

## Actividad 12

José Carlos Sánchez Gómez

2024-09-04

```
M =  
read.csv("C:\\Users\\jcsg6\\Documentos\\Uni\\SeptimoSemestre\\Estadistica  
\\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)  
  
Modelo2 = lm(Peso ~ Estatura + Sexo, M)  
Modelo2  
  
##  
## Call:  
## lm(formula = Peso ~ Estatura + Sexo, data = M)  
##  
## Coefficients:  
## (Intercept)      Estatura      SexoM  
##      -74.75      89.26      -10.56  
  
summary(Modelo2)  
  
##  
## Call:  
## lm(formula = Peso ~ Estatura + Sexo, data = M)  
##  
## 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 ***  
## Estatura    89.2604     4.5635  19.560  <2e-16 ***  
## 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  
  
cat("Variacion explicada (r^2): ", summary(Modelo2)$adj.r.squared)  
  
## Variacion explicada (r^2):  0.7826698
```

```

b0 = Modelo2$coefficients[1]
b1 = Modelo2$coefficients[2]
b2 = Modelo2$coefficients[3]

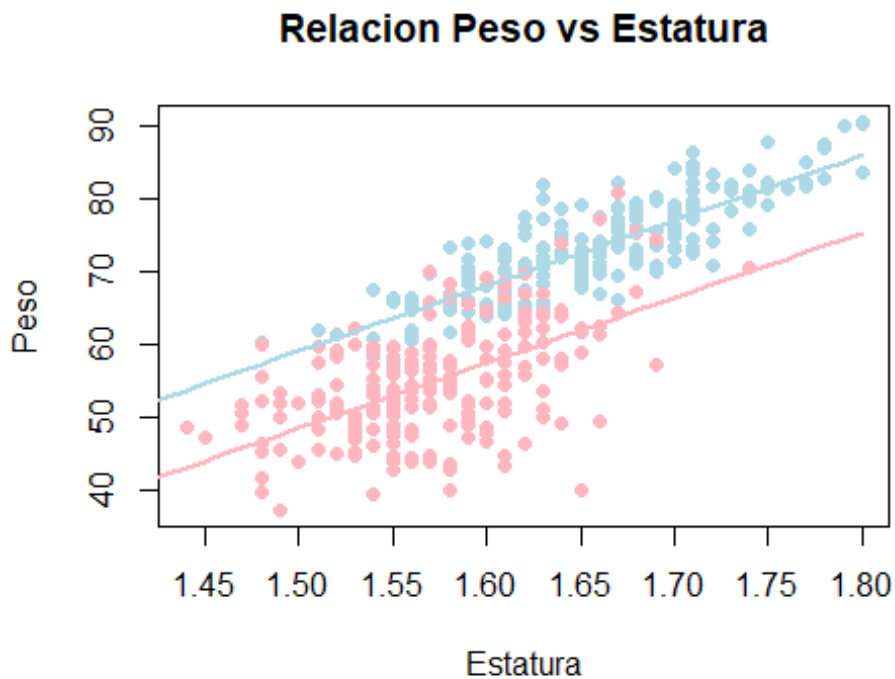
YM = function(x){b0 + b2 + b1 * x}
YH = function(x){b0 + b1 * x}

colores = c("lightblue", "lightpink")
plot(M$Estatura, M$Peso, col = colores[factor(M$Sexo)], pch = 19, ylab =
"Peso", xlab = "Estatura", main = "Relacion Peso vs Estatura")

x = seq(1.40, 1.80, 0.01)
predH = YH(x)
predM = YM(x)

lines(x, predM, col = "lightpink", lwd = 2)
lines(x, predH, col = "lightblue", lwd = 2)

