

Act7_FridaCanoFalcon_A01752953

Frida Cano Falcon

2023-08-29

```
library(nortest)
M = read.csv("Estatura-peso_HyM.csv")
```

Medidas

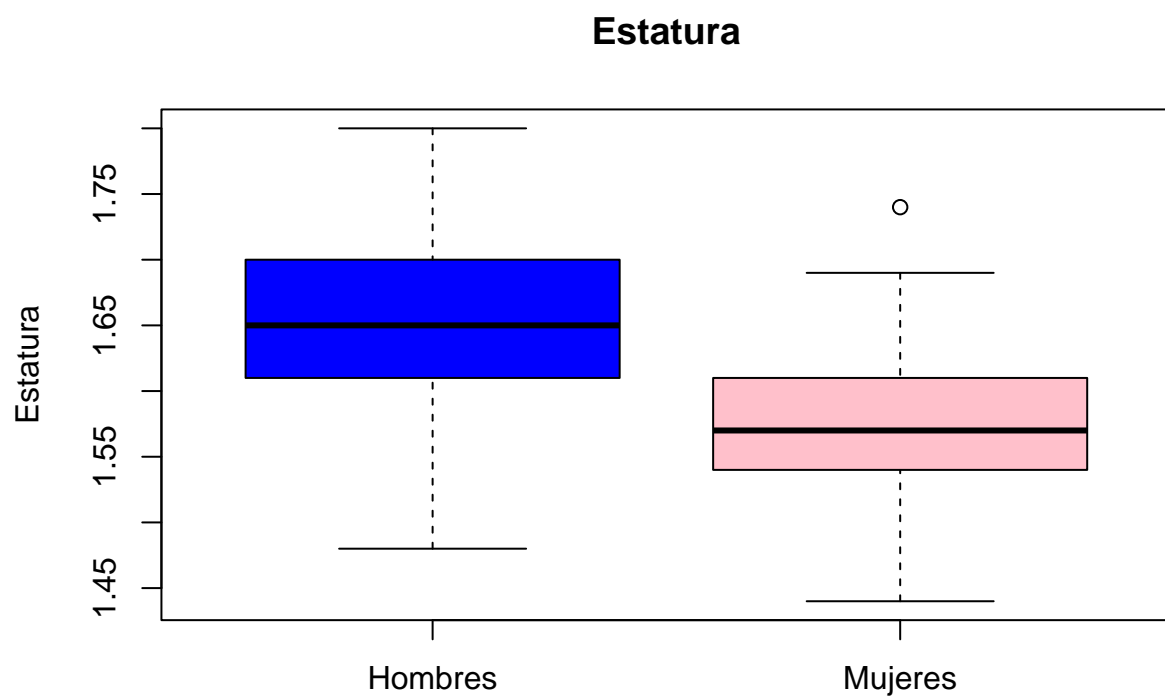
```
MM = subset(M,M$Sexo=="M")
MH = subset(M,M$Sexo=="H")
M1 = data.frame(MH$Estatura,MH$Peso,MM$Estatura,MM$Peso)

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("Mínimo","Q1","Mediana","Media","Q3","Máximo","Desv Est")
m
```

