

AnalisisErrores

Cleber Perez

2024-09-04

Leer archivo

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

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

Validez del modelo / Analisis de los residuos

```
A = lm(Peso~Estatura, M)
A

##
## Call:
## lm(formula = Peso ~ Estatura, data = M)
##
## Coefficients:
## (Intercept)      Estatura
##      -151.9         133.8

summary(A)

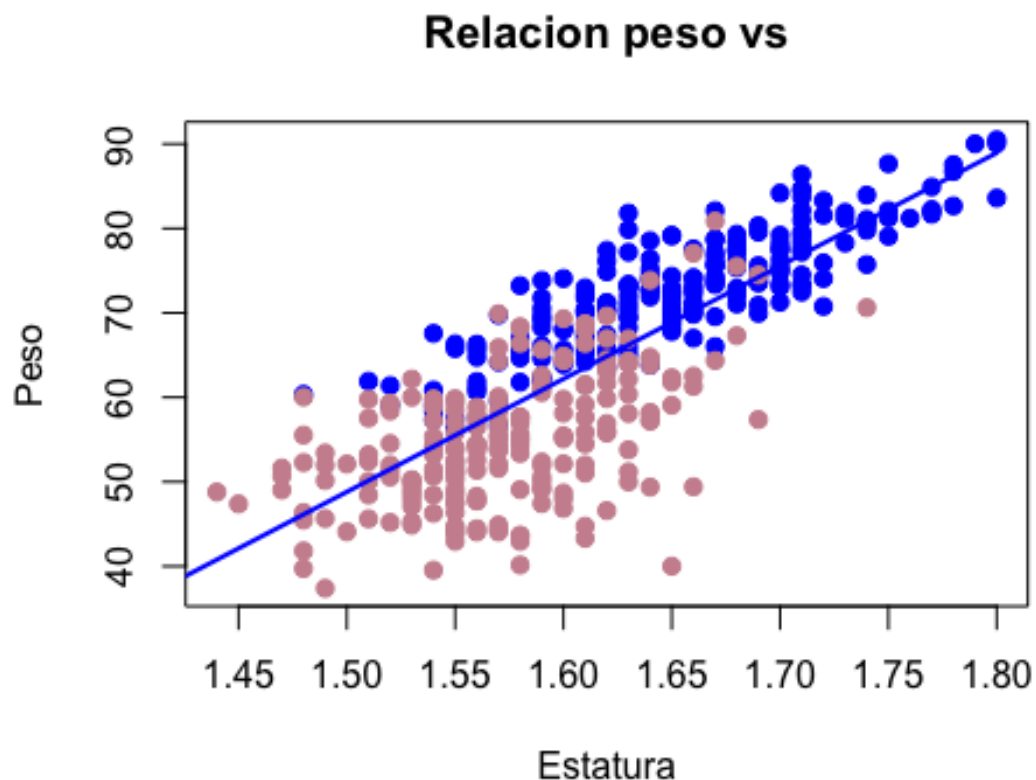
##
## Call:
## lm(formula = Peso ~ Estatura, data = M)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.8653  -3.7654   0.6706   5.0142  15.6006
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -151.883     7.655  -19.84  <2e-16 ***
## Estatura      133.793     4.741   28.22  <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.883 on 438 degrees of freedom
## Multiple R-squared:  0.6452, Adjusted R-squared:  0.6444
## F-statistic: 796.5 on 1 and 438 DF,  p-value: < 2.2e-16

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

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

colores = c("blue", "pink3")
plot(M$Estatura, M$Peso, col=colores[factor(M$Sexo)], pch=19,
     xlab="Estatura", ylab="Peso", main="Relacion peso vs")
x = seq(1.40, 1.80, 0.01)
lines(x, Ym(x), col="pink2", lwd=2)
lines(x, Yh(x), col="blue", lwd=2)
```

