Actividad 10

Saúl Francisco Vázquez del Río

2024-08-30

Primer problema

Analiza la base de datos de estatura y peso Download datos de estatura y pesode los hombres y mujeres en México y obten el mejor modelo de regresión para esos datos

```
M=read.csv("C:\\Users\\saulv\\OneDrive\\Escritorio\\Septimo 
semestre\\Estatura-peso_HyM.csv") #leer la base de datos
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)
```

Matriz de correlacion

Obtén la matriz de correlación de los datos que se te proporcionan. Interpreta.

```
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
```

La matriz de correlacion establece las relaciones entre peso y estatura de hombres y mujeres.

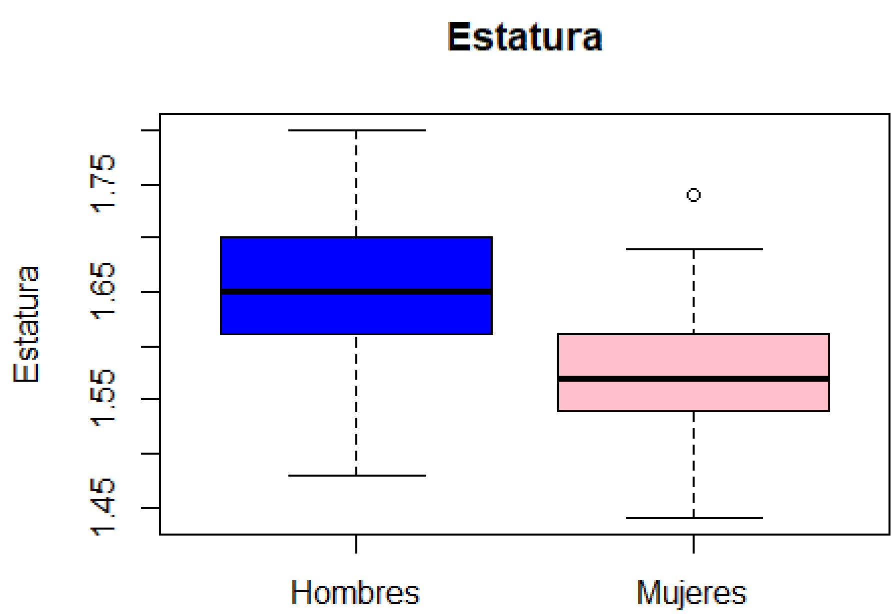
```
n=4 #número de variables
d=matrix(NA,ncol=7,nrow=n)
for(i in 1:n){
  d[i,]<-c(as.numeric(summary(M1[,i])),sd(M1[ ,i]))
}
m=as.data.frame(d)
```

```r
row.names(m)=c("H-Estatura","H-Peso","M-Estatura","M-Peso")
names(m)=c("Minimo","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
```

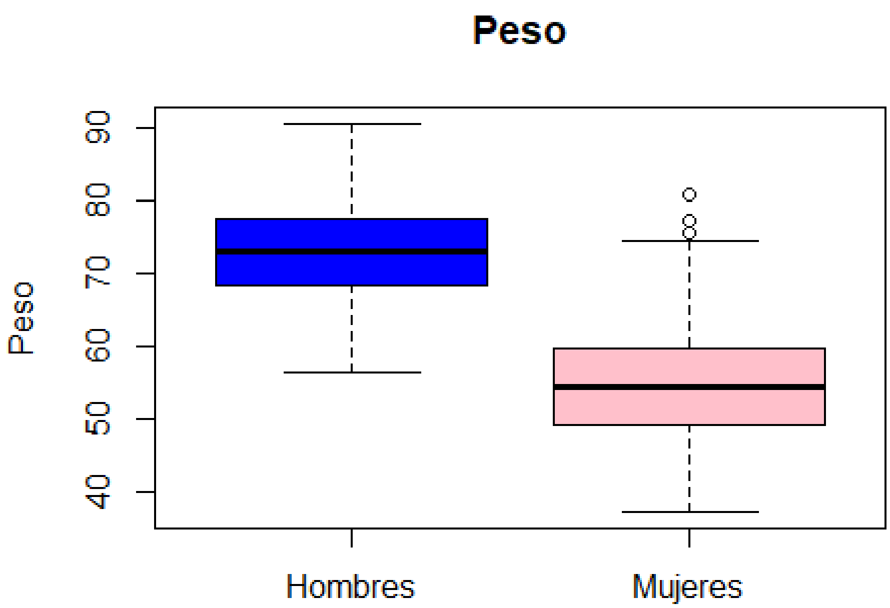
Establecemos la variables como las mediana y la media para entedner mejor los datos

Box plot

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




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



En el box plot se muestran la relacion que tienen los hombres y las mujeres con la estatura y con el peso

Hombres

```
Modelo1H = lm(Peso~Estatura, data = MH)
Modelo1H

##
## Call:
## lm(formula = Peso ~ Estatura, data = MH)
##
## Coefficients:
## (Intercept)     Estatura
##      -83.68        94.66

Modelo1M = lm(Peso~Estatura, data = MM)
Modelo1M

##
## Call:
## lm(formula = Peso ~ Estatura, data = MM)
##
## Coefficients:
## (Intercept)     Estatura
##      -72.56        81.15
```