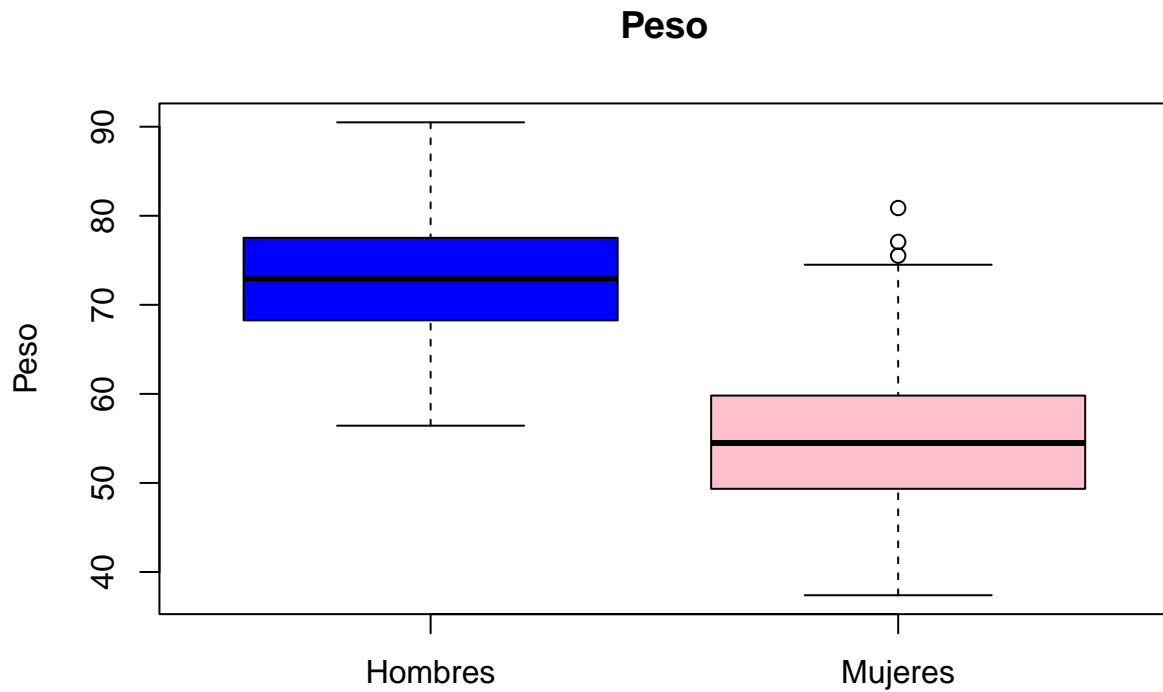
##		Minimo	Q1	Mediana	Media	Q3	Máximo	Desv Est
##	H-Estatura	1.48	1.6100	1.650	1.653727	1.7000	1.80	0.06173088
##	H-Peso	56.43	68.2575	72.975	72.857682	77.5225	90.49	6.90035408
##	M-Estatura	1.44	1.5400	1.570	1.572955	1.6100	1.74	0.05036758
##	M-Peso	37.39	49.3550	54.485	55.083409	59.7950	80.87	7.79278074

Describir las variables

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



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



Regresión Lineal

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

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

Verificación del modelo

- Significancia global
- Significancia individual
- 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
```

Ecuación del modelo

```
# Para mujeres (SexoM=1)
cat("Para mujeres","\n")
```

```
## Para mujeres
```

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

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

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

```
##
```

```
## Para hombres
```

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

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

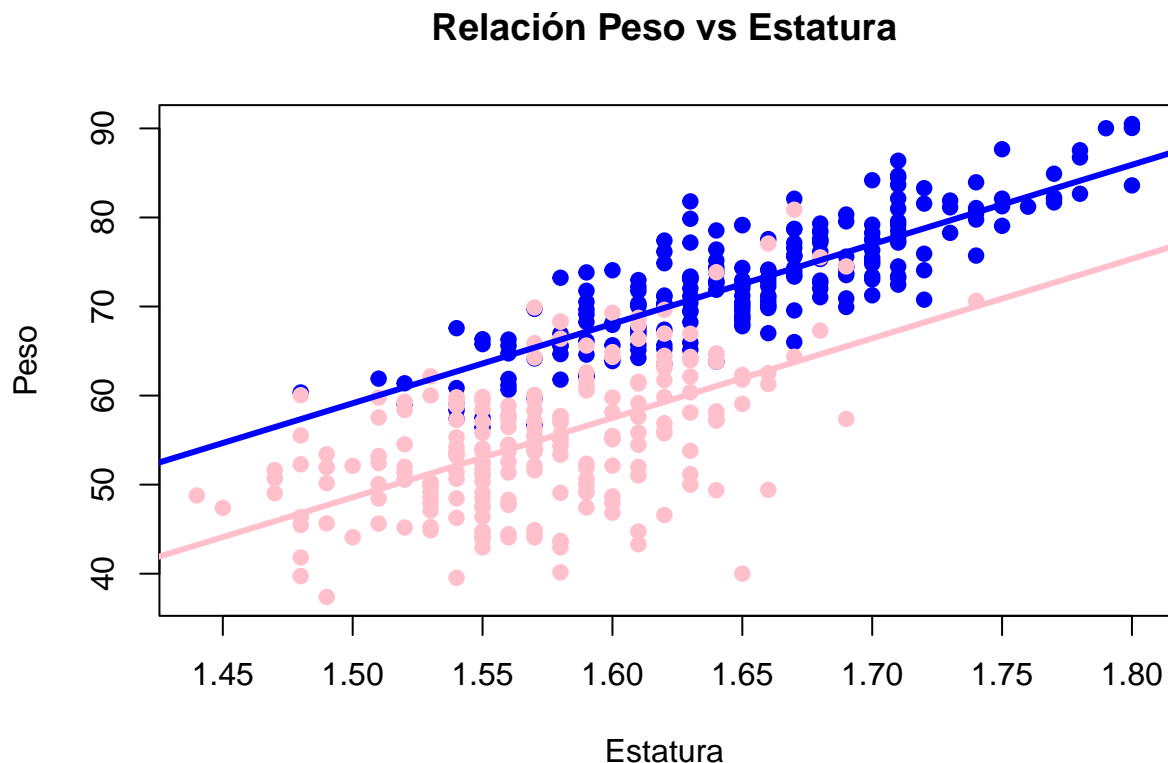
Grafica