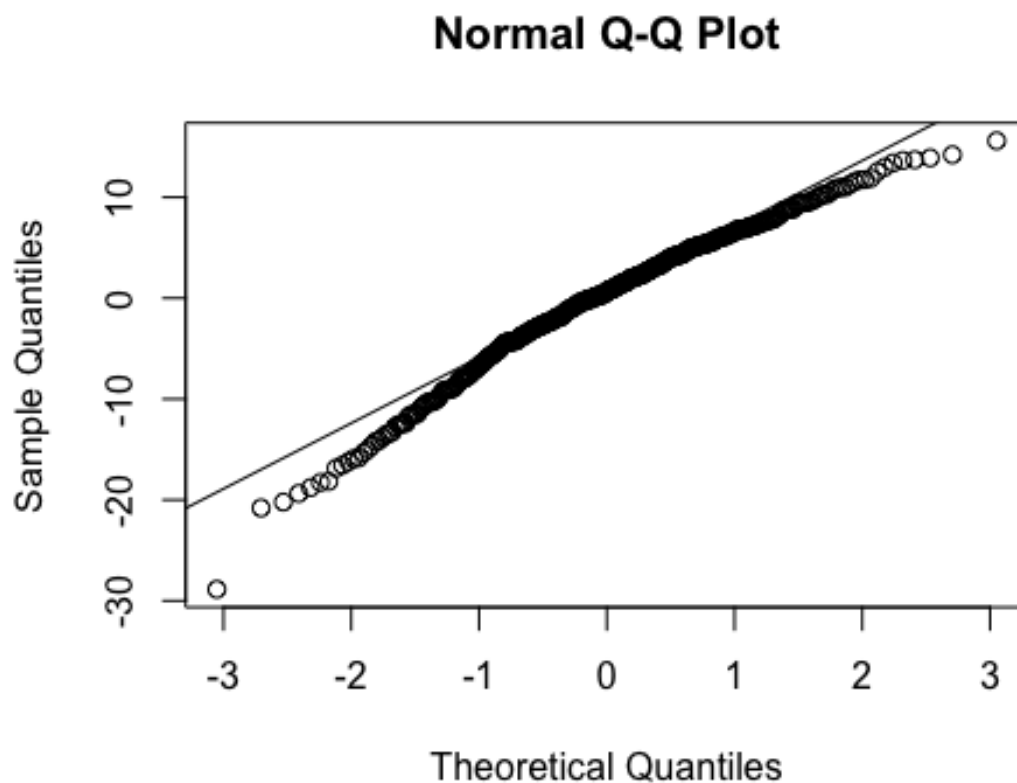


Normalidad de los residuos

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

##
##  Anderson-Darling normality test
##
## data:  A$residuals
## A = 2.4766, p-value = 2.888e-06

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

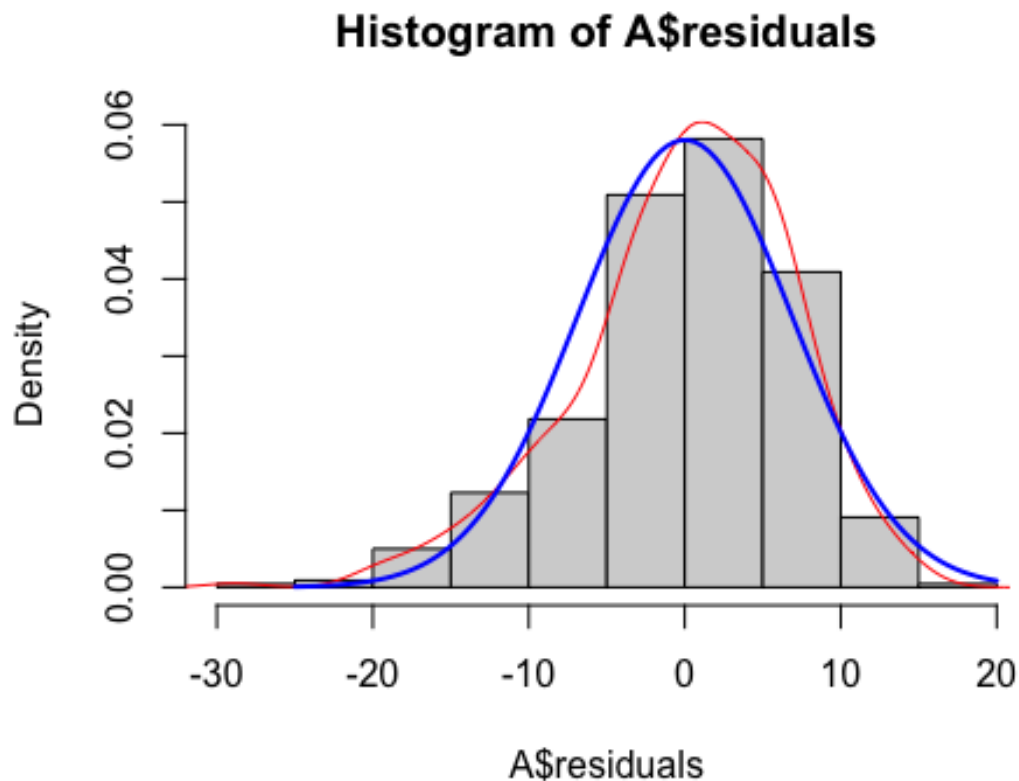


```
t.test(A$residuals)

##
##  One Sample t-test
##
## data:  A$residuals
## t = -2.3405e-15, df = 439, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.6441362  0.6441362
## sample estimates:
```

```
##      mean of x
## -7.670632e-16

hist(A$residuals,freq=FALSE)
lines(density(A$residual),col="red")
curve(dnorm(x,mean=mean(A$residuals),sd=sd(A$residuals)), from=-25, to=20,
add=TRUE, col="blue",lwd=2)
```