Se hacen los modelos de hombres y mujeres para ver la relacion entre el peso y la estatura

Hipotesis

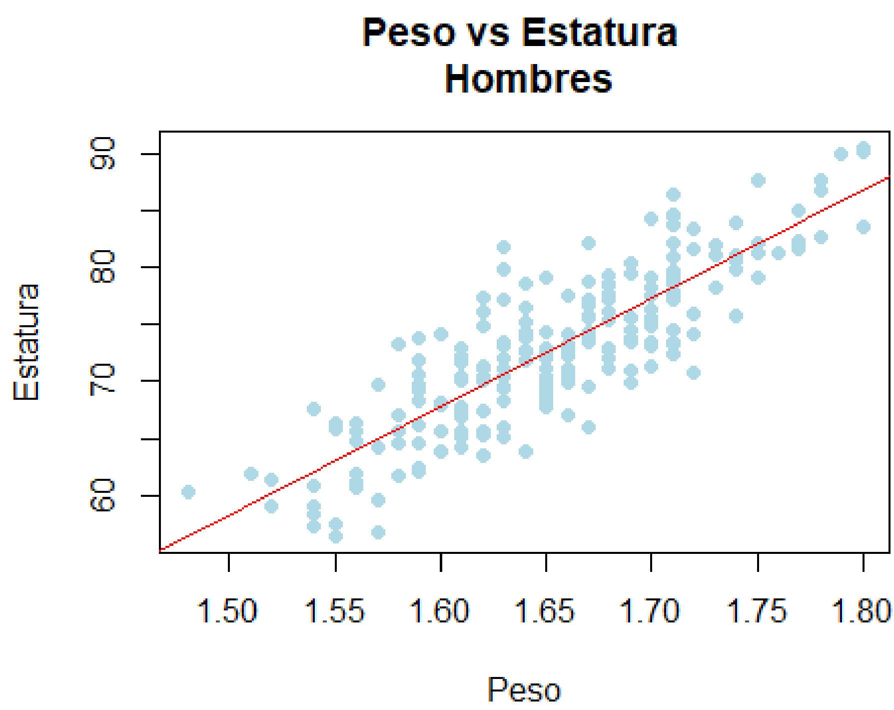
$H_0: \beta_1 = 0$ $H_1: _1 $

```r
summary(Modelo1H)
```

```
##
## Call:
## lm(formula = Peso ~ Estatura, data = MH)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.3881 -2.6073 -0.0665  2.4421 11.1883
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -83.685      6.663  -12.56   <2e-16 ***
## Estatura      94.660      4.027   23.51   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.678 on 218 degrees of freedom
## Multiple R-squared:  0.7171, Adjusted R-squared:  0.7158
## F-statistic: 552.7 on 1 and 218 DF,  p-value: < 2.2e-16
```

Grafico de dispercion para hombres

```r
plot(MH$Estatura, MH$Peso, col='lightblue', main="Peso vs Estatura \n
Hombres", ylab="Estatura", xlab="Peso", pch = 19 )
abline(Modelo1H, col="red", lwd=1.5)
```



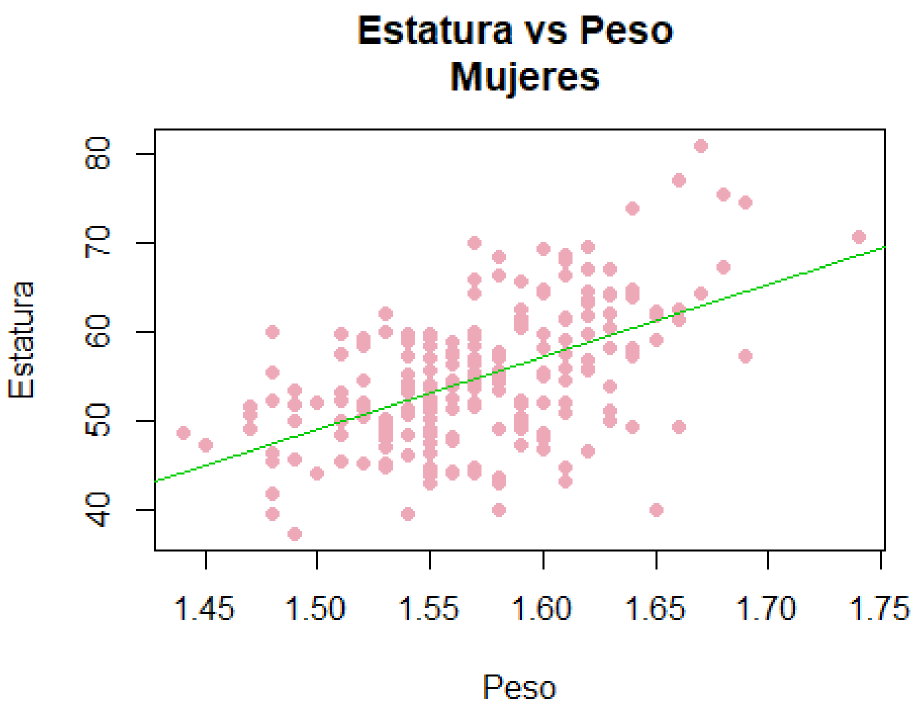
La grafica muestra la relacion entre el peso y la estatura de los hombres

