

Analisis de los errores

Catherine Rojas

2024-09-04

```
library(readr)

## Warning: package 'readr' was built under R version 4.3.3

data <- read_csv("Estatura-peso_HyM.csv")

## Rows: 440 Columns: 3
## — Column specification
##
## Delimiter: ","
## chr (1): Sexo
## dbl (2): Estatura, Peso
##
## i Use `spec()` to retrieve the full column specification for this
## data.
## i Specify the column types or set `show_col_types = FALSE` to quiet
## this message.
```

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: Se considera un alpha de 0.03 para todas las pruebas.

Modelo sin interacción

```
# Ajustar el modelo de regresión sin interacción
modelo <- lm(Peso ~ Estatura + Sexo, data)

summary(modelo)

##
## 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

# Obtener la ecuación de regresión para hombres
cat("Ecuación de regresión para el modelo general: Peso =",
coef(modelo)[1], "+", coef(modelo)[2], "* Estatura\n")

## Ecuación de regresión para el modelo general: Peso = -74.7546 +
89.26035 * Estatura
```

1. Normalidad de los residuos

H_0 : Los datos provienen de una población normal H_1 : Los datos no provienen de una población normal.

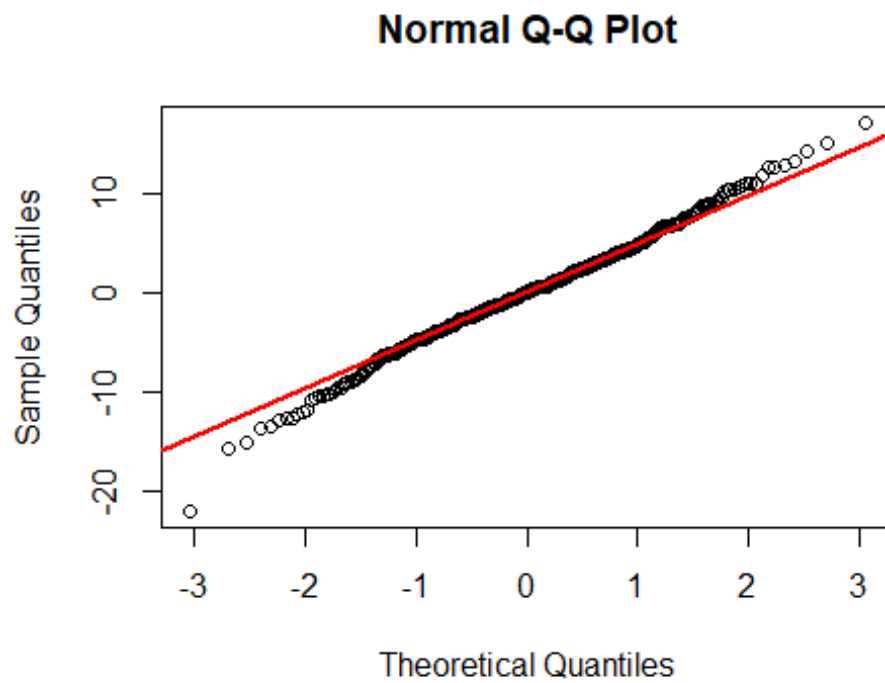
```
# 1. Normalidad de los residuos
# Test de normalidad (Shapiro-Wilk)
shapiro_test <- shapiro.test(residuals(modelo))
cat("Test de Shapiro-Wilk para normalidad de los residuos: \n")

## Test de Shapiro-Wilk para normalidad de los residuos:

print(shapiro_test)

##
## Shapiro-Wilk normality test
##
## data:  residuals(modelo)
## W = 0.99337, p-value = 0.0501

# Gráfico QQ para verificar normalidad visualmente
qqnorm(residuals(modelo))
qqline(residuals(modelo), col = "red", lwd = 2)
```