Verificacion de media cero

```
A = lm(Peso~Estatura, M)
t.test(A$residuals)

##
##  One Sample t-test
##
## data:  A$residuals
## t = -2.3405e-15, df = 439, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.6441362  0.6441362
## sample estimates:
##      mean of x
## -7.670632e-16
```

Homocedastidad e independencia

Homocedastidad (La varianza de los errores no es constante (heterocidad))

```
library(lmtest)

## Loading required package: zoo

##
## Attaching package: 'zoo'

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

bptest(A)

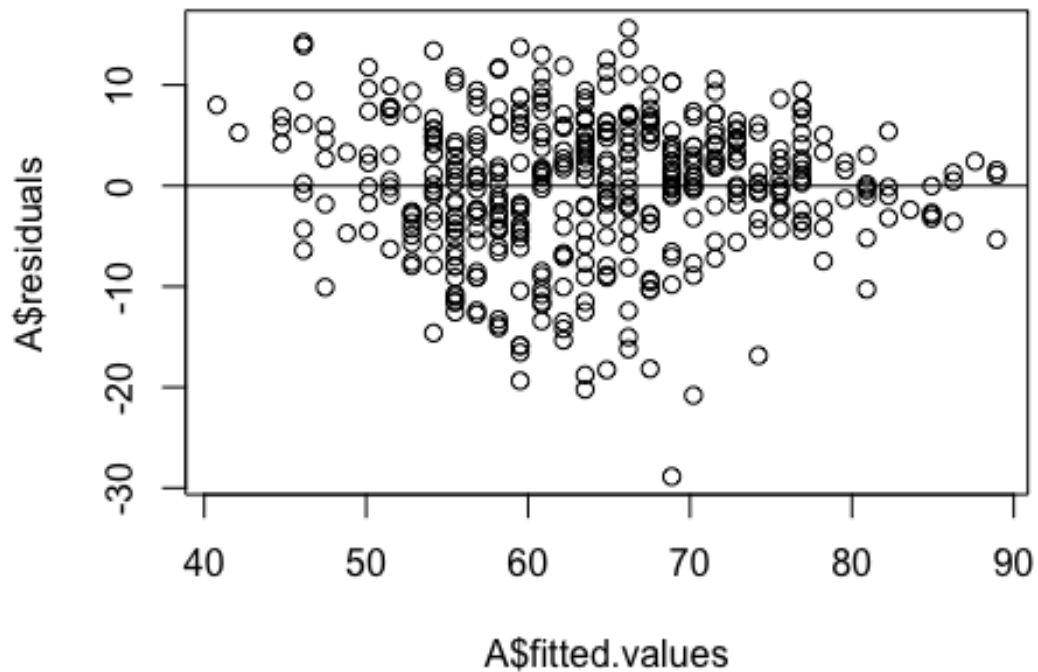
##
## studentized Breusch-Pagan test
##
## data: A
## BP = 5.7194, df = 1, p-value = 0.01678

gqtest(A)

##
## Goldfeld-Quandt test
##
## data: A
## GQ = 3.2684, df1 = 218, df2 = 218, p-value < 2.2e-16
## alternative hypothesis: variance increases from segment 1 to 2

plot(A$fitted.values,A$residuals)
abline(h=0, color="blue")

## Warning in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...):
## "color" is
## not a graphical parameter
```