Mujeres

```
summary(Modelo1M)

##
## Call:
## lm(formula = Peso ~ Estatura, data = MM)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.3256  -4.1942   0.4004   4.2724  17.9114
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -72.560     14.041  -5.168 5.34e-07 ***
## Estatura      81.149      8.922   9.096  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.65 on 218 degrees of freedom
## Multiple R-squared:  0.2751, Adjusted R-squared:  0.2718
## F-statistic: 82.73 on 1 and 218 DF,  p-value: < 2.2e-16
```

Grafico para mujeres

```
plot(MM$Estatura, MM$Peso, col='pink2', main="Estatura vs Peso \n
Mujeres", ylab="Estatura", xlab="Peso", pch = 19 )
abline(Modelo1M, col="Green3", lwd=1.5)
```



La grafica muestra la relacion entre el peso y la estatura de las mujeres

Un solo modelo

```
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
```

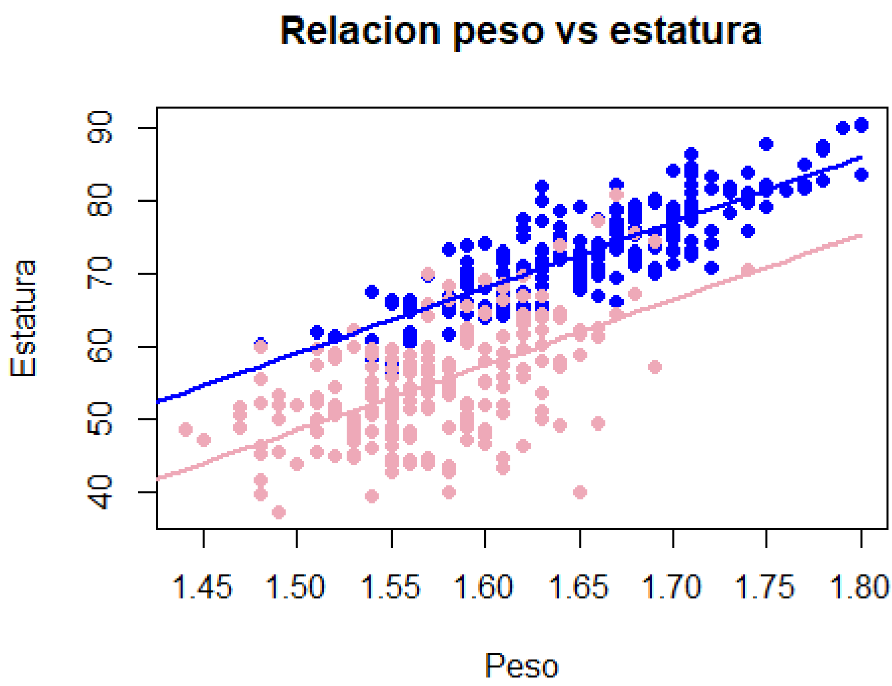
Se crea un modelo en el cual se pueden ver las diferencias del peso y la estatura entre hombres y mujeres

A 0.05 sí es significativo y los modelos quedarian

Hombres: Estatura = 1.3862 + 0.00399P Estatura = 1.2727097 + 0.0052296P + 0.0121799 Estatura = 1.2848 + 0.0052296P

Mujeres: Estatura = 1.3862 + 0.00399P Estatura = 1.2727097 + 0.0052296P + 0.0121799 Estatura = 1.2727097 + 0.0052296P

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



La grafica muestra las diferencias de peso y estatura de hombres y mujeres

Interpreta en el contexto del problema: ¿Qué información proporciona β̂0 sobre la relación entre la estatura y el peso de hombres y mujeres? Los modelos presentan la relacion que los hombres y las mujeres tienen con su peso y estatura, ademas que se nota un control de peso por la estatura.

¿Cómo interpretas β̂1 en la relación entre la estatura y el peso de hombres y mujeres? La relacion se puede interpretar que esta varia dependiendo del sexo que se esta tratando ya que los hombres tienden a ser más altos lo que conlleva a que estos pesen más, al igual en el lado de las mujeres estas tienden a ser más bajitas y por esto tienden a pesar menos.