```



## Residuos del modelo

### Normalidad de residuos

```

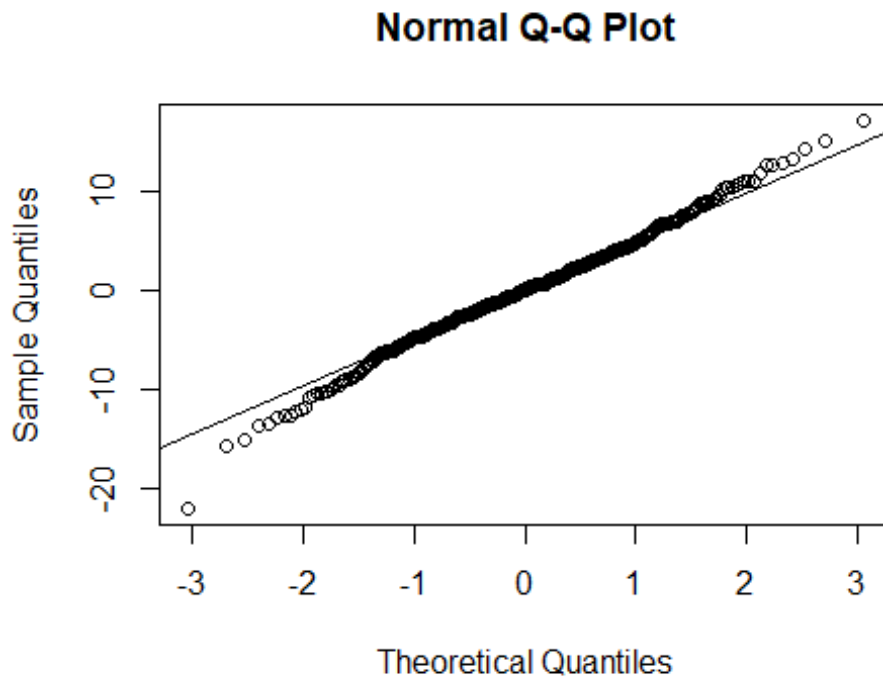
library(nortest)
ad.test(Modelo2$residuals)

##
## Anderson-Darling normality test

```

```
##
## data: Modelo2$residuals
## A = 0.79651, p-value = 0.03879

qqnorm(Modelo2$residuals)
qqline(Modelo2$residuals)
```



En este caso no se rechaza la  $h_0$  porque nuestra  $\alpha = 0.03$  y  $p = 0.038$ , y para poder rechazar esta hipótesis nuestro valor  $p$  tiene que ser menor al alfa.

#### Verificación de media cero

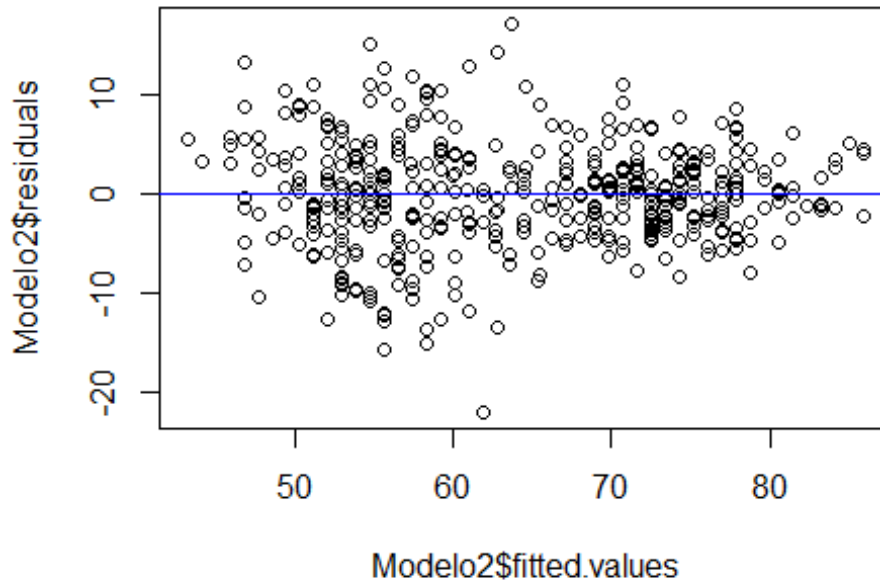
```
t.test(Modelo2$residuals)

##
## One Sample t-test
##
## data: Modelo2$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
```