Independencia (Los errores estan autocorrelacionados)

```
library(lmtest)
bptest(A)

##
## studentized Breusch-Pagan test
##
## data: A
## BP = 5.7194, df = 1, p-value = 0.01678

gqtest(A)

##
## Goldfeld-Quandt test
##
## data: A
## GQ = 3.2684, df1 = 218, df2 = 218, p-value < 2.2e-16
## alternative hypothesis: variance increases from segment 1 to 2
```

Linealidad (Hay una especificacion erronea en el modelo que indica no linealidad)

```
library(lmtest)
resettest(A)
```

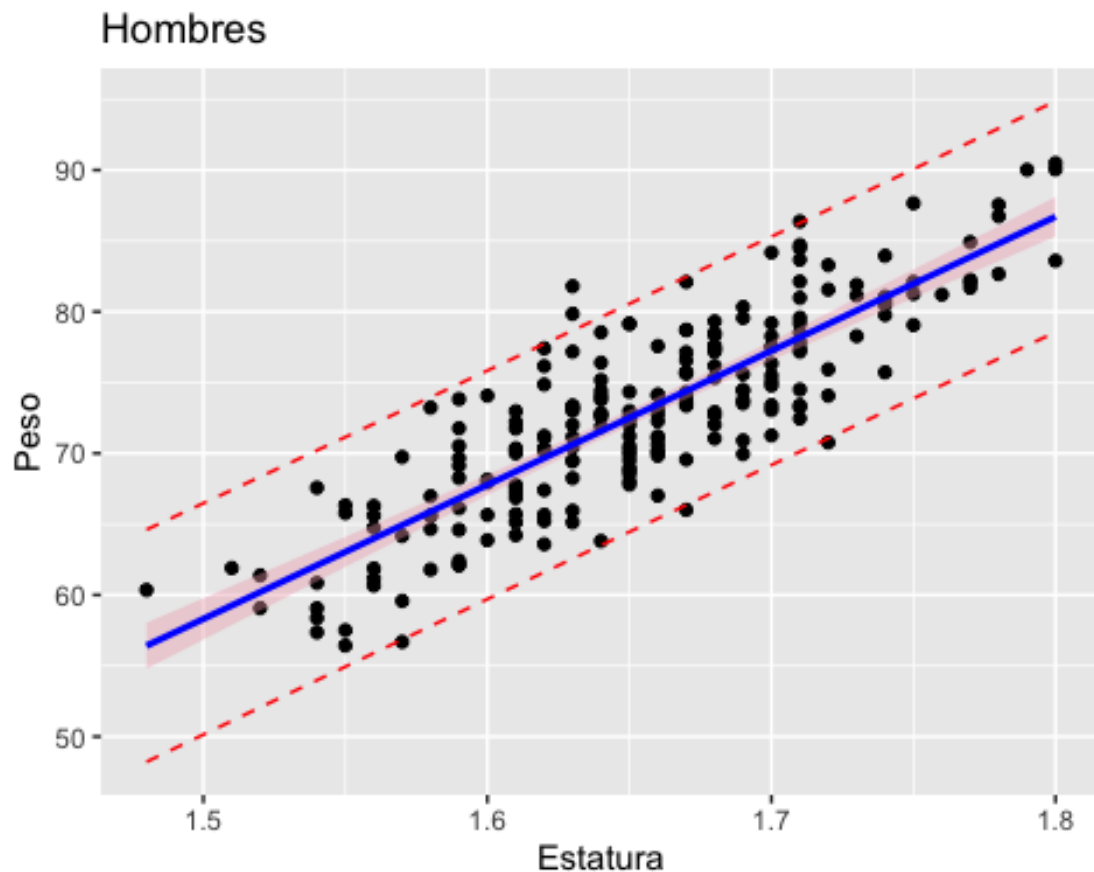
```
##
## RESET test
##
## data: A
## RESET = 6.9691, df1 = 2, df2 = 436, p-value = 0.001049
```