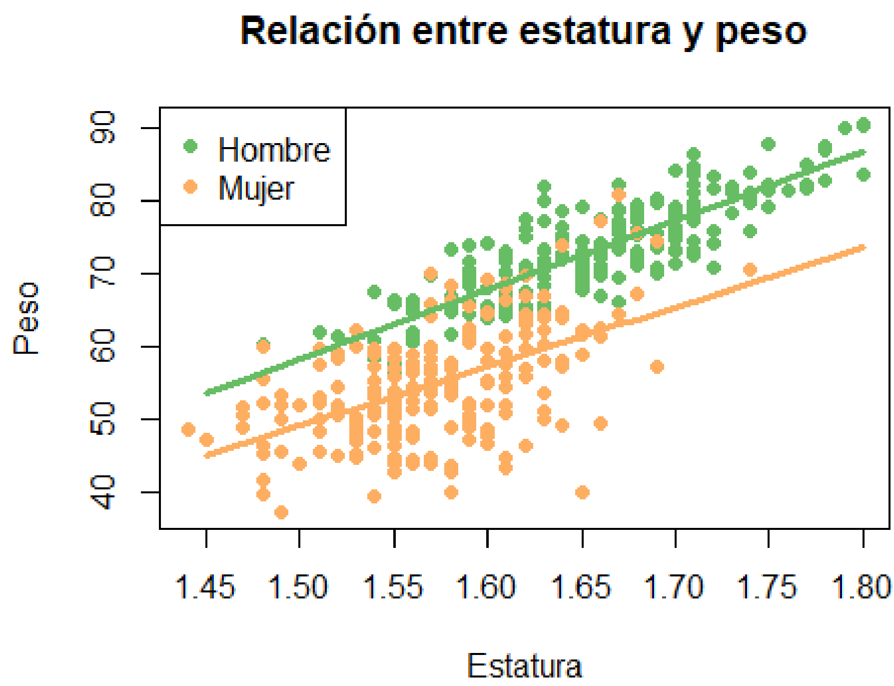
Otro Modelo (segunda entrega)

```
Modelo3 = lm(Peso~Estatura * Sexo,data = M)
Modelo3

##
## Call:
## lm(formula = Peso ~ Estatura * Sexo, data = M)
##
## Coefficients:
##    (Intercept)        Estatura           SexoM  Estatura:SexoM
##         -83.68           94.66           11.12          -13.51
```

Formula de hombres = -83.68454+ 11.12409 +94.66024 Formula de mujeres = -83.68454 +94.66024

```r
summary(Modelo3)

##
## Call:
## lm(formula = Peso ~ Estatura * Sexo, data = M)
##
## 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 ***
## Estatura           94.660      5.882  16.092   <2e-16 ***
## SexoM              11.124     14.950   0.744    0.457
## Estatura: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

anova(Modelo3)

## Analysis of Variance Table
##
## Response: Peso
##                  Df Sum Sq Mean Sq   F value Pr(>F)
## Estatura          1  37731   37731 1306.5938 <2e-16 ***
## Sexo              1   8097    8097  280.3892 <2e-16 ***
## Estatura:Sexo     1     61      61    2.1085 0.1472
## Residuals       436  12590      29
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

rsq_adj <- summary(Modelo3)$adj.r.squared
print(paste("R-squared ajustado: ", rsq_adj))

## [1] "R-squared ajustado:  0.783219679867522"

# Asignar los coeficientes del modelo con interacción
b0_A <- Modelo3$coefficients[1]
b1_A <- Modelo3$coefficients[2]
b2_A <- Modelo3$coefficients[3]
b3_A <- Modelo3$coefficients[4]

# Imprimir los coeficientes
print(b0_A)
```

```
## (Intercept)
##   -83.68454

print(b1_A)

## Estatura
## 94.66024

print(b2_A)

##    SexoM
## 11.12409

print(b3_A)

## Estatura:SexoM
##      -13.51113

# Funciones para las líneas de regresión
Ym <- function(x) { b0_A + b2_A + (b1_A + b3_A) * x }  # Línea para
mujeres
Yh <- function(x) { b0_A + b1_A * x }                  # Línea para
hombres

# Colores para los puntos y las líneas
colores <- c("#66BD63", "#FDAE61")

# Crear la gráfica de dispersión
plot(M$Estatura, M$Peso, col=colores[factor(M$Sexo)], pch=19,
ylab="Peso", xlab="Estatura", main="Relación entre estatura y peso")

# Generar secuencia de valores para las líneas de regresión
x <- seq(1.45, 1.80, 0.01)

# Añadir las líneas de regresión al gráfico
lines(x, Ym(x), col="#FDAE61", lwd=3)  # Línea para mujeres
lines(x, Yh(x), col="#66BD63", lwd=3)  # Línea para hombres

# Añadir leyenda al gráfico
legend("topleft", legend=c("Hombre", "Mujer"), pch=19, col=c("#66BD63",
"#FDAE61"))
```



B_0: respresenta el promedio del peso cuando la estarua de los hombres 0, siendo esto el punto de inicio / referencia para emprezar a comprar las diferencias de sexo.

B_1: Muestra el cambio de peso en los hombres

B_2: Muestra el cambio de peso en las mujeres, pero comparado con el de los hombres

B_3: Muestra la relacion entre la estatura de las mujeres comparado con la de los hombres

Para concluir el modelo3 no aporta un valor adiccional a la relacion entre estatura y peso, ya que tiene falta de significancia ademas de una complejidad inecesaria para la interpretacion de los datos.

Tercer problema

El Validez del modelo

Analiza si el (los) modelo(s) obtenidos anteriormente son apropiados para el conjunto de datos. Realiza el análisis de los residuos:

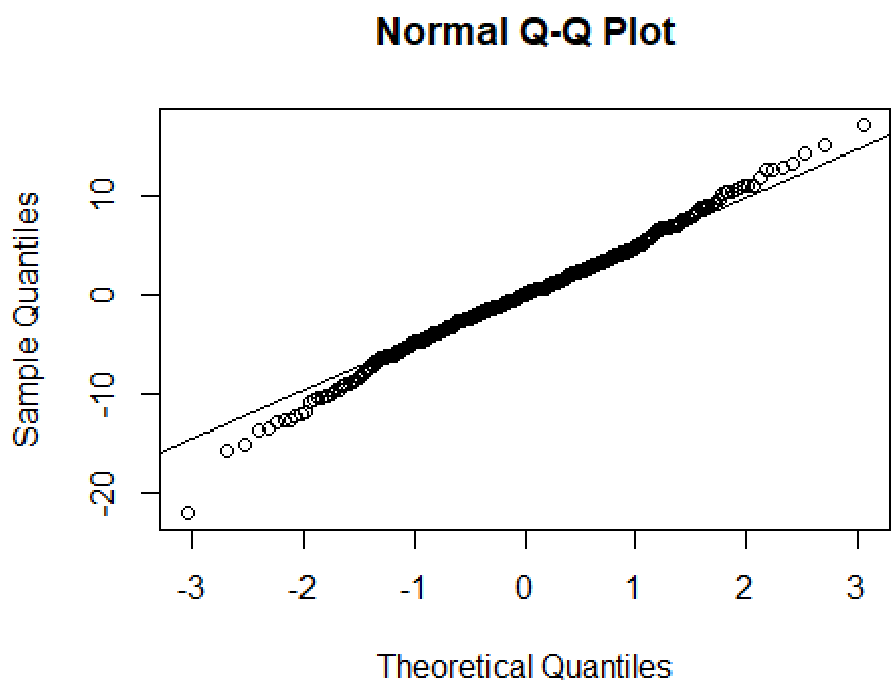
Normalidad de los residuos

H0: Los datos provienen de una población normal H1: Los datos no provienen de una población normal

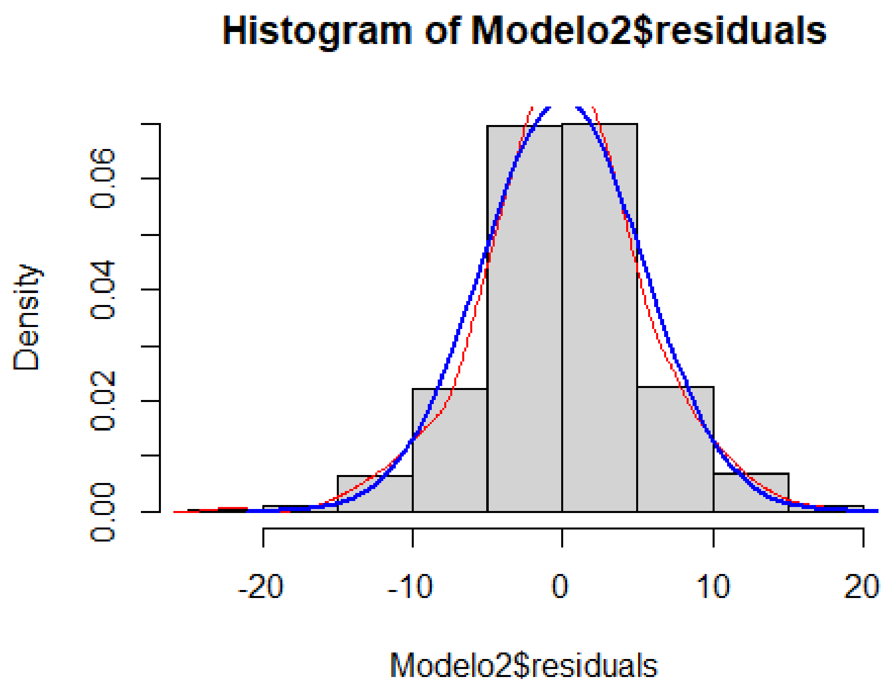
```
library(nortest)
ad.test(Modelo2$residuals)
```

```
##
##  Anderson-Darling normality test
##
## data:  Modelo2$residuals
## A = 0.79651, p-value = 0.03879
```

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




```
hist(Modelo2$residuals,freq=FALSE)
lines(density(Modelo2$residual),col="red")
curve(dnorm(x,mean=mean(Modelo2$residuals),sd=sd(Modelo2$residuals)),
from=-
21, to=21, add=TRUE, col="blue",lwd=2)
```



H0 no se rechaza ya que el valorp es menor que 0.03

Verificación de media cero

H0: $\mu = 0$ H1: $\mu \neq 0$

```
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
```