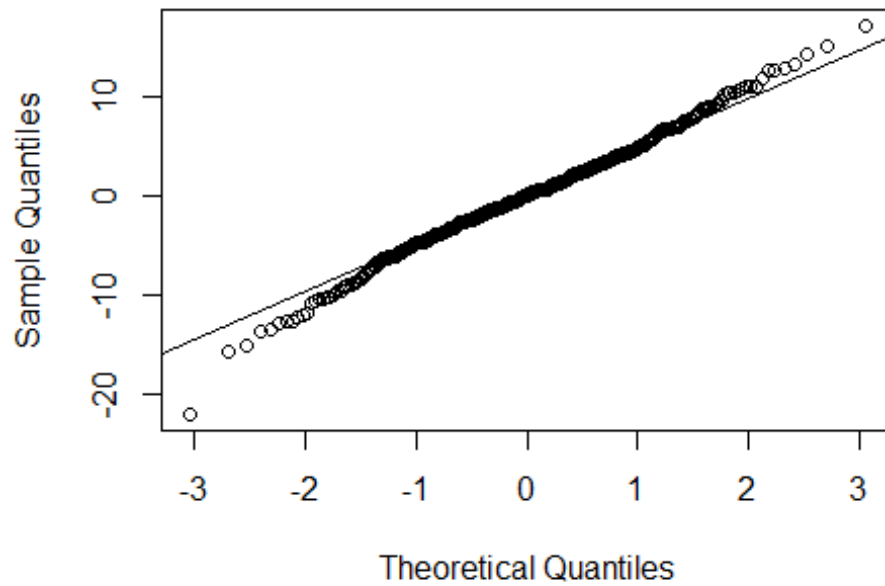


```
library(nortest)
ad.test(modelo$residuals)

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

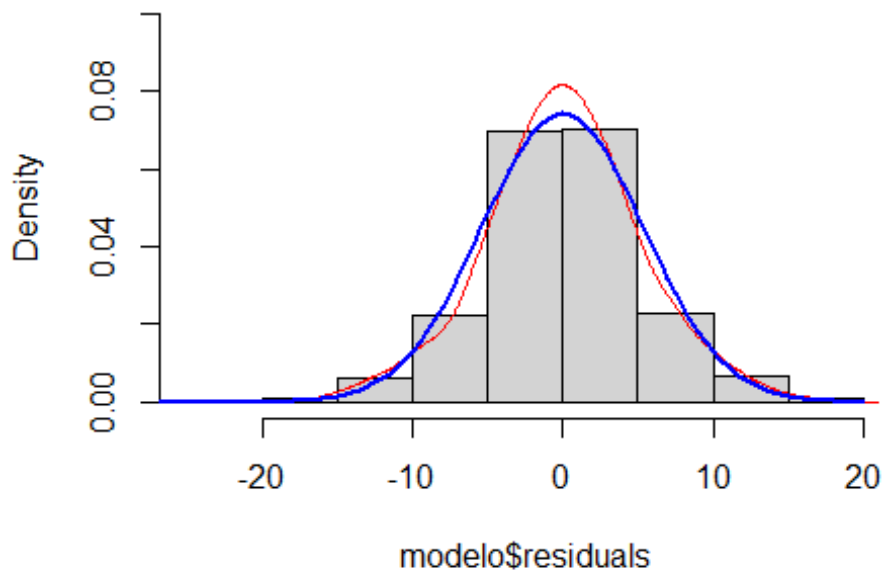
qqnorm(modelo$residuals)
qqline(modelo$residuals)
```

Normal Q-Q Plot



```
hist(modelo$residuals,freq=FALSE, ylim= c(0,0.1))  
lines(density(modelo$residual),col="red")  
curve(dnorm(x,mean=mean(modelo$residuals),sd=sd(modelo$residuals)),  
from=-30, to=20, add=TRUE, col="blue",lwd=2)
```

Histogram of modelo\$residuals



Los residuos se comportan razonablemente bien en términos de normalidad, aunque el test de Anderson-Darling señala una posible ligera desviación.

2. Verificación de media cero

$H_0: \mu = 0$ $H_1: \mu \neq 0$

```
# 2. Verificación de media cero de Los residuos
media_residuos <- mean(residuals(modelo))
cat("\nMedia de los residuos: ", media_residuos, "\n")

##
## Media de los residuos:  6.163788e-17

# Prueba t para verificar si La media de Los residuos es
# significativamente diferente de cero
t.test(modelo$residuals)

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

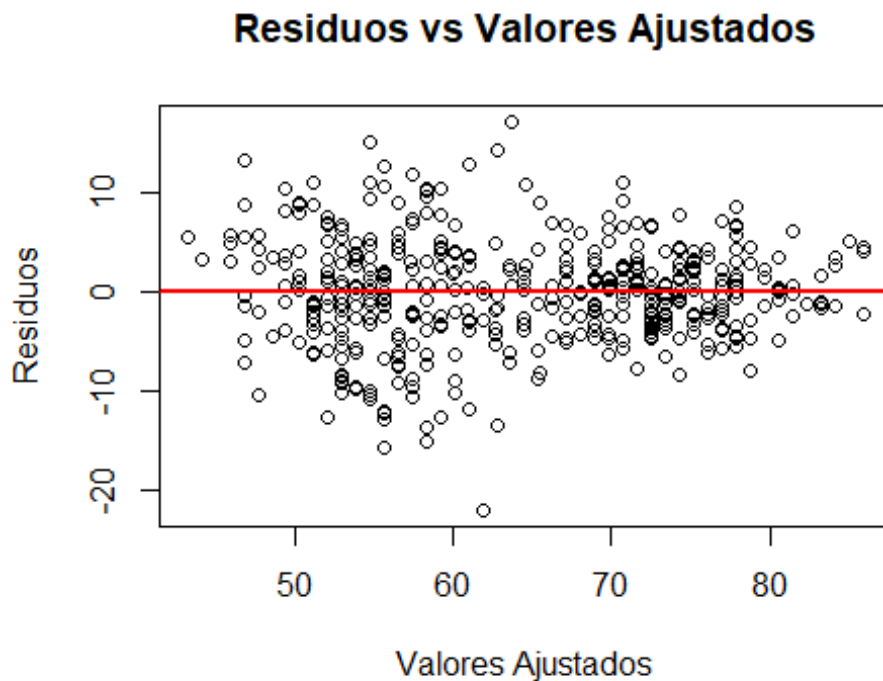
```
## mean of x  
## 6.163788e-17
```

El valor p es 1, lo que significa que no hay evidencia para rechazar la hipótesis de que la media de los residuos es cero. La media estimada es prácticamente cero ($6.16e-17$), lo cual es lo esperado en un modelo de regresión bien ajustado. La media de los residuos es cero, por lo que este supuesto se cumple.

3. Homocedasticidad

Prueba de Breusch-Pagan y White H_0 : La varianza de los errores es constante (homocedasticidad) H_1 : La varianza de los errores no es constante (heterocedasticidad)

```
# 3. Homocedasticidad  
# Gráfico de los residuos vs valores ajustados  
plot(fitted(modelo), residuals(modelo), main="Residuos vs Valores  
Ajustados", xlab="Valores Ajustados", ylab="Residuos")  
abline(h = 0, col = "red", lwd = 2)
```



```
library(lmtest)  
## Warning: package 'lmtest' was built under R version 4.3.3  
## Loading required package: zoo  
## Warning: package 'zoo' was built under R version 4.3.3
```

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

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

bptest(modelo)

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

gqtest(modelo)

##
## Goldfeld-Quandt test
##
## data:  modelo
## GQ = 3.2684, df1 = 217, df2 = 217, p-value < 2.2e-16
## alternative hypothesis: variance increases from segment 1 to 2
```

Breusch-Pagan test: El valor p es extremadamente pequeño ($3.413e-11$), lo que indica una fuerte evidencia de heterocedasticidad. Esto significa que la varianza de los residuos no es constante.

