12. Regresion Lineal

Adrian Pineda Sanchez

2024-09-04

La recta de mejor ajuste (Primera entrega)

```
data = read.csv("Estatura_Peso.csv")

dataM = subset(data, data$Sexo=="M")
dataH = subset(data, data$Sexo == "H")
data1 = data.frame(dataH$Estatura, dataH$Peso, dataM$Estatura, dataM$Peso)
```

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

```
# 1. Obtener la matriz de correlación de los datos
cor_matrix <- cor(data1)
print(cor_matrix)
```

```
##                 dataH.Estatura  dataH.Peso dataM.Estatura dataM.Peso
## dataH.Estatura   1.0000000000 0.846834792   0.0005540612 0.04724872
## dataH.Peso       0.8468347920 1.000000000   0.0035132246 0.02154907
## dataM.Estatura   0.0005540612 0.003513225   1.0000000000 0.52449621
## dataM.Peso       0.0472487231 0.021549075   0.5244962115 1.00000000
```

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

```
# Regresión para hombres
modelo_hombres <- lm(dataH.Peso ~ dataH.Estatura, data = data1)
modelo_hombres
```

```
##
## Call:
## lm(formula = dataH.Peso ~ dataH.Estatura, data = data1)
##
## Coefficients:
##    (Intercept)  dataH.Estatura
##         -83.68           94.66
```

```
summary(modelo_hombres)
```

```
##
## Call:
## lm(formula = dataH.Peso ~ dataH.Estatura, data = data1)
##
## 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 ***
## dataH.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
```

```r
# Regresión para mujeres
modelo_mujeres <- lm(dataM.Peso ~ dataM.Estatura, data = data1)
modelo_mujeres
```

```
##
## Call:
## lm(formula = dataM.Peso ~ dataM.Estatura, data = data1)
##
## Coefficients:
##    (Intercept)  dataM.Estatura
##         -72.56           81.15
```

```r
summary(modelo_mujeres)
```

```
##
## Call:
## lm(formula = dataM.Peso ~ dataM.Estatura, data = data1)
##
## 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 ***
## dataM.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
```

```r
# Modelo sin interaccion
modelo_ambos <- lm(Peso ~ Estatura + Sexo, data = data)
modelo_ambos
```

```
##
## Call:
```

```
## lm(formula = Peso ~ Estatura + Sexo, data = data)
##
## Coefficients:
## (Intercept)     Estatura        SexoM
##      -74.75        89.26       -10.56
```

```
summary(modelo_ambos)
```

```
##
## Call:
## lm(formula = Peso ~ Estatura + Sexo, data = data)
##
## 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
```

Pruebas Hipotesis (2. No te olvides de incluir las hipótesis en la pruebas de hipótesis que realices).

H_0:El modelo sigue una distribución normal

H_1:El modelo no sigue una distribución normal

$\alpha = 0.03$

Por separado

Hombres

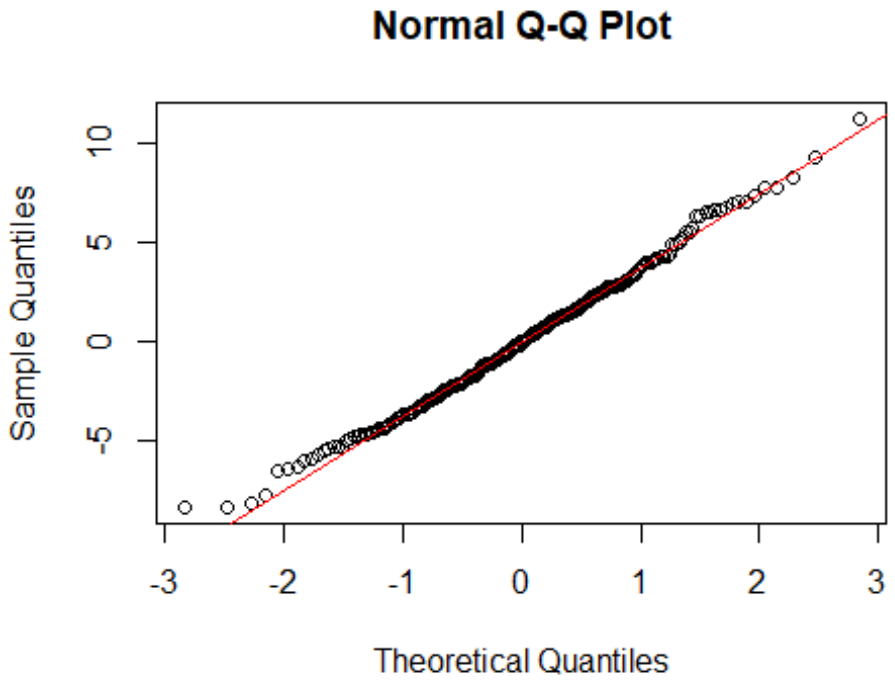
1.1 Normalidad de los residuos

```
library(nortest)
ad.test(modelo_hombres$residuals)  # Anderson-Darling test para normalidad
```

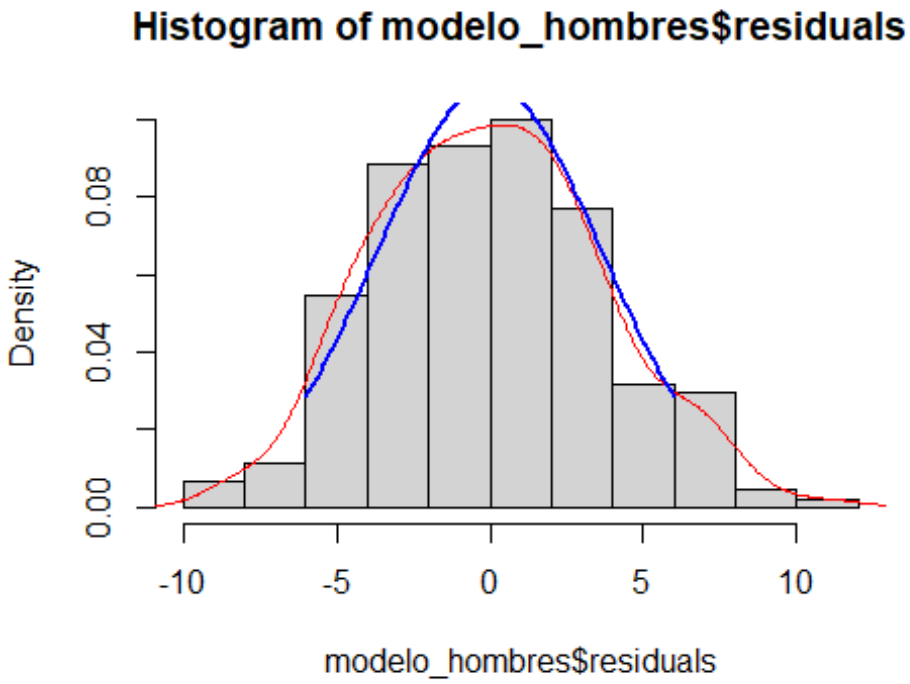
```
##
##  Anderson-Darling normality test
##
## data:  modelo_hombres$residuals
## A = 0.3009, p-value = 0.5771
```

p vlaue > 0.03 por lo tanto pasa la prueba de normalidad y no rechazamos la hipotesis nula

```r
# Graficar la normalidad
qqnorm(modelo_hombres$residuals)
qqline(modelo_hombres$residuals, col = "red")
```




```r
hist(modelo_hombres$residuals, freq = FALSE)
lines(density(modelo_hombres$residuals), col = "red")
curve(dnorm(x, mean = mean(modelo_hombres$residuals), sd =
sd(modelo_hombres$residuals)),
      from = -6, to = 6, add = TRUE, col = "blue", lwd = 2)
```