Se rechaza  $h_0$  porque el valor  $p$  de la media cero no es igual a 0, por lo que se acepta  $h_1$

### Homocedasticidad e Independencia

```
plot(Modelo2$fitted.values, Modelo2$residuals)
abline(h=0, col="blue")
```



### Homocedasticidad

```
library(lmtest)

## Cargando paquete requerido: zoo

##
## Adjuntando el paquete: 'zoo'

## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric

bptest(Modelo2)

##
## studentized Breusch-Pagan test
##
## data:  Modelo2
## BP = 48.202, df = 2, p-value = 3.413e-11
```

Se acepta  $h_0$  debido a que el valor p que nos da la prueba de BP es mucho menor a 0.05, por lo que no hay suficiente evidencia de que existe heterocedasticidad.

### *Independencia*

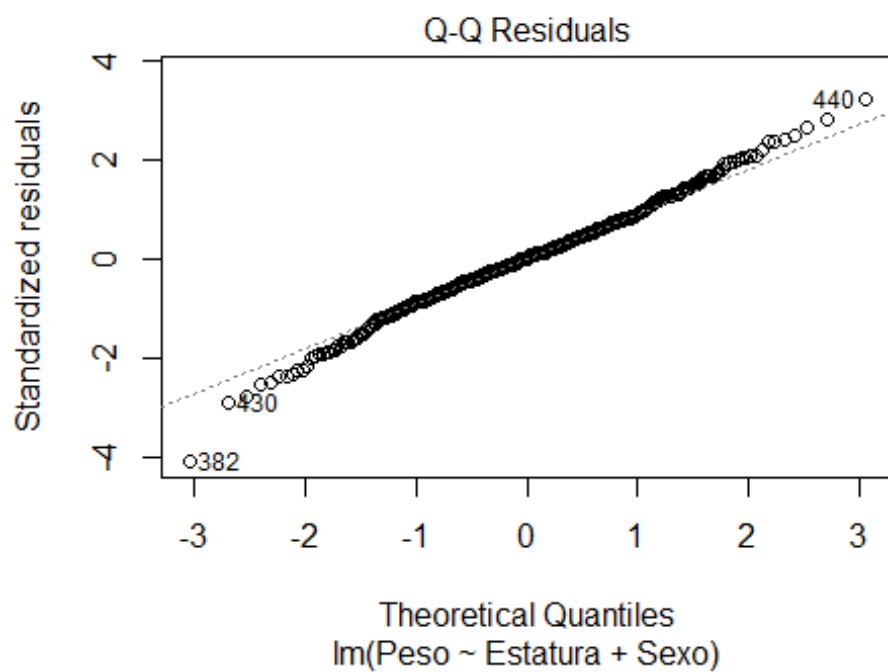
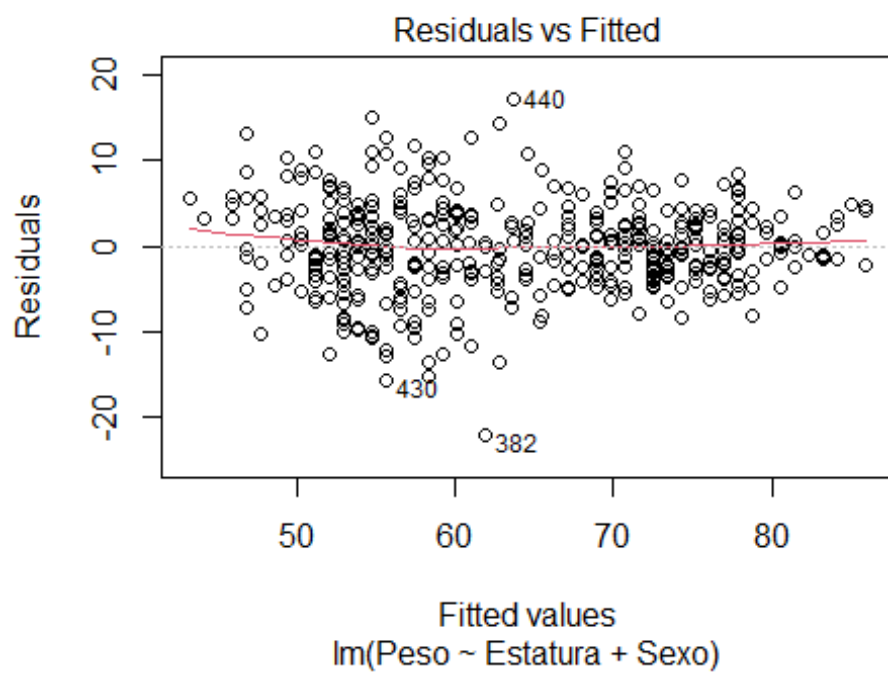
```
library(lmtest)
dwtest(Modelo2)
```