Goldfeld-Quandt test: El valor p también es muy bajo, lo que refuerza la conclusión de que la varianza de los residuos aumenta con los valores ajustados. Conclusión: No se cumple el supuesto de homocedasticidad, ya que los residuos muestran heterocedasticidad significativa.

4. Independencia

Pruebas de hipótesis: Test de Durbin-Watson y Prueba Breusch-Godfrey para independencia H_0 : Los errores no están autocorrelacionados. H_1 Los errores están autocorrelacionados.

```
library(lmtest)

dwtest(modelo)

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

bgtest(modelo)
```

```
##
## Breusch-Godfrey test for serial correlation of order up to 1
##
## data: modelo
## LM test = 1.3595, df = 1, p-value = 0.2436
```

Independencia de los residuos:

Durbin-Watson test: El valor DW de 1.8663 está cerca de 2, lo que sugiere que no hay una autocorrelación significativa de los residuos. El valor p es 0.07325, lo que indica que no hay evidencia suficiente para rechazar la hipótesis nula de no autocorrelación.

Breusch-Godfrey test: El valor p es 0.2436, lo que también sugiere que no hay evidencia de autocorrelación en los residuos. El supuesto de independencia de los residuos se cumple.

Modelos Hombre y Mujer

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

```
# Ajustar el modelo de regresión para hombres
modelo_hombres <- lm(Peso ~ Estatura, data = dataH)
```

```
# Ajustar el modelo de regresión para mujeres
modelo_mujeres <- lm(Peso ~ Estatura, data = dataM)
```

```
# Mostrar los resultados del modelo
summary(modelo_hombres)
```

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

summary(modelo_mujeres)

##
## Call:
## lm(formula = Peso ~ Estatura, data = dataM)
##
## 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

# Obtener la ecuación de regresión para hombres
cat("Ecuación de regresión para hombres modelo: Peso =",
coef(modelo_hombres)[1], "+", coef(modelo_hombres)[2], "* Estatura\n")

## Ecuación de regresión para hombres modelo: Peso = -83.68454 + 94.66024
* Estatura

# Obtener la ecuación de regresión para mujeres
cat("Ecuación de regresión para mujeres modelo : Peso =",
coef(modelo_mujeres)[1], "+", coef(modelo_mujeres)[2], "* Estatura\n")

## Ecuación de regresión para mujeres modelo : Peso = -72.56045 +
81.14911 * Estatura
```

1. Normalidad de los residuos

H_0 : Los datos provienen de una población normal H_1 : Los datos no provienen de una población normal.

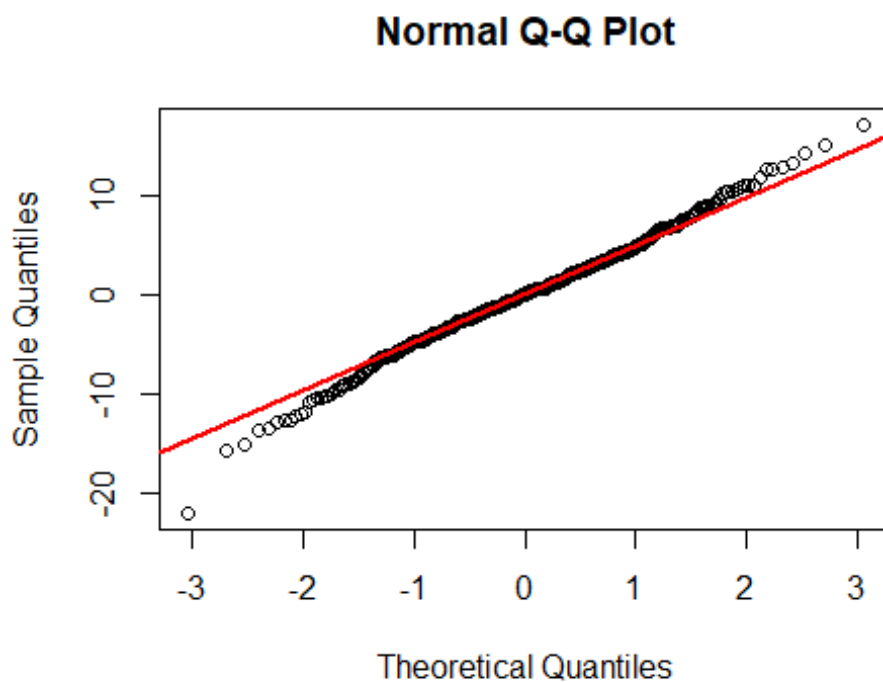
```
# 1. Normalidad de Los residuos
# Test de normalidad (Shapiro-Wilk)
shapiro_test <- shapiro.test(residuals(modelo))
cat("Test de Shapiro-Wilk para normalidad de los residuos: \n")