1.2 Verificación de media cero

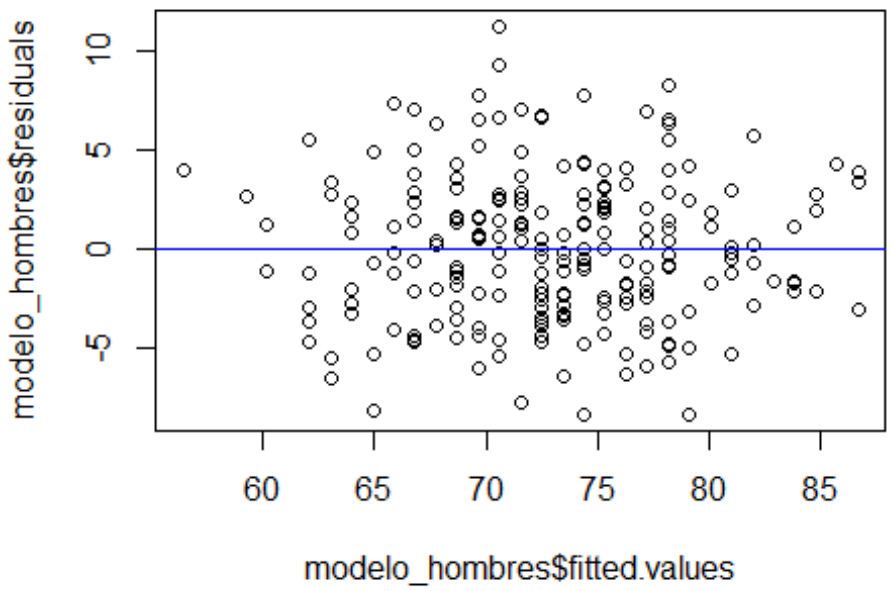
```
t.test(modelo_hombres$residuals)  # Prueba t para ver si la media de los residuos es 0
```

```
##
##  One Sample t-test
##
## data:  modelo_hombres$residuals
## t = 4.5495e-16, df = 219, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.4876507  0.4876507
## sample estimates:
##    mean of x
## 1.125698e-16
```

p vlaue > 0.03 por lo tanto pasa la prueba de normalidad y no rechazamos la hipotesis nula

1.3 Homocedasticidad e independencia

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




```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

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

```
dwtest(modelo_hombres)  # Durbin-Watson test para autocorrelación
```

```
##
##  Durbin-Watson test
##
## data:  modelo_hombres
## DW = 2.0556, p-value = 0.6599
## alternative hypothesis: true autocorrelation is greater than 0
```

```
bptest(modelo_hombres)  # Breusch-Pagan test para heterocedasticidad
```

```
##
##  studentized Breusch-Pagan test
##
## data:  modelo_hombres
## BP = 0.93324, df = 1, p-value = 0.334
```

p vlaue > 0.03 por lo tanto pasa la prueba de normalidad y no rechazamos la hipotesis nula.

3 Interpreta en el contexto del problema cada uno de los análisis que hiciste.

*Normalidad de los residuos: Prueba Anderson-Darling: No se rechazó la hipótesis nula de normalidad (valor p > 0.03), lo que sugiere que los residuos del modelo pueden considerarse normalmente distribuidos. Esto apoya la validez del modelo en términos de suposiciones de regresión lineal.

*Verificación de media cero: La media de los residuos es aproximadamente cero, como lo indicó el t-test (valor p = 1). Esto es importante porque confirma que no hay sesgo en los residuos del modelo, lo que significa que el modelo no tiene errores sistemáticos en la predicción del peso.

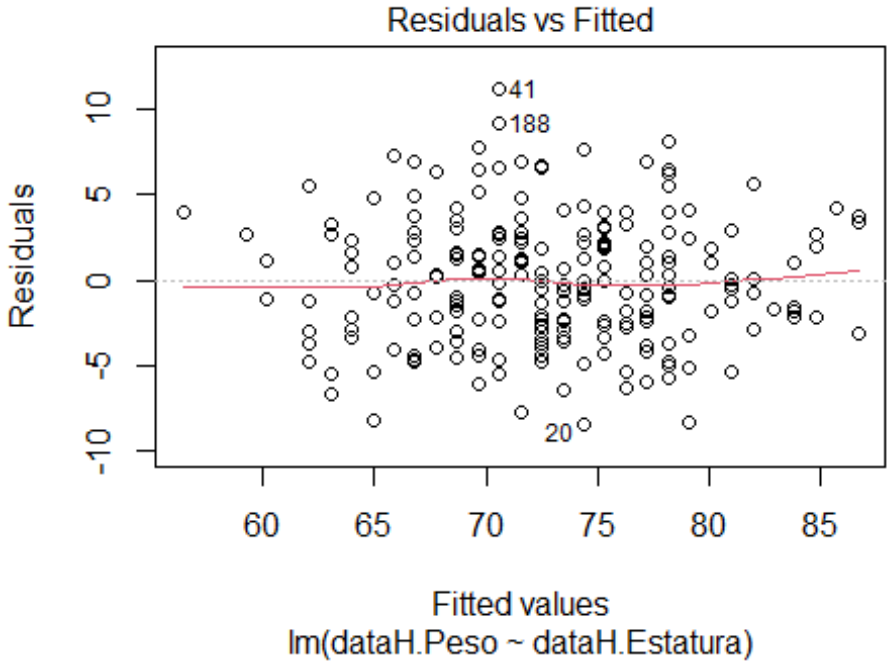
**Homocedasticidad: *Independencia de los errores: El Durbin-Watson test no rechazó la hipótesis nula de independencia de los errores (valor p > 0.03), lo que indica que no hay autocorrelación en los residuos. Esto es clave para la validez del modelo, ya que asegura que los errores no están correlacionados entre sí.

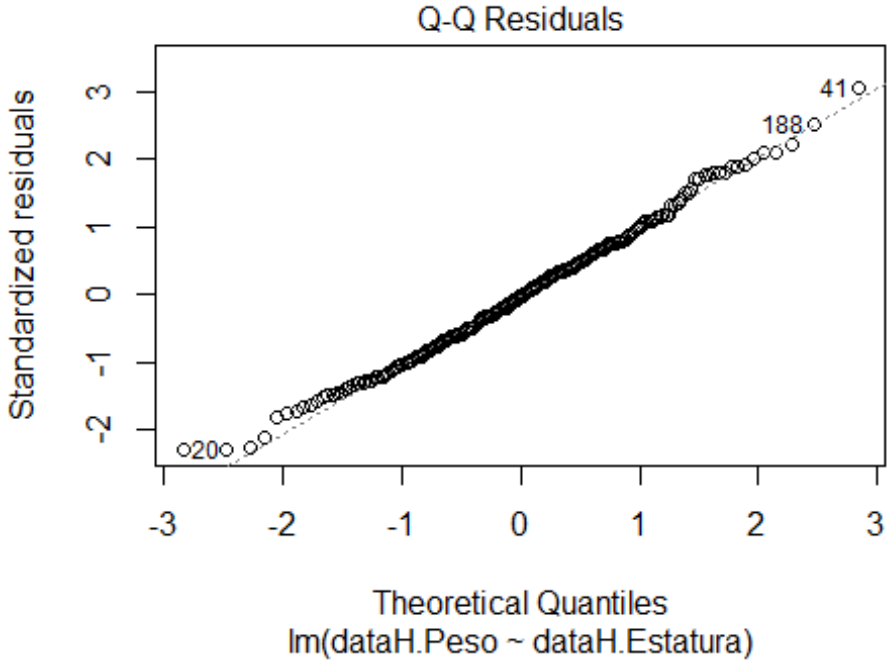
*Heterocedasticidad: La prueba de Breusch-Pagan no rechazó la hipótesis de homocedasticidad (valor p > 0.03), lo que indica que existe no existe heterocedasticidad en los residuos. Esto indicaria que los errores aparentan que tienen una varianza constante.

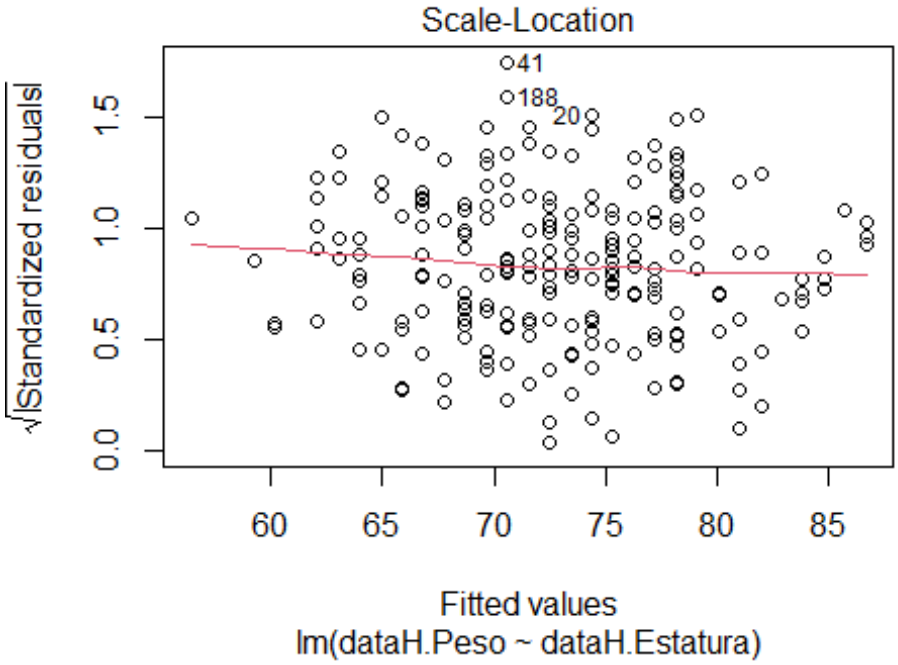
Es un muy buen modelo debido a que pasa todas las pruebas de forma adecuada, y sera uno de los candidatos a continuar con la parte de los intervalos

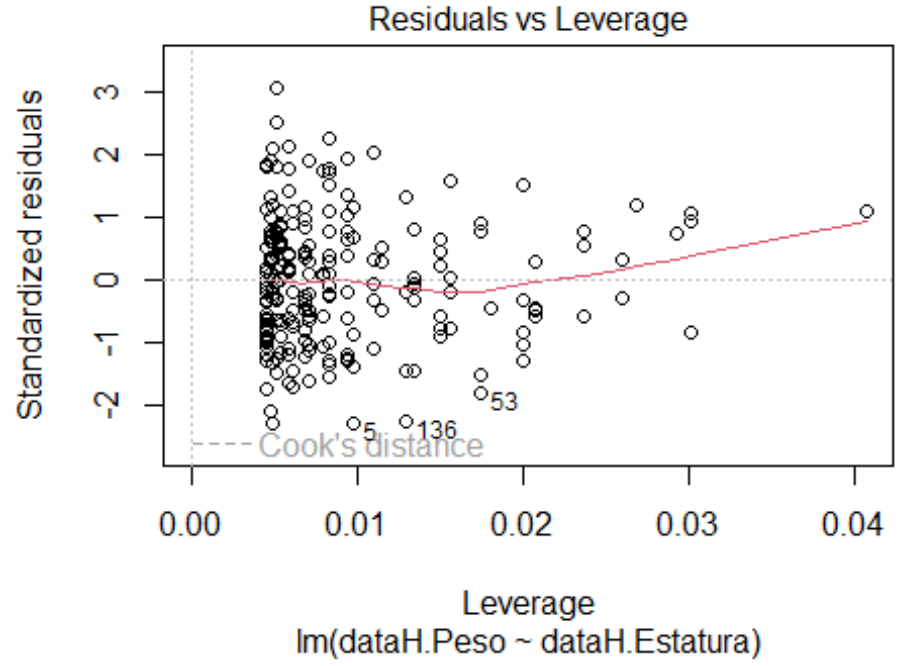
4 Utiliza el comando: plot(modelo). Observa las gráficas obtenidas y contesta:

```
plot(modelo_hombres)
```


Residuals vs Fitted
Residuals
Fitted values
lm(dataH.Peso ~ dataH.Estatura)
41
188
20
-10
-5
0
5
10
60
65
70
75
80
85


Q-Q Residuals
Standardized residuals
Theoretical Quantiles
lm(dataH.Peso ~ dataH.Estatura)
41
188
20
-2
-1
0
1
2
3





4.1 ¿Cuáles son las diferencias y similitudes de estos gráficos con respecto a los que ya habías analizado?

En similitudes debo decir que en cuestion del grafico Q-Q Residuals ya es exactamente igual a la grafica Q-Q plot que habiamos realizado anteriormente y de la misma manera

Residuals vs fitted es excatamente lo mismo que hicimos en nuestras pruebas de homocedaticidad que toman en cuenta la raiz cuadrada de los residuals y los fitted values.

Entre las diferencias podemos observar que el grafico de Scale location y Residuals vs leverage nos dan una perspectiva nueva a traves de los residuos de forma estandarizada o incluso la raiz cuadrada de los mismos comparados con leverage o con los fitted values, que podrian ser formas neuvas de visualizar los errores o residuos existentes sin verse afectado por tomar en cuenta escalas diferentes

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

Realmente solo reafirman las conclusiones que ya habia hecho con anterioridad para concluir con el modelo

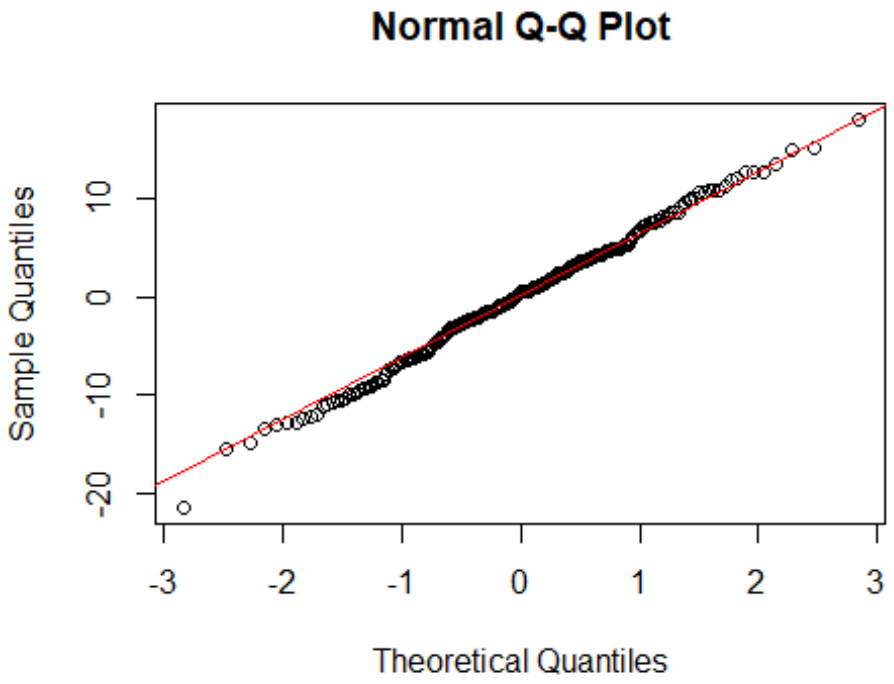
Mujeres