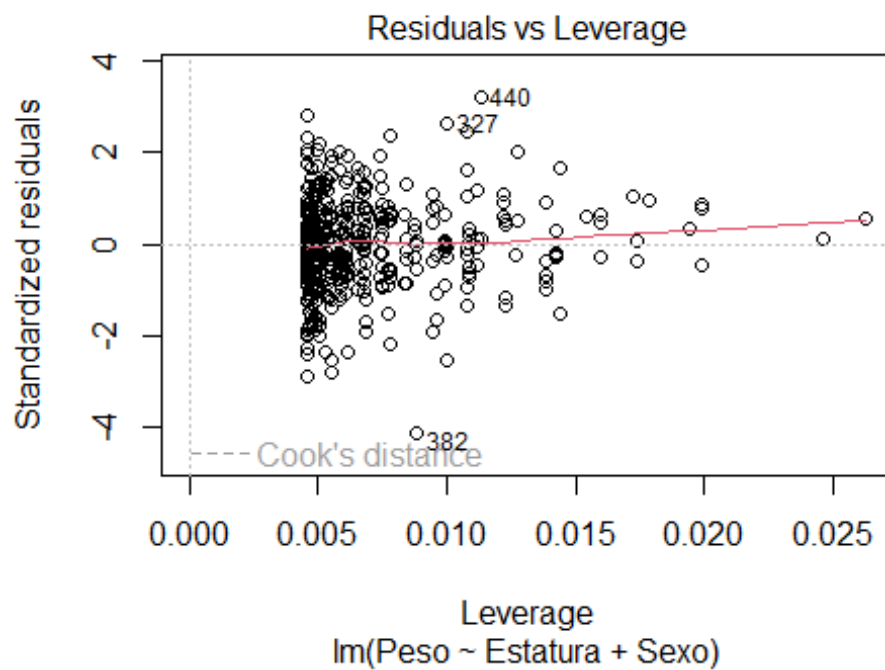
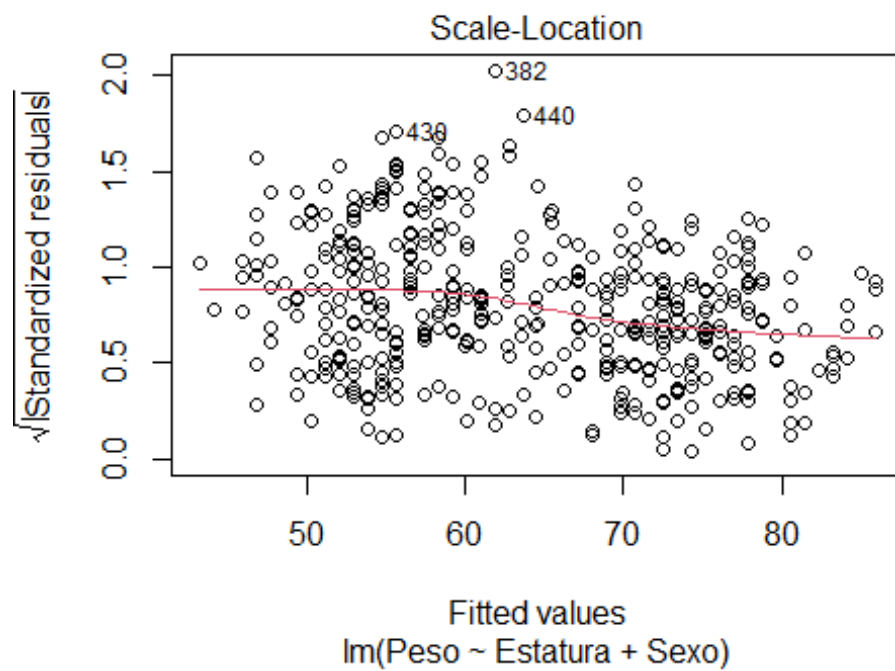
```
##
## Durbin-Watson test
##
## data: Modelo2
## DW = 1.8663, p-value = 0.07325
## alternative hypothesis: true autocorrelation is greater than 0
```