## Test de Shapiro-Wilk para normalidad de los residuos:

print(shapiro_test)
```

```
##
## Shapiro-Wilk normality test
##
## data: residuals(modelo)
## W = 0.99337, p-value = 0.0501

# Gráfico QQ para verificar normalidad visualmente
qqnorm(residuals(modelo))
qqline(residuals(modelo), col = "red", lwd = 2)
```

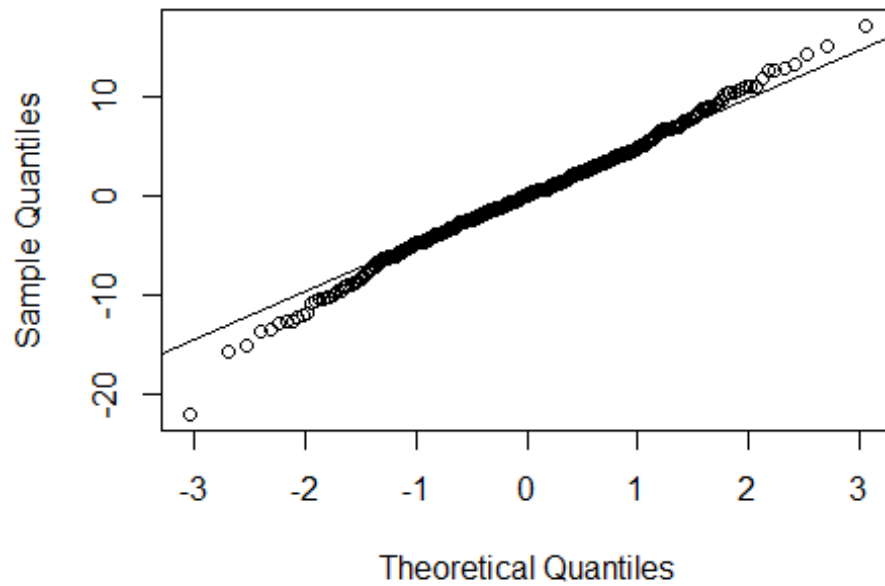


```
library(nortest)
ad.test(modelo$residuals)

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

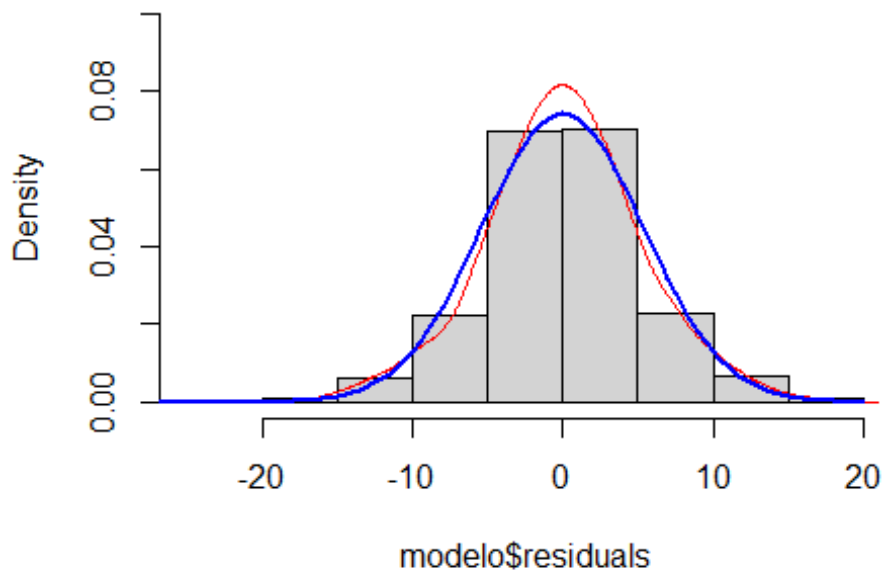
qqnorm(modelo$residuals)
qqline(modelo$residuals)
```

Normal Q-Q Plot



```
hist(modelo$residuals,freq=FALSE, ylim= c(0,0.1))  
lines(density(modelo$residual),col="red")  
curve(dnorm(x,mean=mean(modelo$residuals),sd=sd(modelo$residuals)),  
from=-30, to=20, add=TRUE, col="blue",lwd=2)
```

Histogram of modelo\$residuals



Los residuos se comportan razonablemente bien en términos de normalidad, aunque el test de Anderson-Darling señala una posible ligera desviación.

2. Verificación de media cero

$H_0: \mu = 0$ $H_1: \mu \neq 0$

```
# 2. Verificación de media cero de Los residuos
media_residuos <- mean(residuals(modelo))
cat("\nMedia de los residuos: ", media_residuos, "\n")