```
library(nortest)
ad.test(modelo_mujeres$residuals)  # Anderson-Darling test para normalidad
```

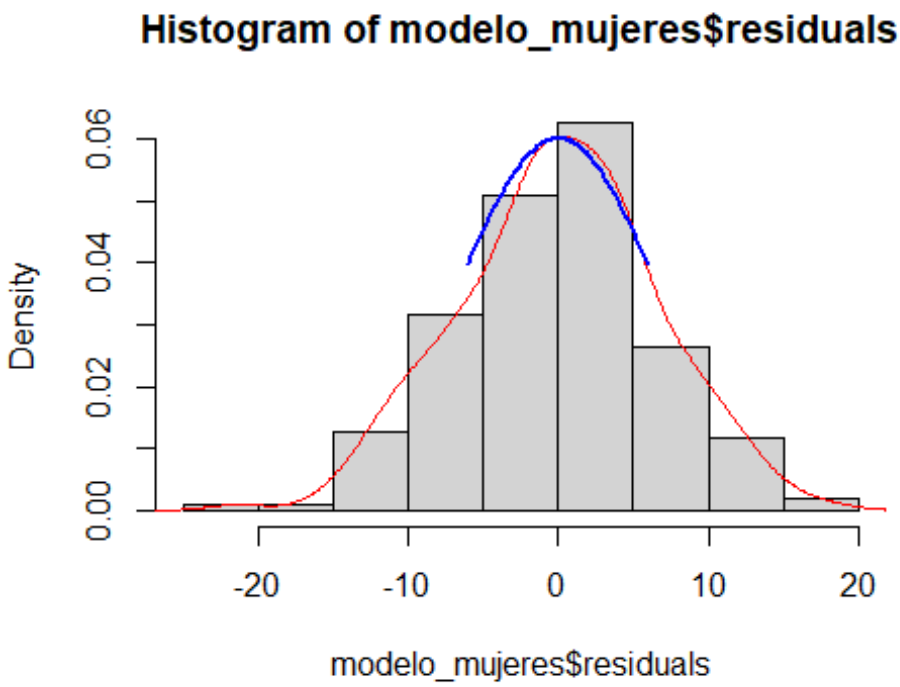
```
##
##  Anderson-Darling normality test
##
## data:  modelo_mujeres$residuals
## A = 0.24899, p-value = 0.7451
```

p vlaue > 0.03 por lo tanto pasa la prueba de normalidad y no rechazamos la hipotesis nula

```
# Graficar la normalidad
qqnorm(modelo_mujeres$residuals)
qqline(modelo_mujeres$residuals, col = "red")
```




```r
hist(modelo_mujeres$residuals, freq = FALSE)
lines(density(modelo_mujeres$residuals), col = "red")
curve(dnorm(x, mean = mean(modelo_mujeres$residuals), sd =
sd(modelo_mujeres$residuals)),
      from = -6, to = 6, add = TRUE, col = "blue", lwd = 2)
```



1.2 Verificación de media cero

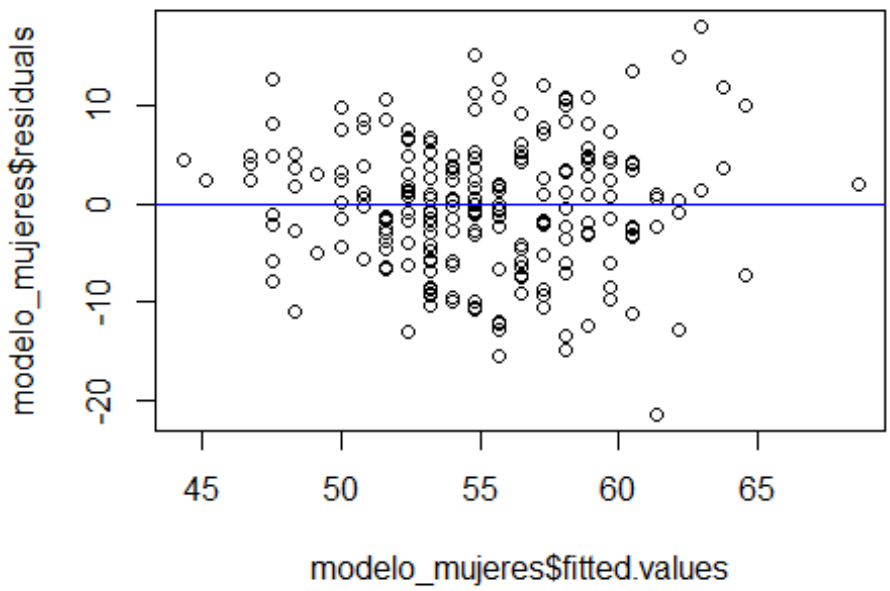
```
t.test(modelo_mujeres$residuals)  # Prueba t para ver si la media de los residuos es 0
```

```
##
##  One Sample t-test
##
## data:  modelo_mujeres$residuals
## t = -3.9979e-16, df = 219, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.881609  0.881609
## sample estimates:
##     mean of x
## -1.788342e-16
```

p vlaue > 0.03 por lo tanto pasa la prueba de normalidad y no rechazamos la hipotesis nula

1.3 Homocedasticidad e independencia

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




```
library(lmtest)
dwtest(modelo_mujeres)  # Durbin-Watson test para autocorrelación

##
##  Durbin-Watson test
##
## data:  modelo_mujeres
## DW = 1.8062, p-value = 0.07532
## alternative hypothesis: true autocorrelation is greater than 0