```
x = seq(1.4, 2.0, by=0.2)

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="Relación Peso vs Estatura")

lines(x, Yh(x), col="blue", lwd=3)
lines(x, Ym(x), col="pink", lwd=3)
```



Interpretación Vemos que para cada sexo existe una relación entre la estatura y peso, es decir, entre más alto sea la persona más peso tiene. En el contexto de tu modelo de regresión lineal donde estás prediciendo el peso a partir de la estatura y el género (codificado como una variable binaria, por ejemplo, 0 para mujeres y 1 para hombres), el coeficiente B_0 se refiere al valor de la variable dependiente (peso) cuando la estatura es cero y el género es “mujer”. En el caso del género, al codificarlo como 0 para mujeres y 1 para hombres, la interpretación es que el coeficiente B_0 representa la diferencia en el peso promedio entre mujeres y hombres cuando la estatura es cero, lo cual no tiene una interpretación práctica o realista.

El coeficiente B_1 se refiere al cambio promedio en el peso cuando la estatura aumenta en una unidad, considerando ambos géneros. Si el coeficiente B_1 es positivo y significativo, significa que, en promedio, a medida que la estatura aumenta, se espera que el peso también aumente, manteniendo constante el género. Si el coeficiente B_1 es negativo y significativo, indica que a medida que la estatura aumenta, se espera que el peso disminuya, nuevamente manteniendo constante el género.

Modelo de interacción

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

Significancia

```
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
```

NO HAY MODELO DE INTERACCIÓN

Porque nos dimos cuenta que el sexo si es significativo

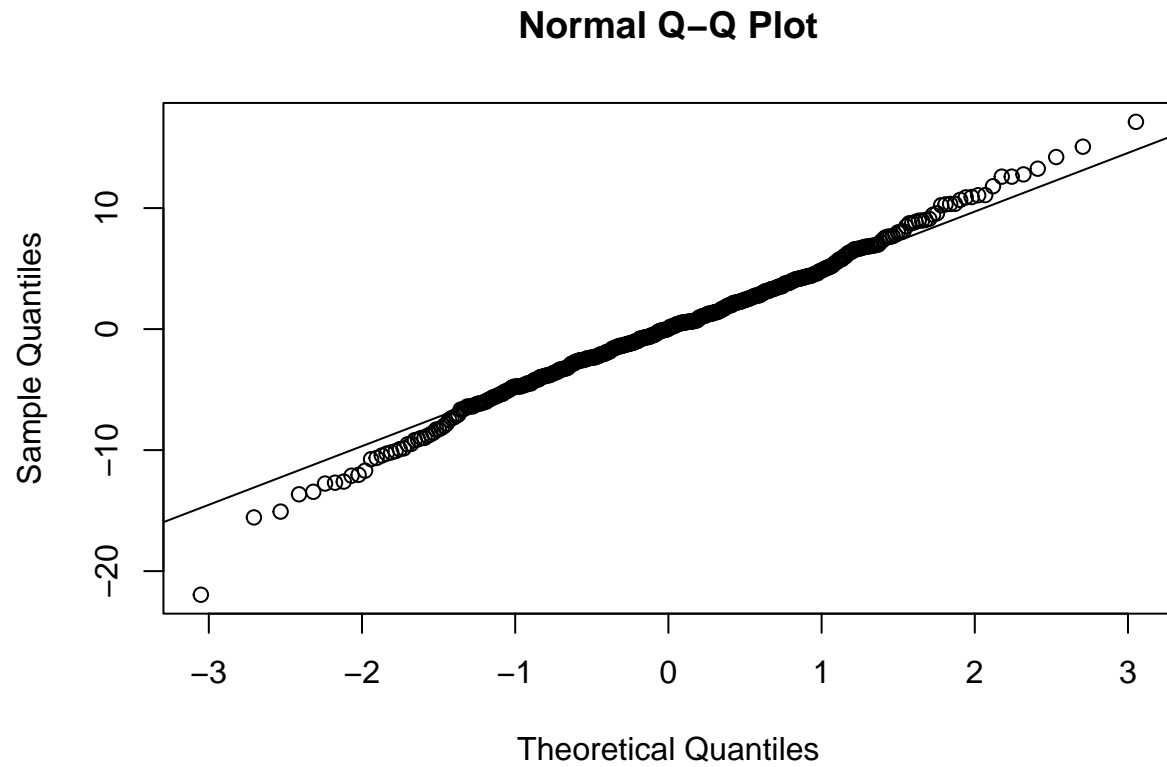
Validez del modelo

Análisis de residuos, normalidad de los residuos