##
## Media de los residuos:  6.163788e-17

# Prueba t para verificar si La media de Los residuos es
# significativamente diferente de cero
t.test(modelo$residuals)

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

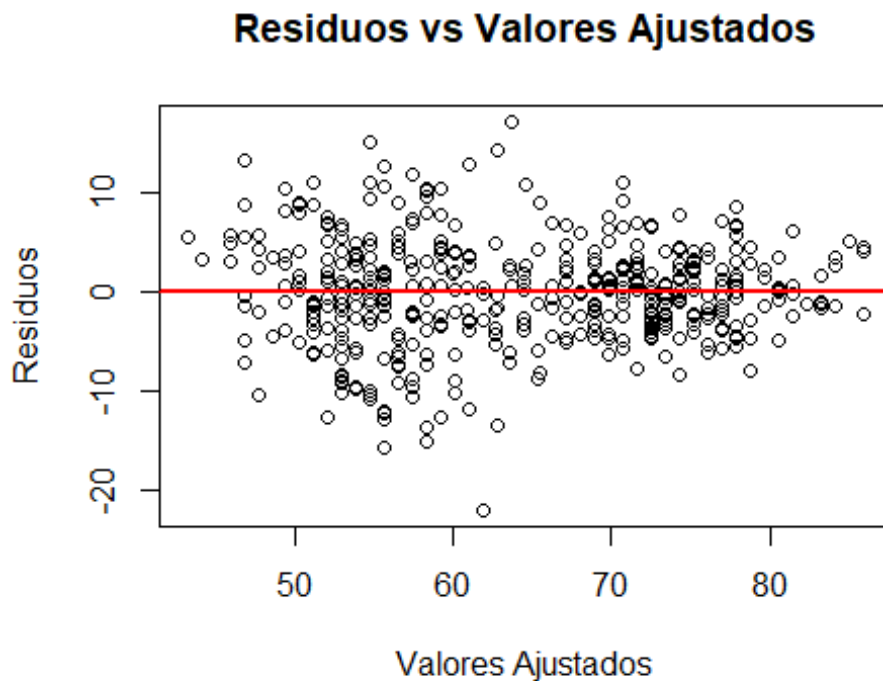
```
## mean of x  
## 6.163788e-17
```

El valor p es 1, lo que significa que no hay evidencia para rechazar la hipótesis de que la media de los residuos es cero. La media estimada es prácticamente cero (6.16e-17), lo cual es lo esperado en un modelo de regresión bien ajustado. La media de los residuos es cero, por lo que este supuesto se cumple.

3. Homocedasticidad

Prueba de Breusch-Pagan y White H_0 : La varianza de los errores es constante (homocedasticidad) H_1 : La varianza de los errores no es constante (heterocedasticidad)

```
# 3. Homocedasticidad  
# Gráfico de los residuos vs valores ajustados  
plot(fitted(modelo), residuals(modelo), main="Residuos vs Valores  
Ajustados", xlab="Valores Ajustados", ylab="Residuos")  
abline(h = 0, col = "red", lwd = 2)
```



```
library(lmtest)  
  
bptest(modelo)  
  
##  
## studentized Breusch-Pagan test  
##
```

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

gqtest(modelo)

##
## Goldfeld-Quandt test
##
## data: modelo
## GQ = 3.2684, df1 = 217, df2 = 217, p-value < 2.2e-16
## alternative hypothesis: variance increases from segment 1 to 2
```

Breusch-Pagan test: El valor p es extremadamente pequeño (3.413e-11), lo que indica una fuerte evidencia de heterocedasticidad. Esto significa que la varianza de los residuos no es constante.

Goldfeld-Quandt test: El valor p también es muy bajo, lo que refuerza la conclusión de que la varianza de los residuos aumenta con los valores ajustados. Conclusión: No se cumple el supuesto de homocedasticidad, ya que los residuos muestran heterocedasticidad significativa.

4. Independencia

Pruebas de hipótesis: Test de Durbin-Watson y Prueba Breusch-Godfrey para independencia H_0 : Los errores no están autocorrelacionados. H_1 Los errores están autocorrelacionados.

```
library(lmtest)

dwtest(modelo)

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

bgtest(modelo)

##
## Breusch-Godfrey test for serial correlation of order up to 1
##
## data: modelo
## LM test = 1.3595, df = 1, p-value = 0.2436
```

Independencia de los residuos:

Durbin-Watson test: El valor DW de 1.8663 está cerca de 2, lo que sugiere que no hay una autocorrelación significativa de los residuos. El valor p es 0.07325, lo que

indica que no hay evidencia suficiente para rechazar la hipótesis nula de no autocorrelación.

Breusch-Godfrey test: El valor p es 0.2436, lo que también sugiere que no hay evidencia de autocorrelación en los residuos. El supuesto de independencia de los residuos se cumple.

Modelos Hombre y Mujer

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

# Ajustar el modelo de regresión para hombres
modelo_hombres <- lm(Peso ~ Estatura, data = dataH)

# Ajustar el modelo de regresión para mujeres
modelo_mujeres <- lm(Peso ~ Estatura, data = dataM)

# Mostrar los resultados del modelo
summary(modelo_hombres)

##
## Call:
## lm(formula = Peso ~ Estatura, data = dataH)
##
## 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

summary(modelo_mujeres)

##
## Call:
## lm(formula = Peso ~ Estatura, data = dataM)
##
## 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

# Obtener la ecuación de regresión para hombres
cat("Ecuación de regresión para hombres modelo: Peso =",
coef(modelo_hombres)[1], "+", coef(modelo_hombres)[2], "* Estatura\n")

## Ecuación de regresión para hombres modelo: Peso = -83.68454 + 94.66024
* Estatura

# Obtener la ecuación de regresión para mujeres
cat("Ecuación de regresión para mujeres modelo : Peso =",
coef(modelo_mujeres)[1], "+", coef(modelo_mujeres)[2], "* Estatura\n")

## Ecuación de regresión para mujeres modelo : Peso = -72.56045 +
81.14911 * Estatura
```