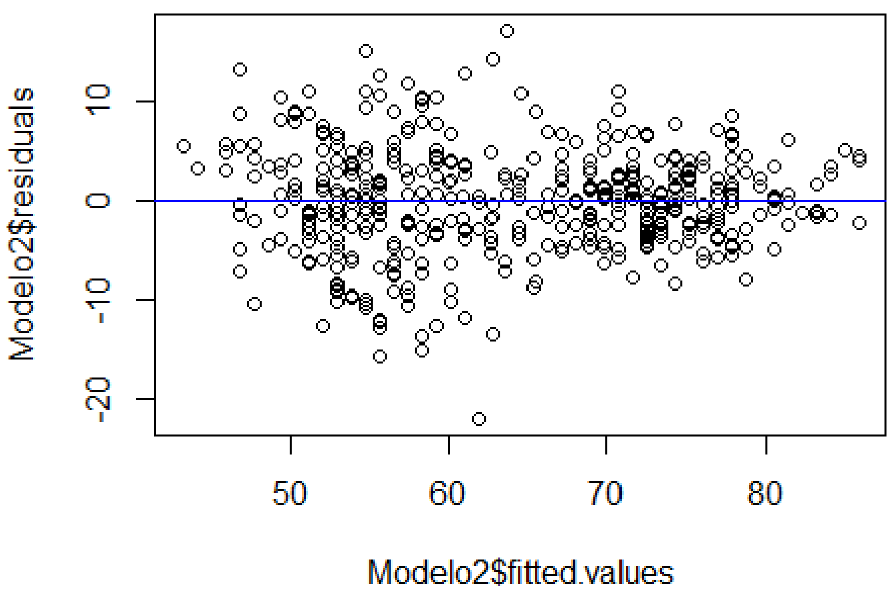
Se rechaza H0 porque es mayor a 0

Homocedasticidad e independencia

Homocedasticidad H0: La varianza de los errores es constante (homocedasticidad) H1: La varianza de los errores no es constante (heterocedasticidad)

Independencia H0: Los errores no están correlacionados H1: Los errores están correlacionados

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

library(lmtest)

## Cargando paquete requerido: zoo

##
## Adjuntando el paquete: 'zoo'

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




```r
dwtest(Modelo2)

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

bptest(Modelo2)

##
##  studentized Breusch-Pagan test
##
```

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

Homocestacidad

H1 se acepta porque la distribucion no es homegenia

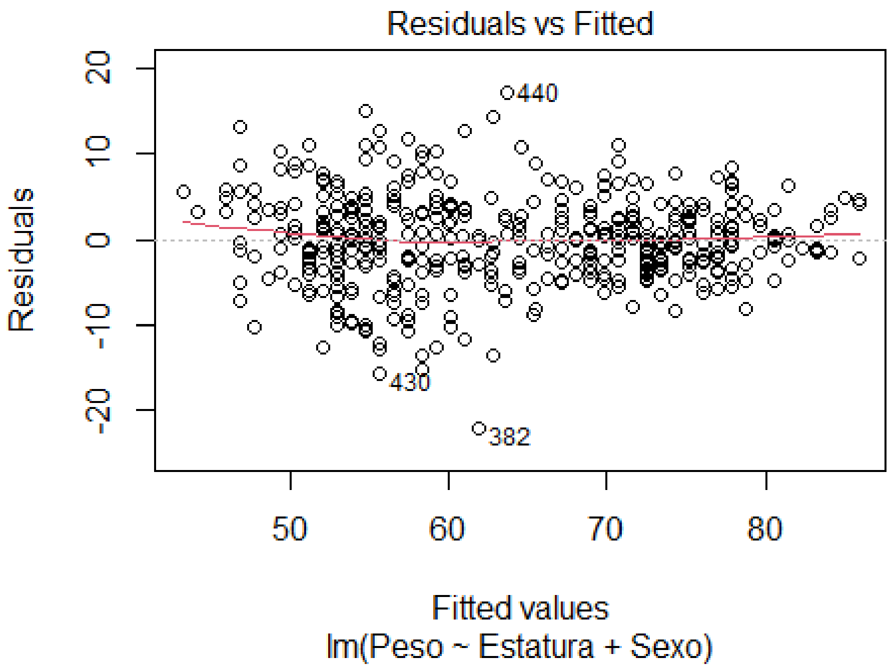
Independencia

H0 se acepta porque df es un valor cercano a 2

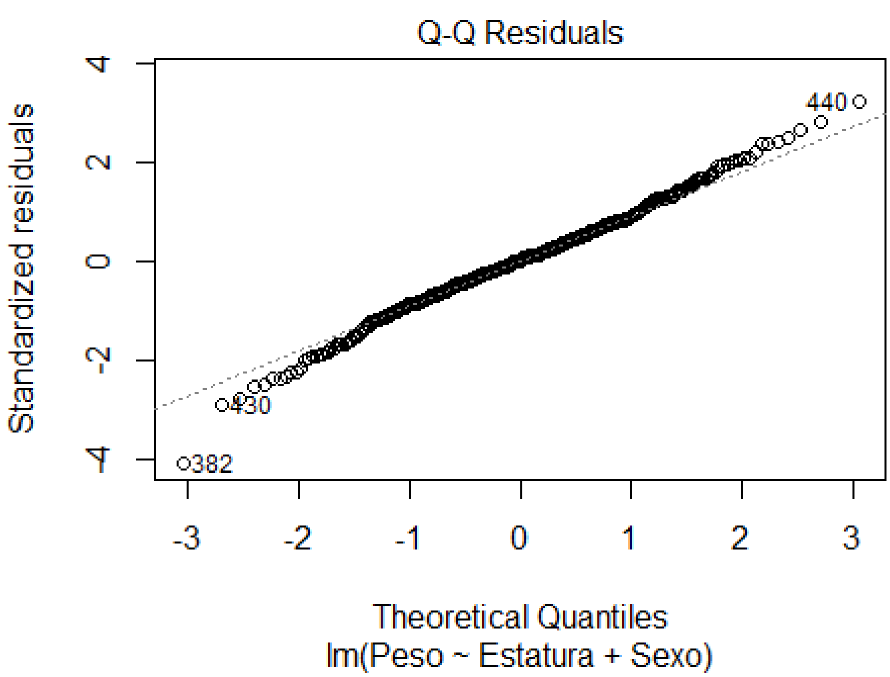
No te olvides de incluir las hipótesis en la pruebas de hipótesis que realices. Interpreta en el contexto del problema cada uno de los análisis que hiciste.