bptest(modelo_mujeres)  # Breusch-Pagan test para heterocedasticidad

##
##  studentized Breusch-Pagan test
##
## data:  modelo_mujeres
## BP = 8.4976, df = 1, p-value = 0.003556
```

p vlaue > 0.03 en la prueba de independencia por Drubin Watson por lo tanto pasa la prueba y no tenemos suficiente evidencia para rechazar la hipotesis nula

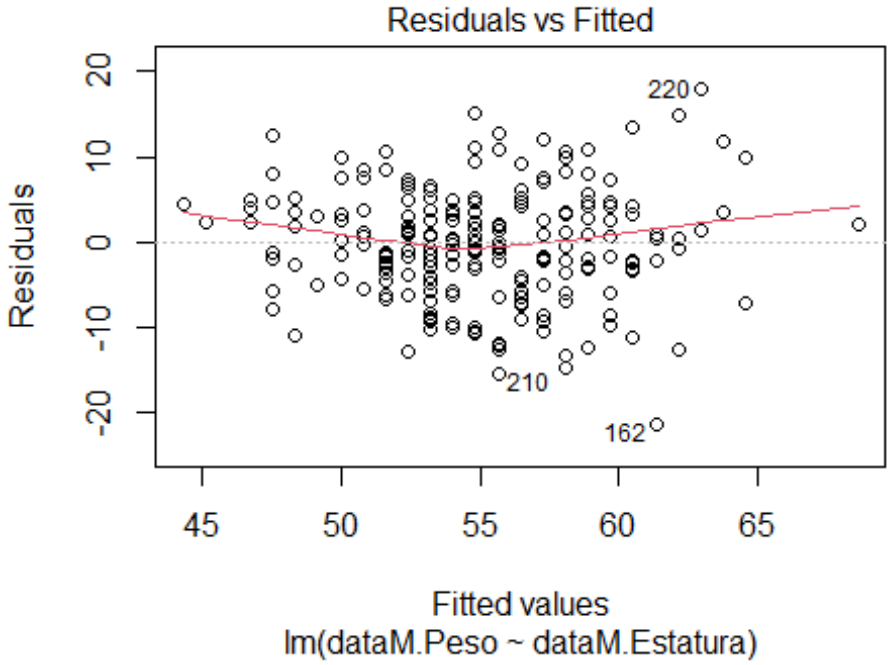
p value < 0.03 en homocedasticidad con Breusch-Pagan lo que señala rechazar la hipotesis nula

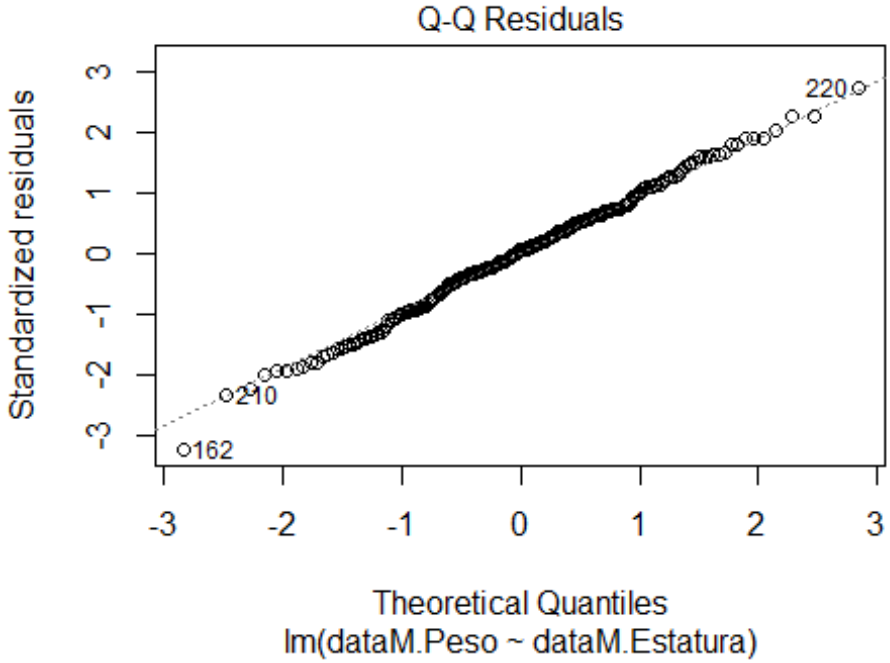
3 Interpreta en el contexto del problema cada uno de los análisis que hiciste.

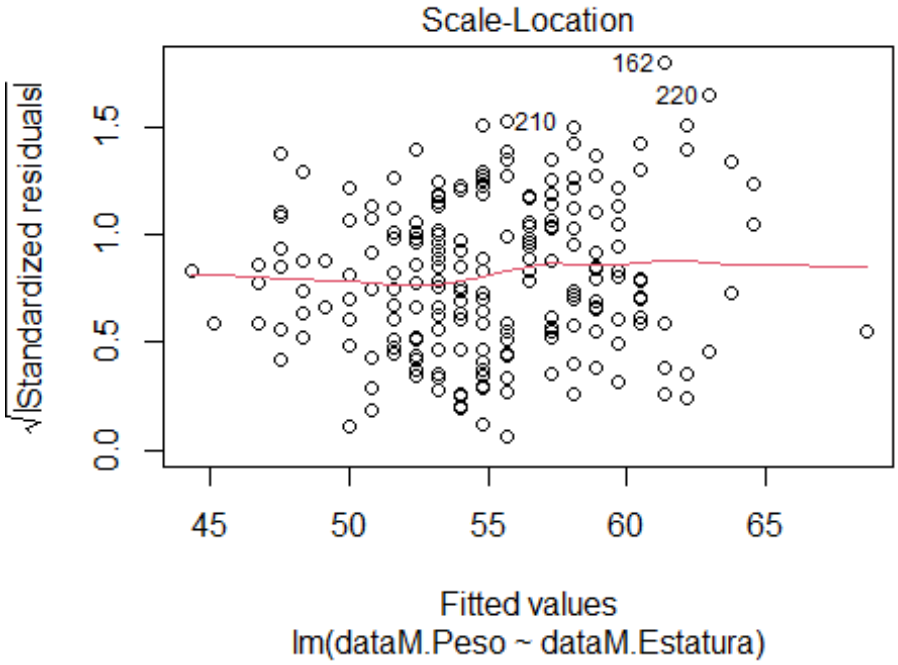
De forma similar y aunque pasan prueba de normalidad similar a los hombres en cuestion de conclusiones finales, si nos podemos a ver los valores o estadisicos de cada parte en la prueba de independencia vemos que no pasa tan facilmente la prueba como los hombres, debido a que aqui el p-value = 0.07532 < 0.6599 (hombres) sin mencionar que no cumple homocedasticidad en su totalidad debido a que p value < 0.03 y aunque pasa las pruebas de normalidad sobradamente, incluyendo que se ve bien el Q-Q plot asi como el histograma y dsitribucion de residuos, si recordamos actividades anteriores nuestro R^2 no fue de los mejores.

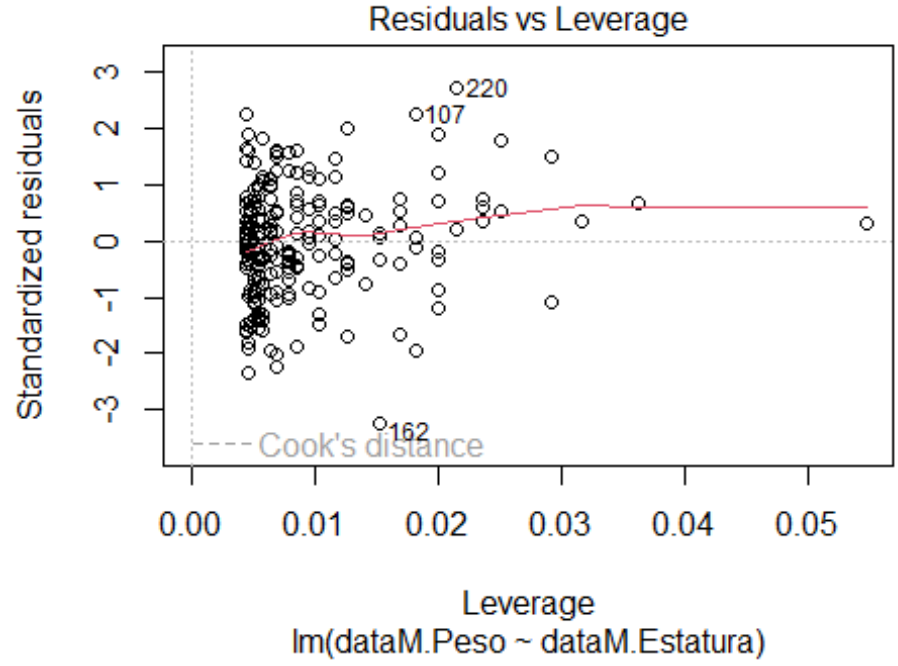
4 Utiliza el comando: plot(modelo). Observa las gráficas obtenidas y contesta:

```
plot(modelo_mujeres)
```


Residuals vs Fitted
Residuals
Fitted values
lm(dataM.Peso ~ dataM.Estatura)
220
210
162
20
10
0
-10
-20
45
50
55
60
65


Q-Q Residuals
Standardized residuals
Theoretical Quantiles
lm(dataM.Peso ~ dataM.Estatura)
220
210
162
3
2
1
0
-1
-2
-3





4.1 ¿Cuáles son las diferencias y similitudes de estos gráficos con respecto a los que ya habías analizado?

En similitudes debo decir que en cuestion del grafico Q-Q Residuals ya es exactamente igual a la grafica Q-Q plot que habiamos realizado anteriormente y de la misma manera

Residuals vs fitted es excatamente lo mismo que hicimos en nuestras pruebas de homocedaticidad que toman en cuenta la raiz cuadrada de los residuals y los fitted values.

Entre las diferencias podemos observar que el grafico de Scale location y Residuals vs leverage nos dan una perspectiva nueva a traves de los residuos de forma estandarizada o incluso la raiz cuadrada de los mismos comparados con leverage o con los fitted values, que podrian ser formas neuvas de visualizar los errores o residuos existentes sin verse afectado por tomar en cuenta escalas diferentes

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

Realmente solo reafirman las conclusiones que ya habia hecho con anterioridad para concluir con el modelo

Ambos

1.1 Normalidad de los residuos

Anderson Darling (n>50)

```
library(nortest)

ad.test(modelo_ambos$residuals)

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

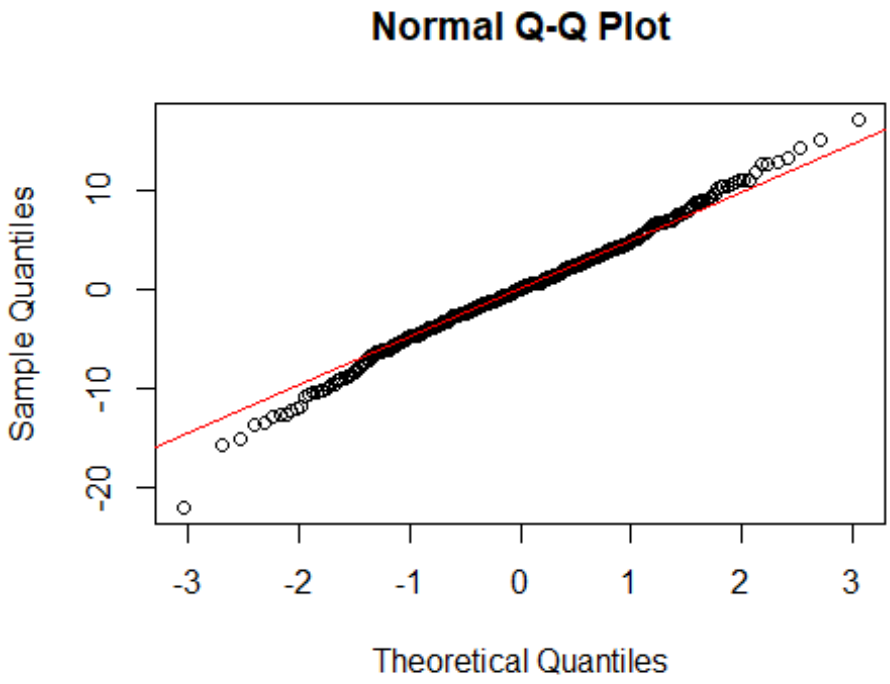
cuando p-value > 0.03 no tenemos suficiente evidencia para descartar normalidad