1. Normalidad de los residuos

H_0 : Los datos provienen de una población normal H_1 : Los datos no provienen de una población normal.

```
# 1. Normalidad de Los residuos Hombres
# Test de normalidad (Shapiro-Wilk)
shapiro_test <- shapiro.test(residuals(modelo_hombres))
cat("Test de Shapiro-Wilk para normalidad de los residuos: \n")

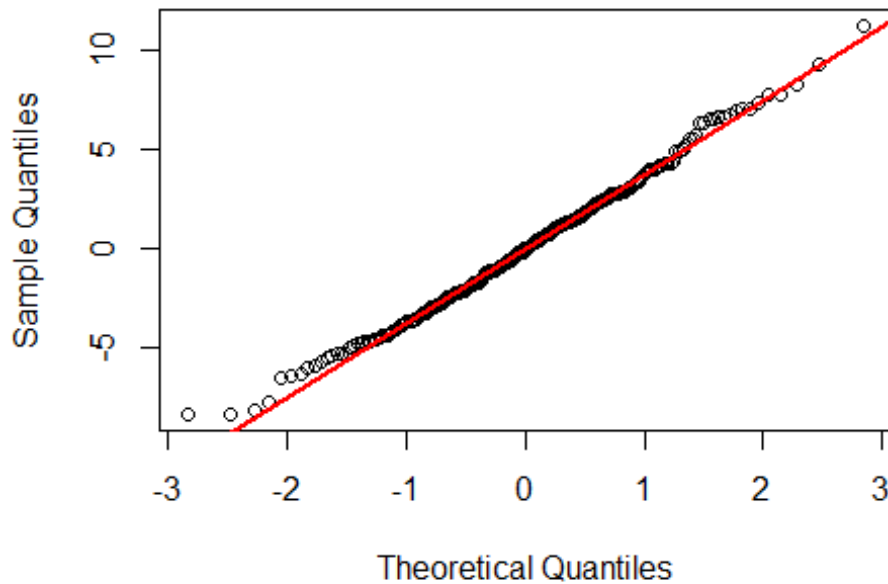
## Test de Shapiro-Wilk para normalidad de los residuos:

print(shapiro_test)

##
## Shapiro-Wilk normality test
##
## data:  residuals(modelo_hombres)
## W = 0.99356, p-value = 0.4597

# Gráfico QQ para verificar normalidad visualmente
qqnorm(residuals(modelo_hombres))
qqline(residuals(modelo_hombres), col = "red", lwd = 2)
```


Normal Q-Q Plot



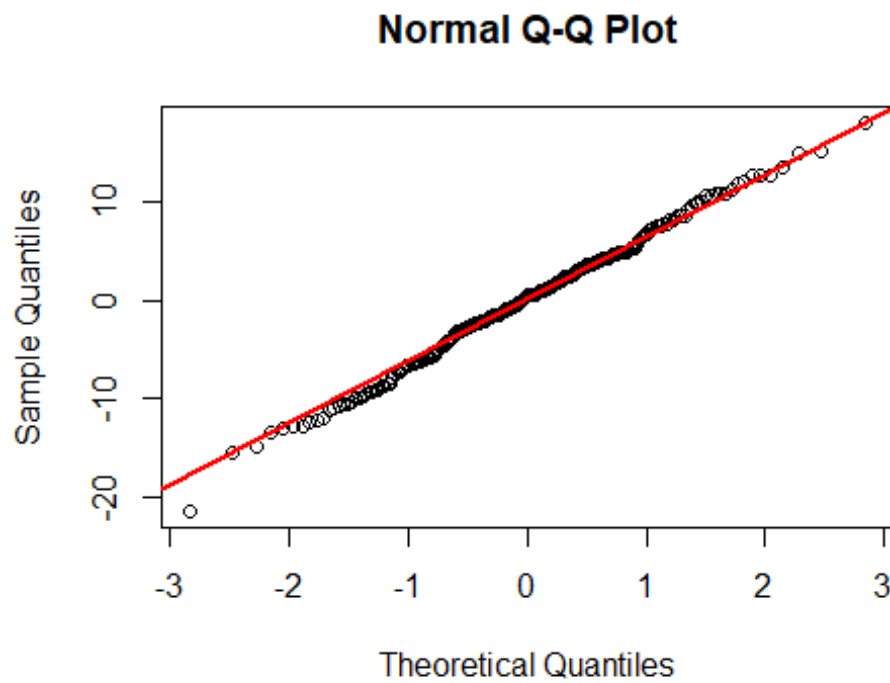
```
# 1. Normalidad de Los residuos Mujeres
# Test de normalidad (Shapiro-Wilk)
shapiro_test <- shapiro.test(residuals(modelo_mujeres))
cat("Test de Shapiro-Wilk para normalidad de los residuos: \n")

## Test de Shapiro-Wilk para normalidad de los residuos:

print(shapiro_test)

##
##  Shapiro-Wilk normality test
##
## data:  residuals(modelo_mujeres)
## W = 0.99659, p-value = 0.9144

# Gráfico QQ para verificar normalidad visualmente
qqnorm(residuals(modelo_mujeres))
qqline(residuals(modelo_mujeres), col = "red", lwd = 2)
```



```
library(nortest)

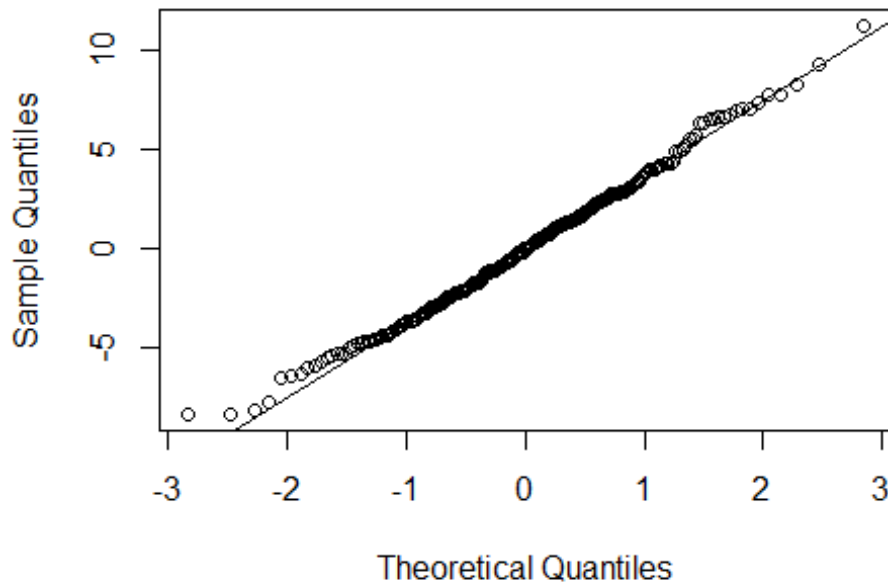
# Hombres

ad.test(modelo_hombres$residuals)

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

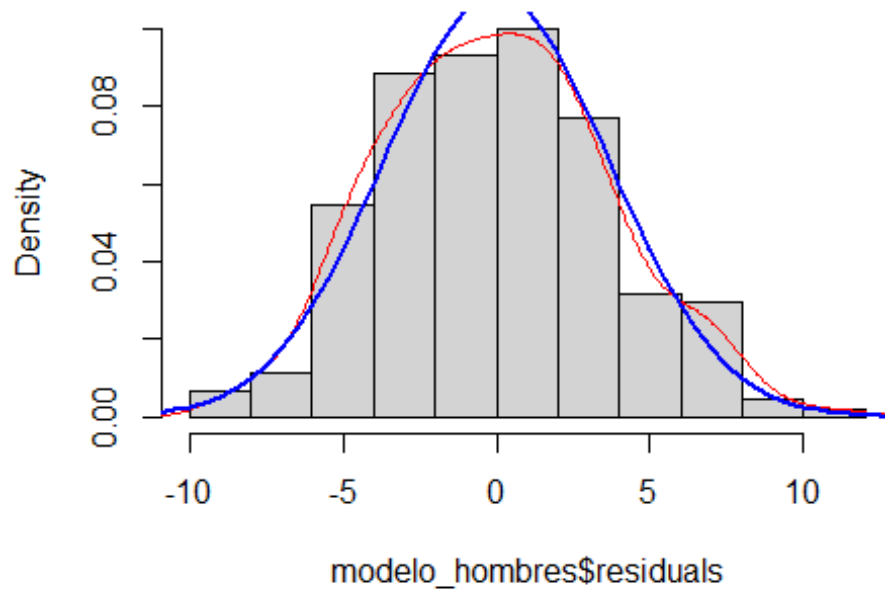
qqnorm(modelo_hombres$residuals)
qqline(modelo_hombres$residuals)
```