Graficas de los intervalos en regresion

```
Modelo2H = lm(Peso~Estatura, MH)
Ip=predict(object=Modelo2H,interval="prediction",level=0.97)

## Warning in predict.lm(object = Modelo2H, interval = "prediction", level =
0.97): predictions on current data refer to _future_ responses

datos1=cbind(MH,Ip)
library(ggplot2)
ggplot(datos1,aes(x=Estatura,y=Peso))+
  geom_point()+
  geom_line(aes(y=lwr), color="red", linetype="dashed")+
  geom_line(aes(y=upr), color="red", linetype="dashed")+
  geom_smooth(method=lm, formula=y~x, se=TRUE, level=0.97, col="blue",
fill="pink2")+
  ggtitle("Hombres")
```



```
theme_light()
```

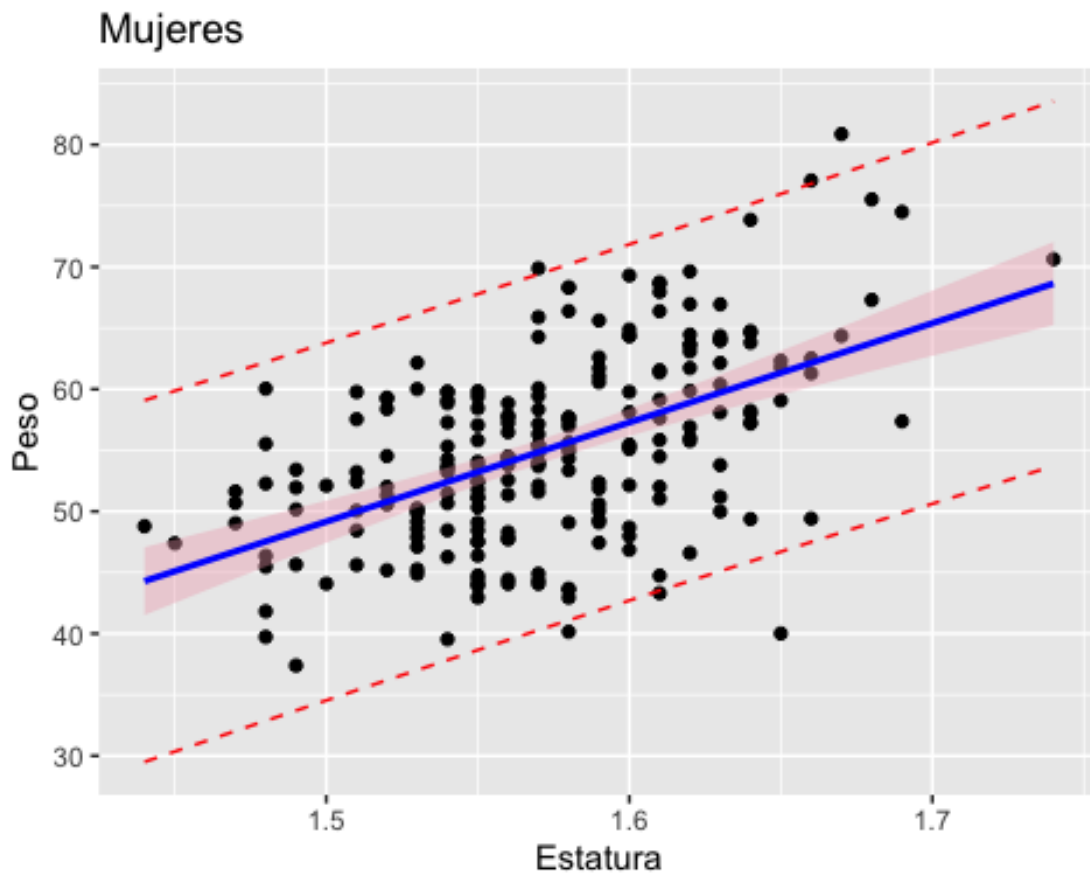
```

Modelo2M = lm(Peso~Estatura, MM)
Ip=predict(object=Modelo2M,interval="prediction",level=0.97)

## Warning in predict.lm(object = Modelo2M, interval = "prediction", level =
0.97): predictions on current data refer to _future_ responses

datos1=cbind(MM,Ip)
library(ggplot2)
ggplot(datos1,aes(x=Estatura,y=Peso))+
  geom_point()+
  geom_line(aes(y=lwr), color="red", linetype="dashed")+
  geom_line(aes(y=upr), color="red", linetype="dashed")+
  geom_smooth(method=lm, formula=y~x, se=TRUE, level=0.97, col="blue",
fill="pink2")+
  ggtitle("Mujeres")