Debido a que el valor de la prueba de Durbin-Watson es cercano a 2, podemos entender que los residuos son independientes, por lo que aceptamos  $h_0$  ya que los errores no están correlacionados.

### **Gráficas del modelo**

```
plot(Modelo2)
```





### Intervalos de confianza

A = Modelo2

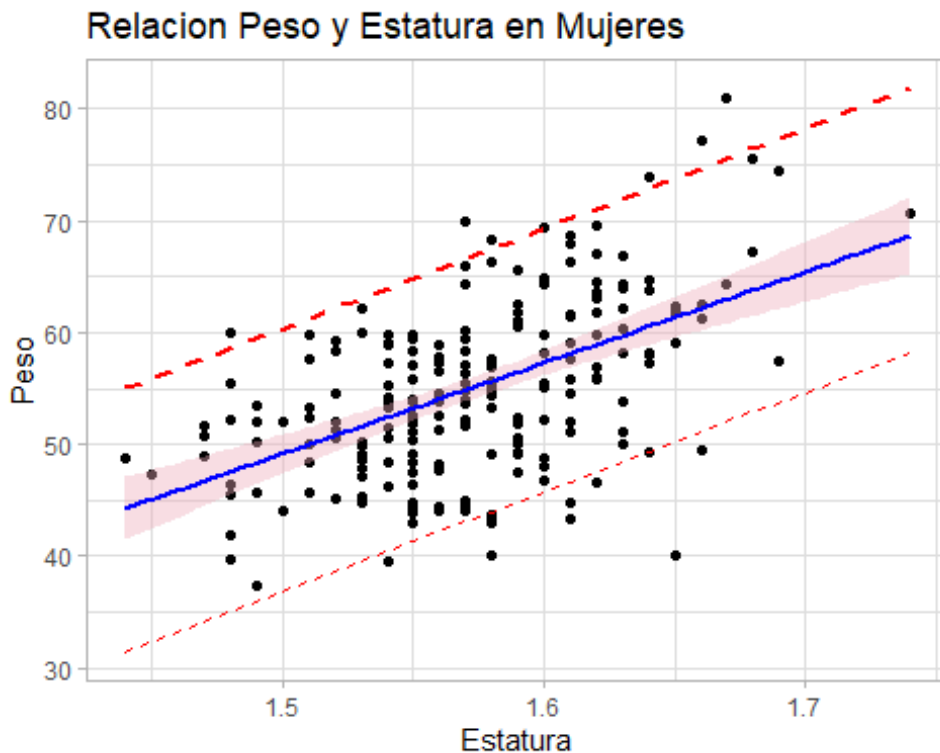
Ip=predict(object=A,interval="prediction",level=0.97)

```
## Warning in predict.lm(object = A, interval = "prediction", level =
0.97): predictions on current data refer to _future_ responses
```

```
M2=cbind(M,Ip)
M2m = subset(M2, Sexo=="M")
M2h = subset(M2, Sexo=="H")

library(ggplot2)
ggplot(M2m,aes(x=Estatura,y=Peso))+
  ggtitle("Relacion Peso y Estatura en Mujeres")+
  geom_point()+
  geom_line(aes(y=lwr), color="red", linetype="dashed")+
  geom_smooth(aes(y=upr), color="red", linetype="dashed")+