Normal Q-Q Plot



```
hist(modelo_hombres$residuals,freq=FALSE, ylim= c(0,0.1))  
lines(density(modelo_hombres$residual),col="red")  
curve(dnorm(x,mean=mean(modelo_hombres$residuals),sd=sd(modelo_hombres$residuals)), from=-30, to=20, add=TRUE, col="blue",lwd=2)
```

Histogram of modelo_hombres\$residuals



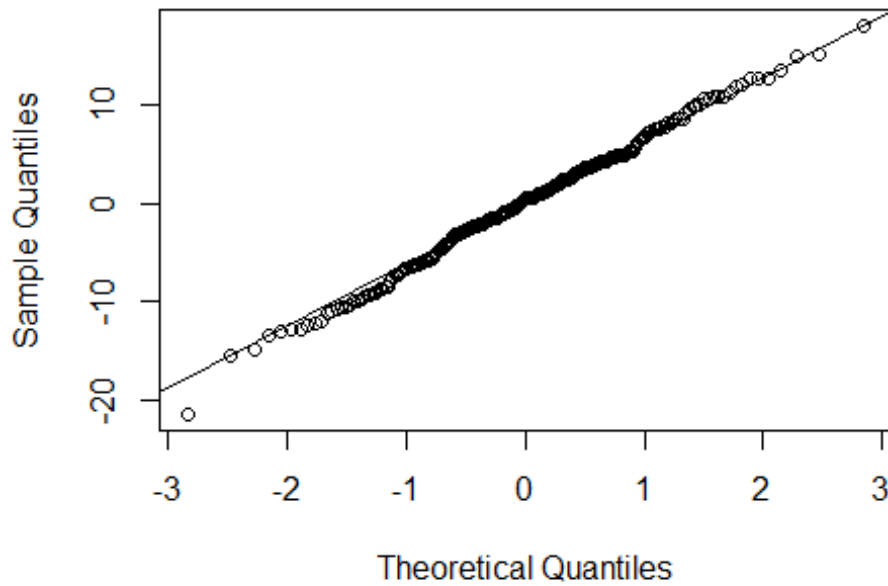
```
# Mujeres
```

```
ad.test(modelo_mujeres$residuals)

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

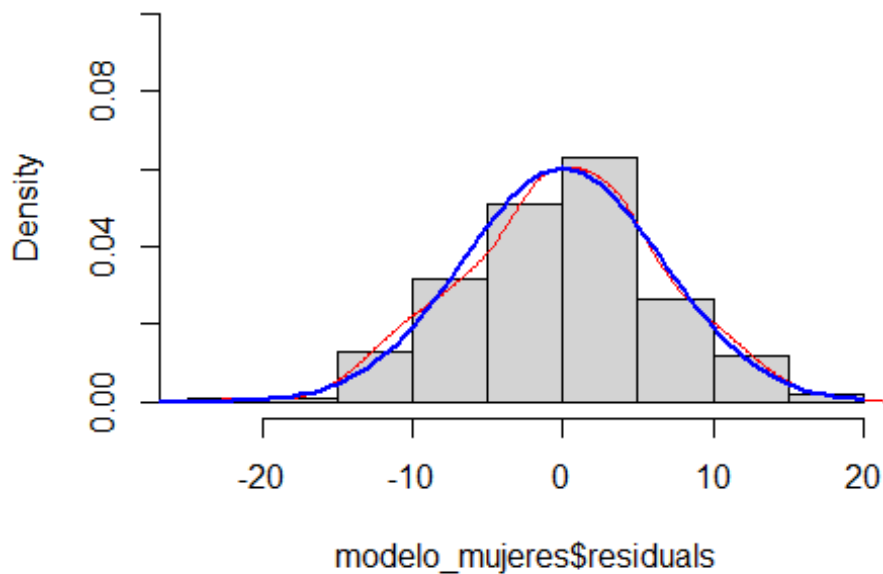
qqnorm(modelo_mujeres$residuals)
qqline(modelo_mujeres$residuals)
```