Utiliza el comando: plot(modelo). Observa las gráficas obtenidas y contesta: ¿Cuáles son las diferencias y similitudes de estos gráficos con respecto a los que ya habías analizado?

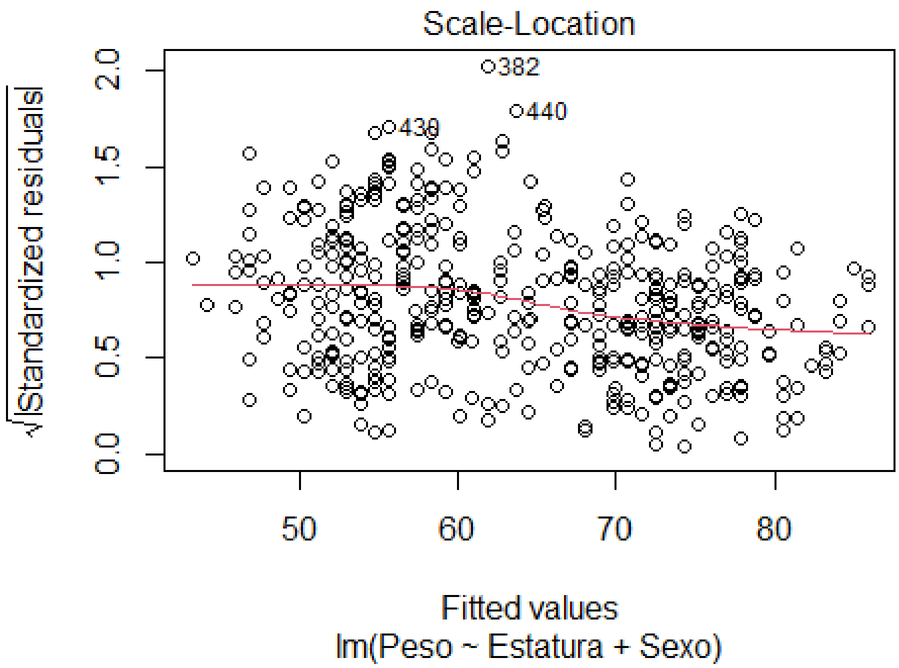
```
plot(Modelo2)
```

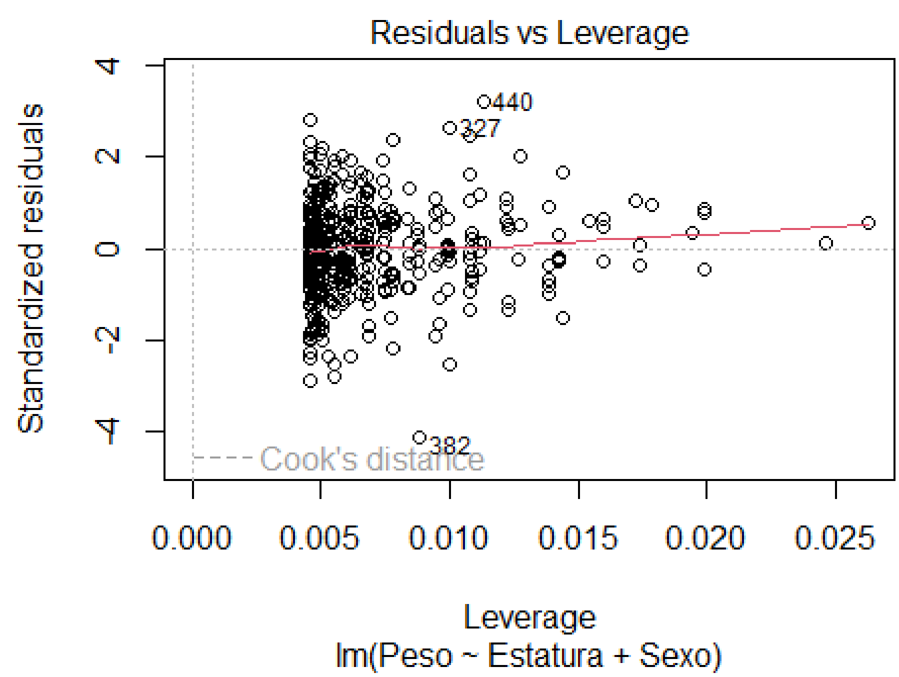


Residuals vs Fitted
Residuals
Fitted values
lm(Peso ~ Estatura + Sexo)
440
430
382
20
10
0
-10
-20
50
60
70
80



Q-Q Residuals
Standardized residuals
Theoretical Quantiles
lm(Peso ~ Estatura + Sexo)
440
430
382
4
2
0
-2
-4
-3
-2
-1
0
1
2
3





Similitudes: Los graficos no tienen un gran error

Diferencias: No noto una gran diferencia

Estos gráficos, ¿cambian en algo las conclusiones que ya habías obtenido?