Kolmogorov_Simirnov

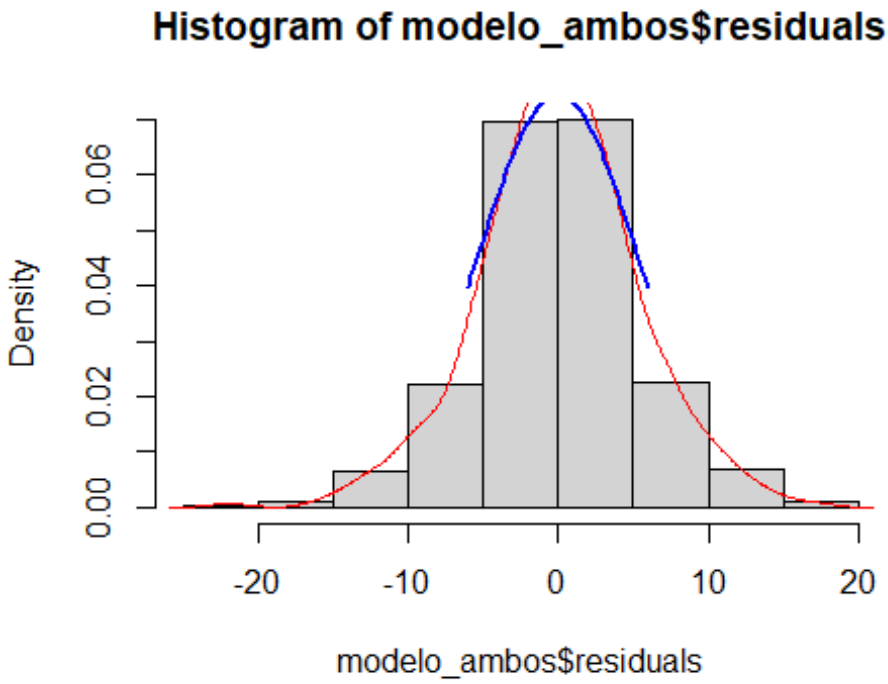
```
lillie.test(modelo_ambos$residuals)

##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  modelo_ambos$residuals
## D = 0.03675, p-value = 0.1583
```

cuando p-value > 0.03

```
# Graficar la normalidad
qqnorm(modelo_ambos$residuals)
qqline(modelo_ambos$residuals, col = "red")
```




```r
hist(modelo_ambos$residuals, freq = FALSE)
lines(density(modelo_ambos$residuals), col = "red")
curve(dnorm(x, mean = mean(modelo_ambos$residuals), sd =
sd(modelo_ambos$residuals)),
      from = -6, to = 6, add = TRUE, col = "blue", lwd = 2)
```



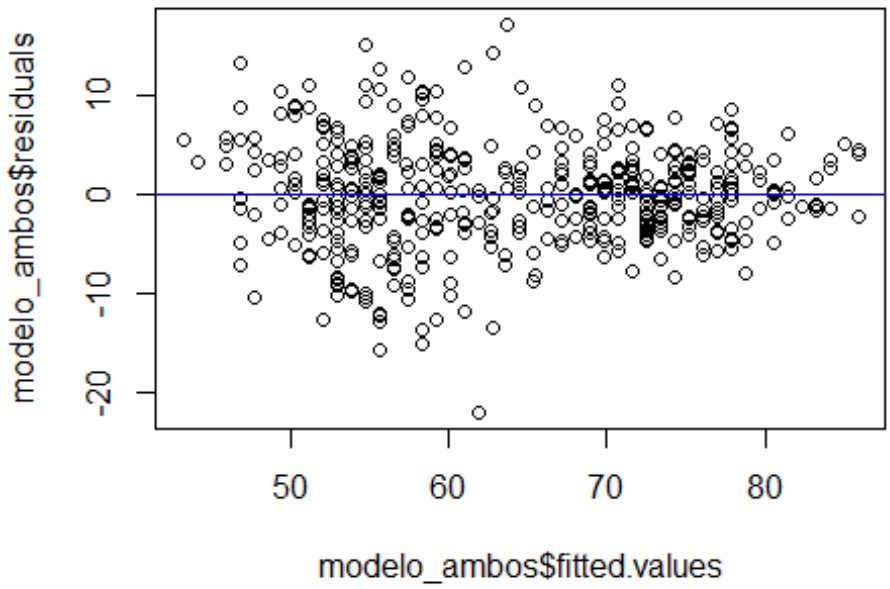
1.2 Verificación de media cero

```
t.test(modelo_ambos$residuals)  # Prueba t para ver si la media de los
residuos es 0

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

Podemos observar que pasa la prueba de t student con un p value > 0.03

1.3 Homocedasticidad e independencia

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




```
library(lmtest)
```

Independencia de las variables

```
dwtest(modelo_ambos)  # Durbin-Watson test para autocorrelación

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

p vlaue > 0.03 en la prueba de independencia por Drubin Watson por lo tanto pasa la prueba y no tenemos suficiente evidencia para rechazar la hipotesis nula