```



```

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

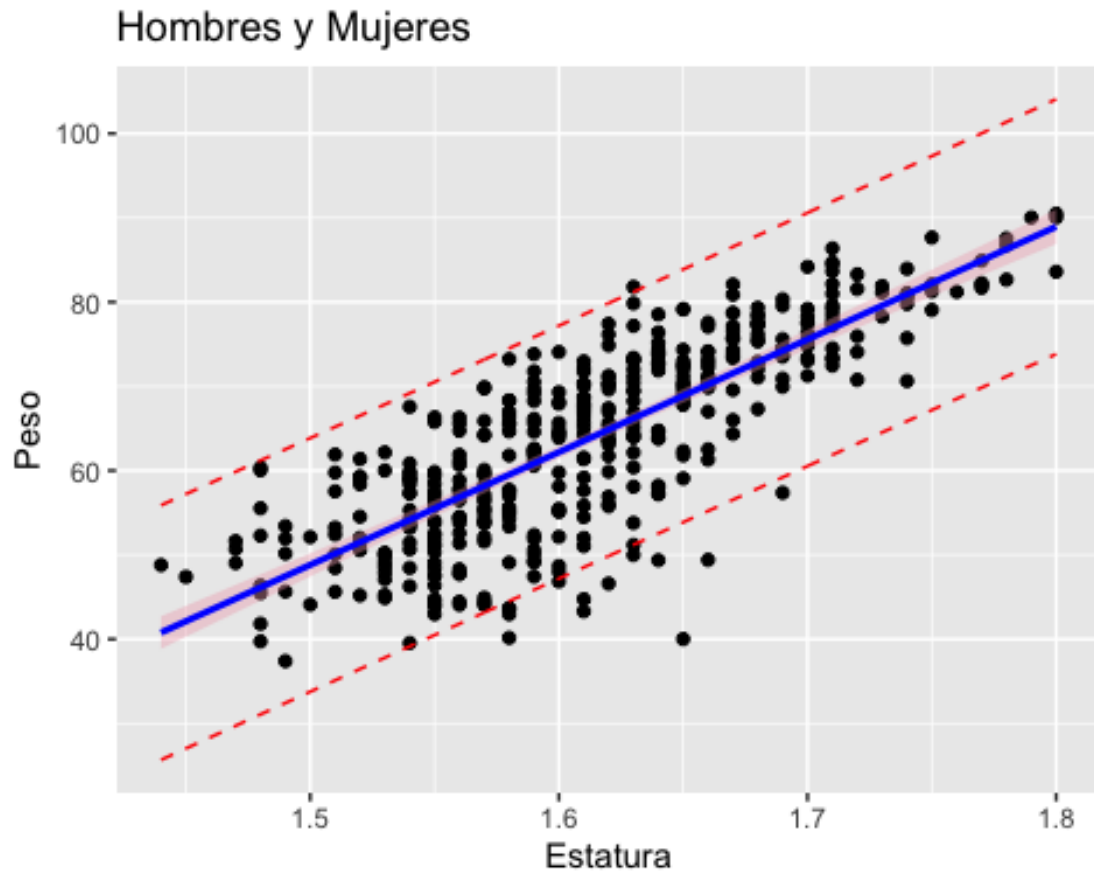
## Warning in predict.lm(object = A, interval = "prediction", level = 0.97):
predictions on current data refer to _future_ responses

datos1=cbind(M,Ip)
library(ggplot2)
ggplot(datos1,aes(x=Estatura,y=Peso))+

```



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



Con estas 3 ultimas graficas podemos observar que si partimos el modelo en hombre y mujeres nos daran respuestas diferentes a que si los tomamos en conjunto, si tomamos todas las variables como una nos da resultados de heterocidad, al contrario de probarlas por separado, el de hombres si demuestra una homocedastidad y el mujeres dependiendo la prueba da homocedastidad y heterocidadidad