```
library(nortest)
ad.test(A$residuals) # Porque tiene más de 50 valores
```

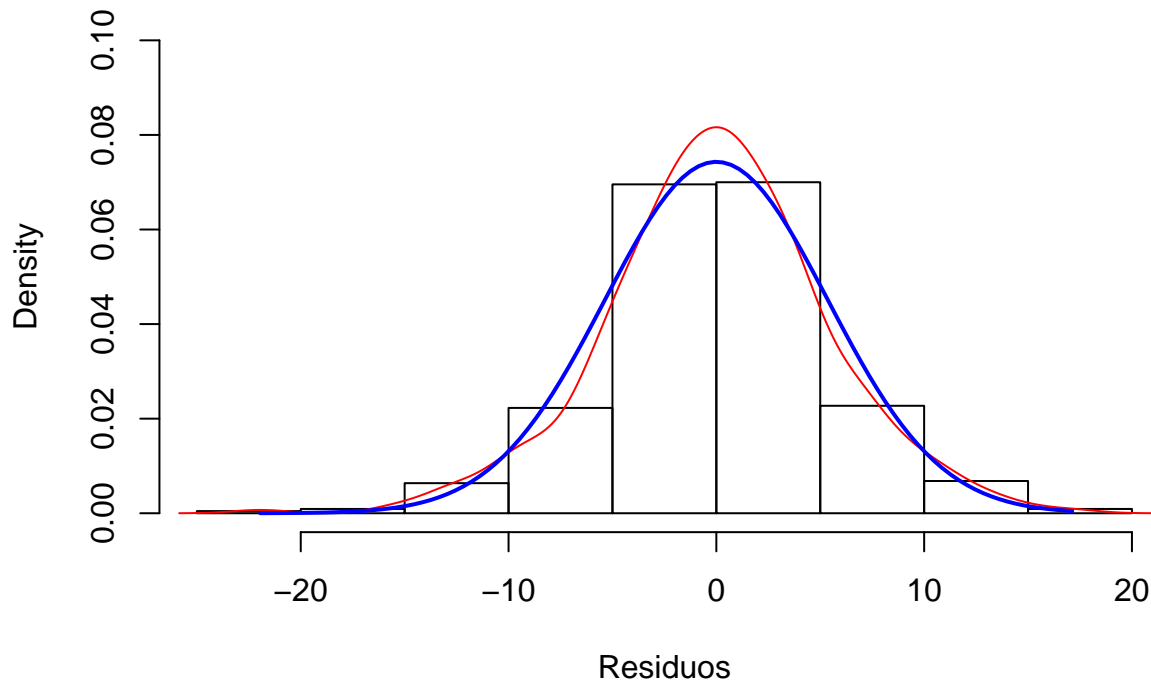
```
##
## Anderson-Darling normality test
##
## data:  A$residuals
## A = 0.79651, p-value = 0.03879
```

```
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")
curve(dnorm(x,mean=mean(A$residuals),sd=sd(A$residuals)), from=min(A$residuals), to=max(A$residuals), a
```

Histogram of A\$residuals



Normalidad de los residuos En este caso se rechaza la hipótesis de normalidad debido a que $p < \alpha$, pero solo por muy poco.

Media de los errores

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

```
##
## One Sample t-test
##
## data: A$residuals
## t = 2.4085e-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
## 6.163788e-17
```

Homocedasticidad e independencia

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