  geom_smooth(method=lm, formula=y~x, se=TRUE, level=0.97, col="blue",
fill="pink2")+
  theme_light()

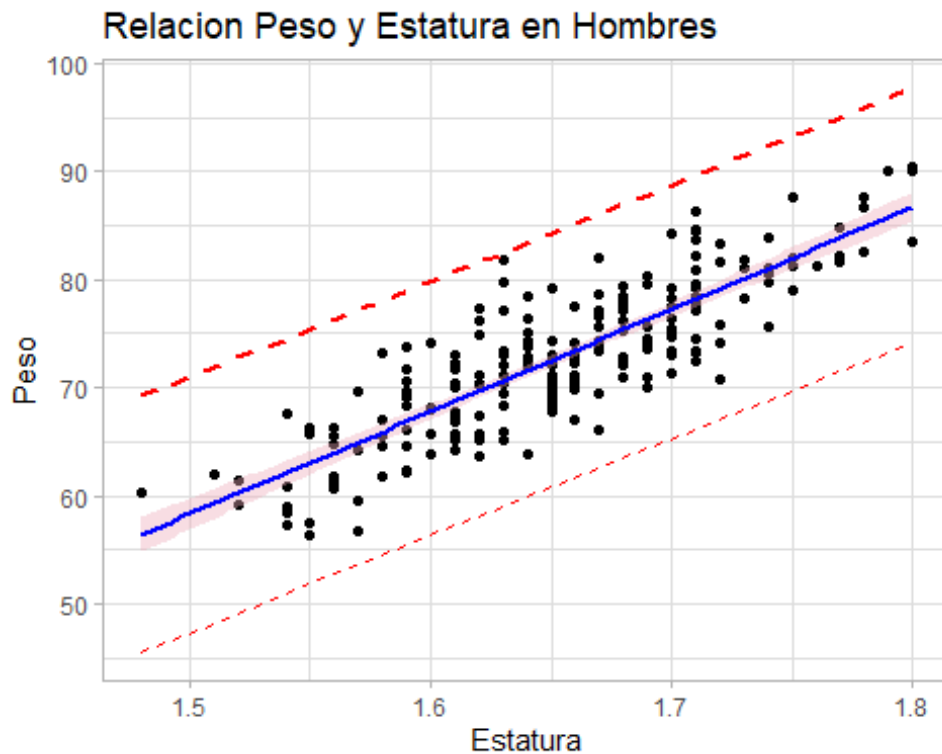
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



```
ggplot(M2h,aes(x=Estatura,y=Peso))+
  ggtitle("Relacion Peso y Estatura en Hombres")+
  geom_point()+
  geom_line(aes(y=lwr), color="red", linetype="dashed")+
  geom_smooth(aes(y=upr), color="red", linetype="dashed")+
  geom_smooth(method=lm, formula=y~x, se=TRUE, level=0.97, col="blue",
fill="pink2")+
  theme_light()
```



```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



En la gráfica de Hombres no existen datos atípicos, y podemos entender que el modelo para predecir peso con base a la estatura de un hombre es muy preciso, a diferencia de la gráfica de las mujeres, la cuál si posee varios datoa atípicos, y no es muy precisa para predecir pesos con base a la altura.