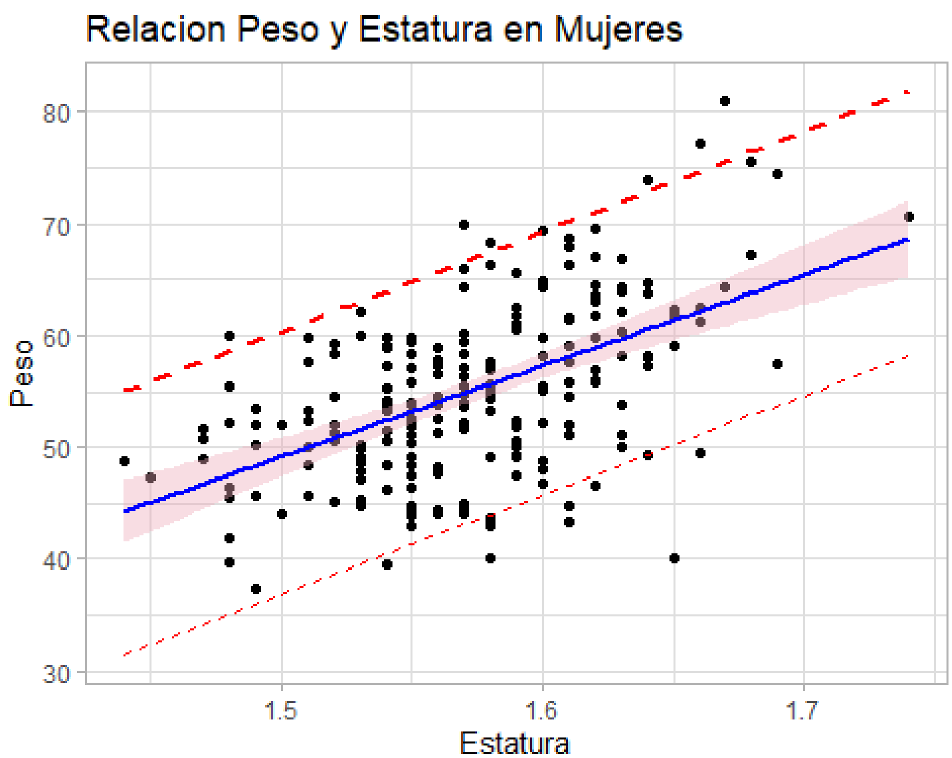
Emite una conclusión final sobre el mejor modelo de regresión lineal que conjunte lo que hiciste en las tres partes de esta actividad.

Como conclusion final ya obervando e interprentando los varios graficos el modelo2 sigue siendo el mejor modelo gracias a que este no tiene un gran error comprado con los otros modelos.

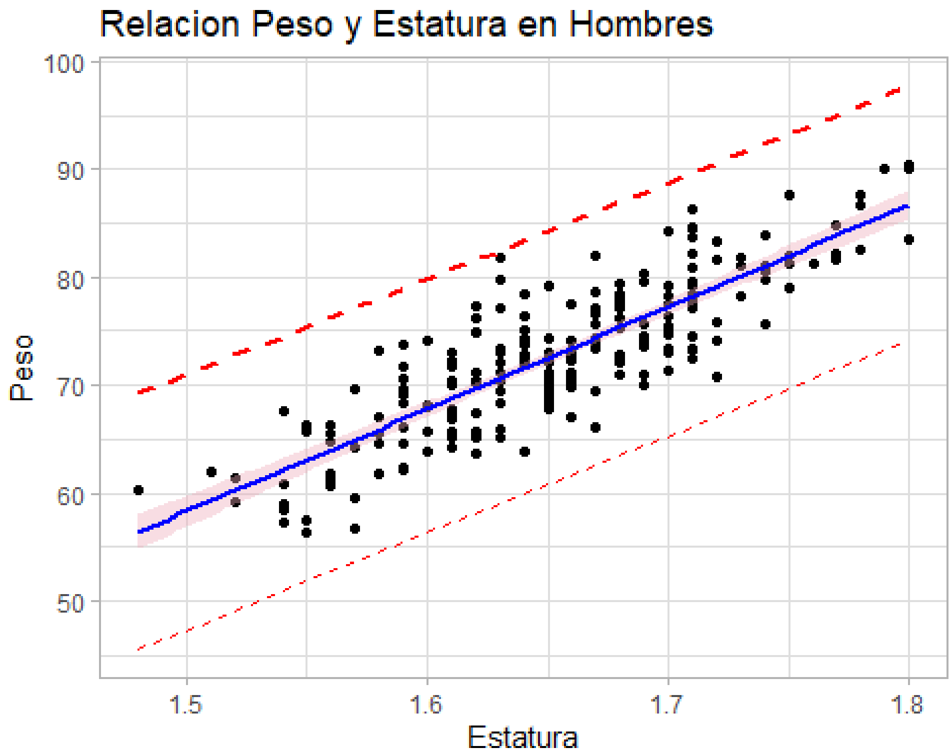
Intervalos de confianza

Con los datos de las estaturas y pesos de los hombres y las mujeres construye la gráfica de los intervalos de confianza y predicción para la estimación y predicción de Y para el mejor modelo seleccionado.

```
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'
```




```r
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'
```



Interpreta y comenta los resultados obtenidos

Se puede observar que en le grafico de hombres la relacion de estatura y peso más significativa ya que este no tiene datos atipicos y su modelo es muy exacto comparado con el modelo de las mujeres.