peso = ",b0+b2,"+",b1,"Estatura","\n")
#Para Hombres SexoM = 0
cat("Para Hombres","\n")
cat("Peso =",b0,"+",b1,"Estatura")
```
Para mujeres
Peso = -85.31907 + 89.26035 Estatura
Para Hombres
Peso = -74.7546 + 89.26035 Estatura
```



```
##Modelo con interacción
```{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

Significancia
```{r}
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
```

```
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
```

se concluye que no hay efecto de interacción  
Si dejé sin interacción el sexo, se hace significativo  
No vamos a quitar de soteton varias variables  
Si quitamos interacción, obtenemos el modelo de arriba  
Quitar una a una, no varias  
Nos quedamos con el modelo A  
No hay interacción entre sexo y estatura.

Betas significativas:

```
Valides del modelo.
```

Los residuos explican la validez del modelo, son todas las variables que no conozco y mando al error.  
Los errores me reflejan el residuo.

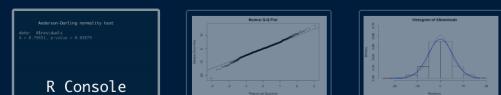
La linea recta están formada por las medias.

Los errores idealmente deben de estar muy pegados a la recta, su varianza no debe de cambiar

No queremos rechazar  $H_0$ , queremos que alfa sea grande

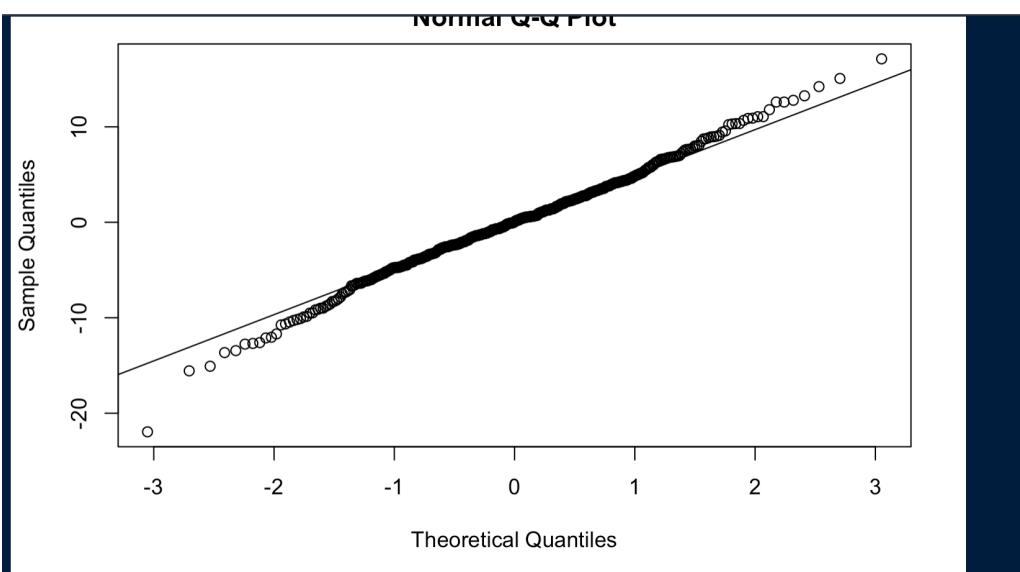
No queremos rechazar  $H_0$ , queremos que alfa sea grande

```
##Normalidad
Histograma
Prueba de Hipótesis
QQplot
```{r}  
ad.test(A$residuals)  
qqnorm(A$residuals)  
qqline(A$residuals)  
hist(A$residuals,freq=FALSE,ylim = c(0,0.1),xlab = "Residuos",col = 0)  
lines(density(A$residuals),col="red") ##LA CURVA AJUSTADA  
curve(dnorm(x,mean=mean(A$residuals),sd=sd(A$residuals)), from=min(A$residuals), to=max(A$residuals),  
add=TRUE, col="blue",lwd=2) ##Teórica de la media de la media de los residuos, la ideal  
```
```



Anderson-Darling normality test

```
data: A$residuals
A = 0.79651, p-value = 0.03879
```



Al tener un p-valor de 0.03879 no rechazamos  $H_0$  ya que resulta que si es normal dado que tenemos un alfa de 0.03.

Las colas de la distribución están fallando un poco o sea que hay una curtosis alta.