Homocedasticidad

```
bptest(modelo_ambos)  # Breusch-Pagan test para heterocedasticidad

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

p value < 0.03 en homocedasticidad con Breusch-Pagan lo que señala rechazar la hipotesis nula

3 Interpreta en el contexto del problema cada uno de los análisis que hiciste.

*Normalidad de los residuos: Prueba Anderson-Darling: No se rechazó la hipótesis nula de normalidad (valor p > 0.03), lo que sugiere que los residuos del modelo pueden considerarse normalmente distribuidos. Esto apoya la validez del modelo en términos de suposiciones de regresión lineal.

*Verificación de media cero: La media de los residuos es aproximadamente cero, como lo indicó el t-test (valor p = 1). Esto es importante porque confirma que no hay sesgo en los residuos del modelo, lo que significa que el modelo no tiene errores sistemáticos en la predicción del peso.

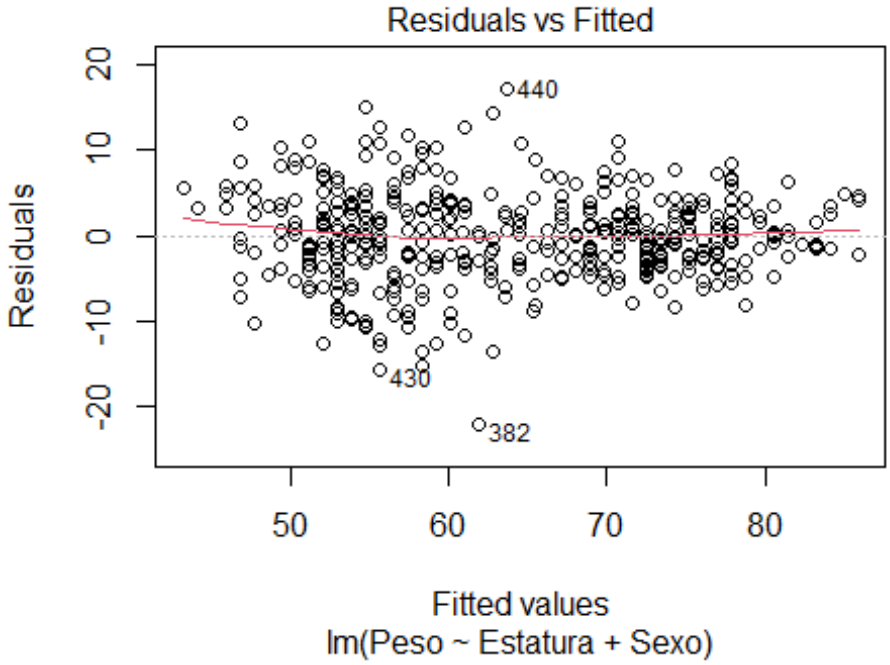
**Homocedasticidad: *Independencia de los errores: El Durbin-Watson test no rechazó la hipótesis nula de independencia de los errores (valor p > 0.03), aunque no de forma tan amplia como en el modelo de los hombres, sino mucho mas similar al caso de las mujeres lo que indica que no hay autocorrelación en los residuos. Esto es clave para la validez del modelo, ya que asegura que los errores no están correlacionados entre sí.

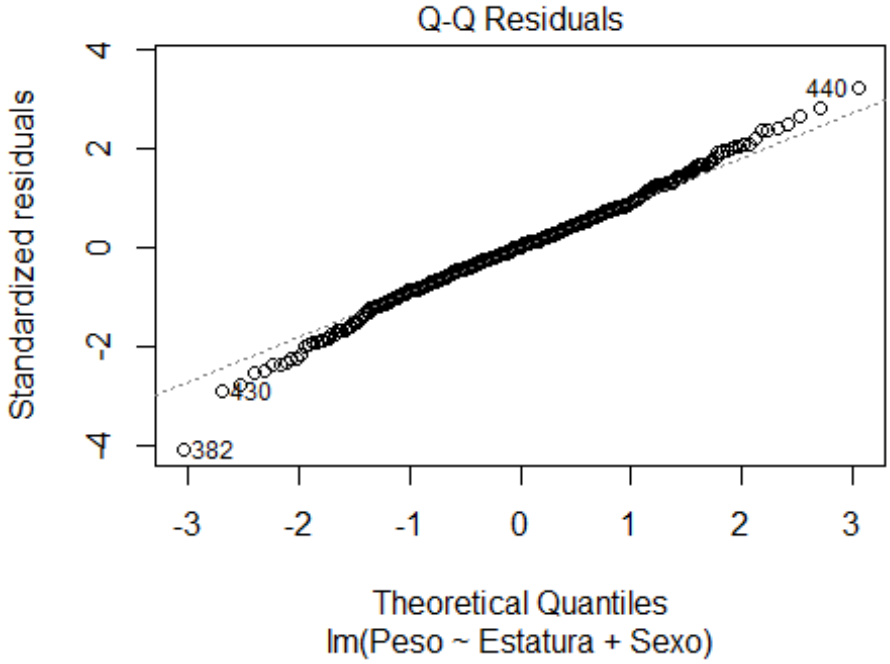
*Heterocedasticidad: La prueba de Breusch-Pagan rechazó la hipótesis de homocedasticidad (valor p < 0.03), lo que indica que existe heterocedasticidad en los residuos. Esto significa que los errores no tienen una varianza constante, lo que podría afectar la eficiencia de los coeficientes del modelo.

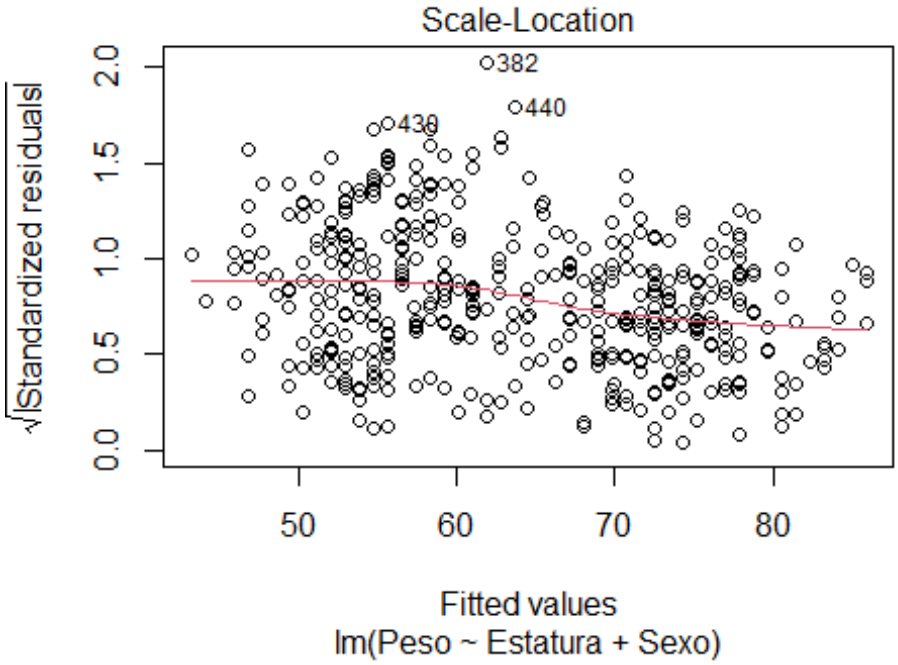
Aunque sigue siendo un muy buen modelo mis conclusiones son que no calificaria para continuar en la prueba por el efecto de la heterocedasticidad sobre la homocedasticidad del modelo

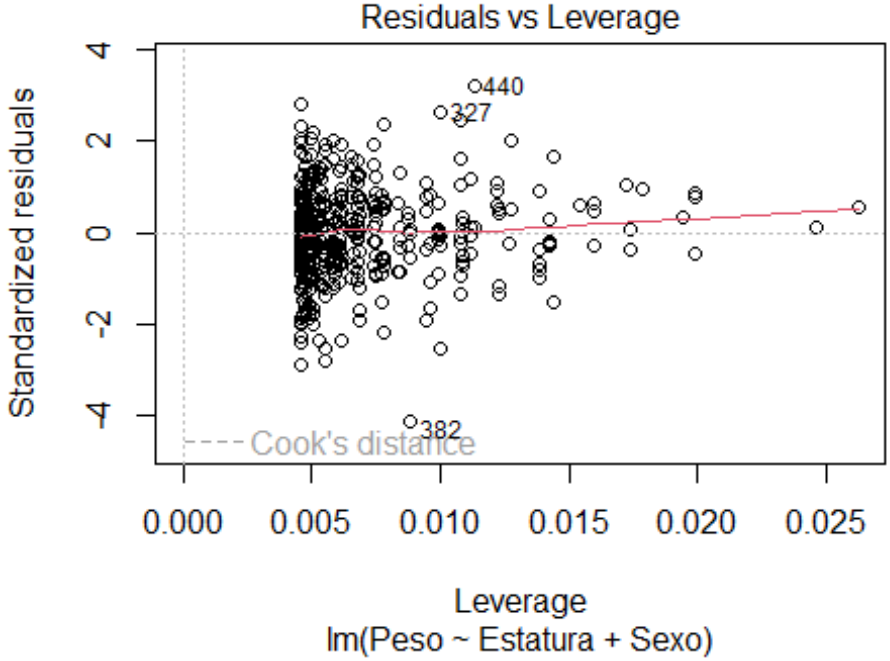
4 Utiliza el comando: plot(modelo). Observa las gráficas obtenidas y contesta:

```
plot(modelo_ambos)
```


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


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


Scale-Location
382
440
430
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
50
60
70
80
Fitted values
lm(Peso ~ Estatura + Sexo)


Residuals vs Leverage
440
327
382
Cook's distance
Standardized residuals
-4
-2
0
2
4
0.000
0.005
0.010
0.015
0.020
0.025
Leverage
lm(Peso ~ Estatura + Sexo)

4.1 ¿Cuáles son las diferencias y similitudes de estos gráficos con respecto a los que ya habías analizado?

En similitudes debo decir que en cuestion del grafico Q-Q Residuals ya es exactamente igual a la grafica Q-Q plot que habiamos realizado anteriormente y de la misma manera Residuals vs fitted es excatamente lo mismo que hicimos en nuestras pruebas de homocedaticidad que toman en cuenta la raiz cuadrada de los residuals y los fitted values.

Entre las diferencias podemos observar que el grafico de Scale location y Residuals vs leverage nos dan una perspectiva nueva a traves de los residuos de forma estandarizada o incluso la raiz cuadrada de los mismos comparados con leverage o con los fitted values, que podrian ser formas neuvas de visualizar los errores o residuos existentes sin verse afectado por tomar en cuenta escalas diferentes

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

Reafirman las conclusiones que fui formando en todo mi analisis

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

Sere breve ya que las conclusiones especificas estan en cada punto de cada modelo, pero en mi opinion, debido a que el modelo de los hombres es el unico que cumple todos los puntos de normalidad, nos iremos por la desicion de la division por separado de sexo, debido a que si escogemos el modelo de hombres tendremos que asumir la division por sexo con mujeres pero estudiar la poblacion en su totalidad.

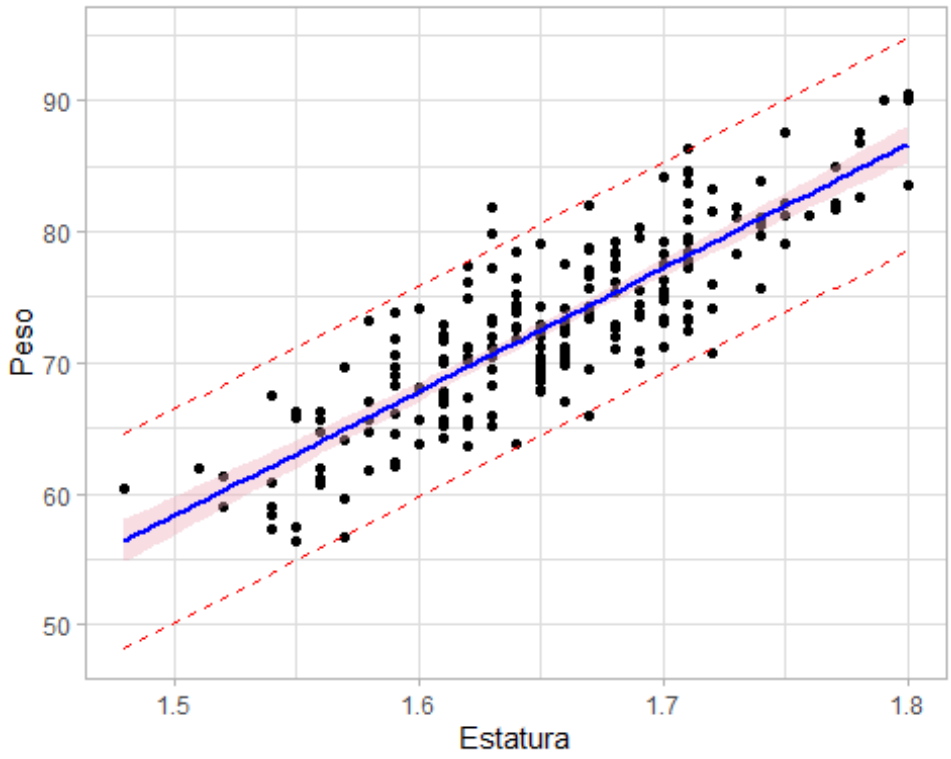
Intervalos de confianza

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.

Hombres

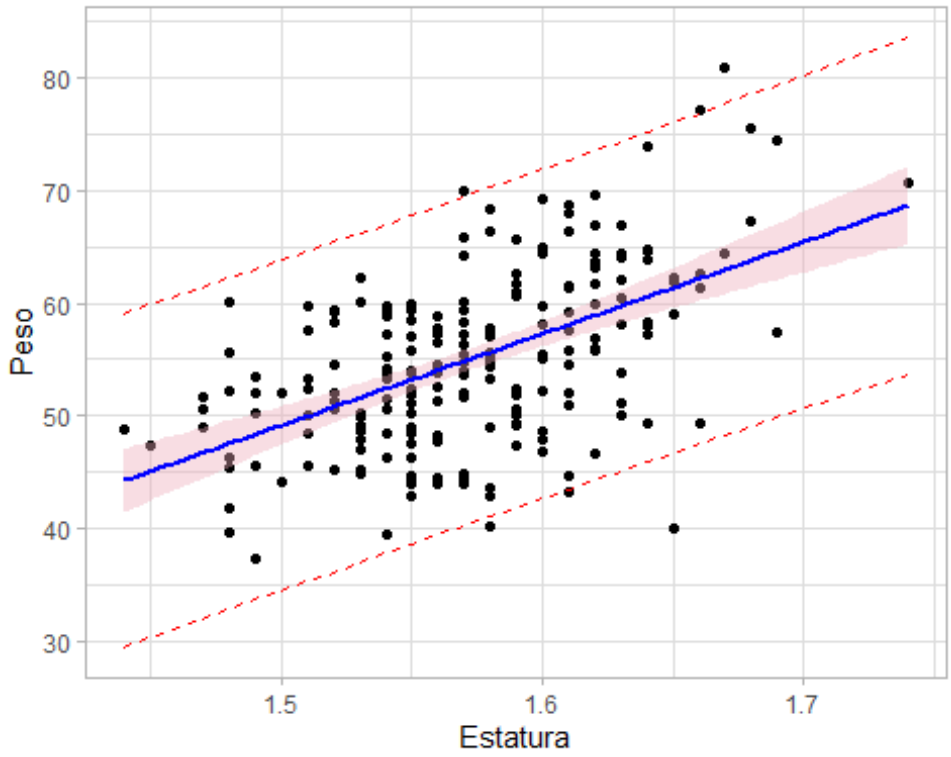
```
Ip=predict(object=modelo_hombres,interval="prediction",level=0.97)

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

datos=cbind(dataH,Ip)
library(ggplot2)
ggplot(datos,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")+
theme_light()
```



Mujeres

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

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

datos=cbind(dataM,Ip)
library(ggplot2)
ggplot(datos,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")+
theme_light()
```



Interpreta y comenta los resultados obtenidos

Mi conclusión sobre el modelo para hombres y mujeres:

Ajuste lineal: El ajuste lineal que veo (representado por la línea azul) parece capturar bien la tendencia central de los datos. A medida que la estatura aumenta, el peso también lo hace de manera lineal, lo que coincide con el coeficiente positivo de estatura en el modelo para hombres y mujeres que ajusté.

Intervalos de confianza: El área sombreada alrededor de la línea de regresión representa los intervalos de confianza para las predicciones del modelo, y las líneas rojas marcan los intervalos de predicción. Esto me indica que el modelo tiene un buen grado de confianza al predecir el peso para diferentes estaturas, entre 1.5 y 1.8 metros y 1.5 y 1.75 aprox en mujeres.

En el caso de los hombres Los puntos observados (en negro) es decir los datos por los cuales evaluamos los residuos del modelo, no sobrepasan la linea roja en el intervalo de prediccion a excepcion de algunos pocos pero están distribuidos de manera relativamente homogénea alrededor de la línea de regresión, no en su totalidad y obviamente algo mucho mas alejado de la linea sombreada pero recordemos que el modelo de hombres fue el unico y apenas paso la prueba de homocedasticidad, aun asi, me sugiere que no hay un patrón claro de heterocedasticidad, lo cual es positivo para la robustez del modelo. Esto se ve un poco mas afectado en el tema de mujeres debido a que recordemos que no paso la prueba

Residuos dentro de los intervalos de predicción: La mayoría de los puntos se encuentran dentro del intervalo de predicción, lo que confirma que el modelo es capaz de predecir de

manera adecuada el peso dentro del rango esperado. Esto es un buen indicio de la calidad del ajuste.

Selección del modelo basado en las pruebas de normalidad: Aun considerando todo, Como mencioné anteriormente, seleccioné este modelo de hombres (y por default mujeres) porque fue el único que pasó todas las pruebas de normalidad. Al observar esta gráfica, puedo ver que el ajuste lineal es sólido en el modelo seleccionado , y que los residuos están bien distribuidos alrededor de la línea de regresión, sin señales de problemas mayores.