Normal Q-Q Plot



```
hist(modelo_mujeres$residuals,freq=FALSE, ylim= c(0,0.1))  
lines(density(modelo_mujeres$residual),col="red")  
curve(dnorm(x,mean=mean(modelo_mujeres$residuals),sd=sd(modelo_mujeres$residuals)), from=-30, to=20, add=TRUE, col="blue",lwd=2)
```

Histogram of modelo_mujeres\$residuals



Normalidad

de los residuos:

Hombres:

Shapiro-Wilk test: p-valor = 0.4597, lo que indica que no hay evidencia para rechazar la normalidad de los residuos.

Anderson-Darling test: p-valor = 0.5771, también apoya la normalidad de los residuos.

Mujeres:

Shapiro-Wilk test: p-valor = 0.9144, lo que indica que los residuos son normales.

Anderson-Darling test: p-valor = 0.7451, confirma la normalidad.

Se cumple el supuesto de normalidad de los residuos para ambos modelos (hombres y mujeres).

2. Verificación de media cero

$H_0: \mu = 0$ $H_1: \mu \neq 0$

```
# Verificación de media cero de Los residuos Hombres
media_residuos <- mean(residuals(modelo_hombres))
cat("\nMedia de los residuos: ", media_residuos, "\n")
```

```
##
## Media de los residuos: 1.125698e-16
```

```

# Verificación de media cero de Los residuos Mujeres
media_residuos <- mean(residuals(modelo_mujeres))
cat("\nMedia de los residuos: ", media_residuos, "\n")

##
## Media de los residuos: -1.788342e-16

# Prueba t para verificar si La media de Los residuos es
significativamente diferente de cero Hombres
t.test(modelo_hombres$residuals)

##
## 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

# Prueba t para verificar si La media de Los residuos es
significativamente diferente de cero Mujeres
t.test(modelo_mujeres$residuals)

##
## 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

```

Media cero de los residuos:

Hombres: One-sample t-test: p-valor = 1, lo que indica que la media de los residuos es cero, como se espera.

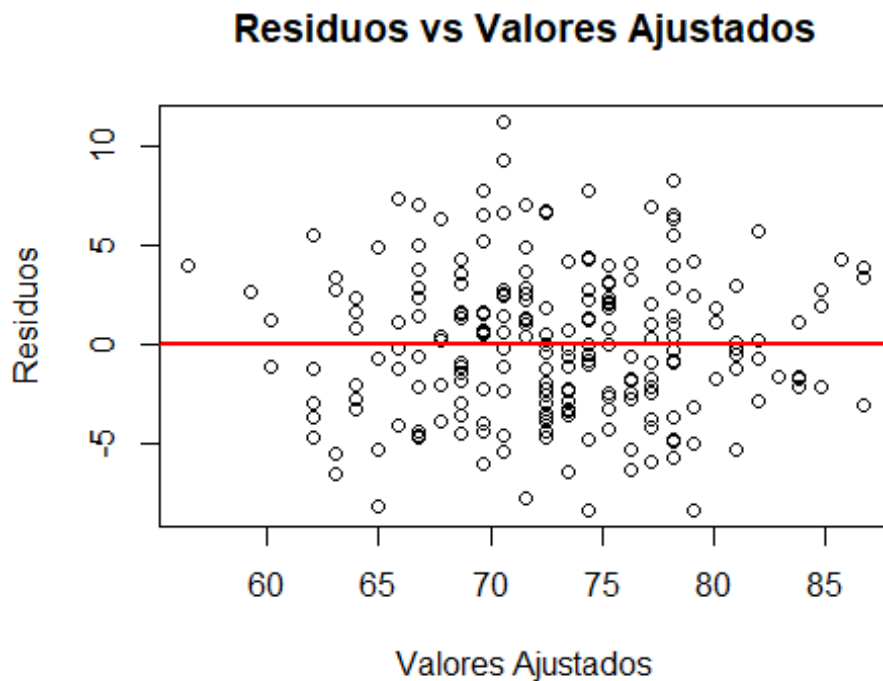
Mujeres: One-sample t-test: p-valor = 1, también sugiere que la media de los residuos es cero.

Se cumple el supuesto de que la media de los residuos es cero en ambos modelos.

3. Homocedasticidad

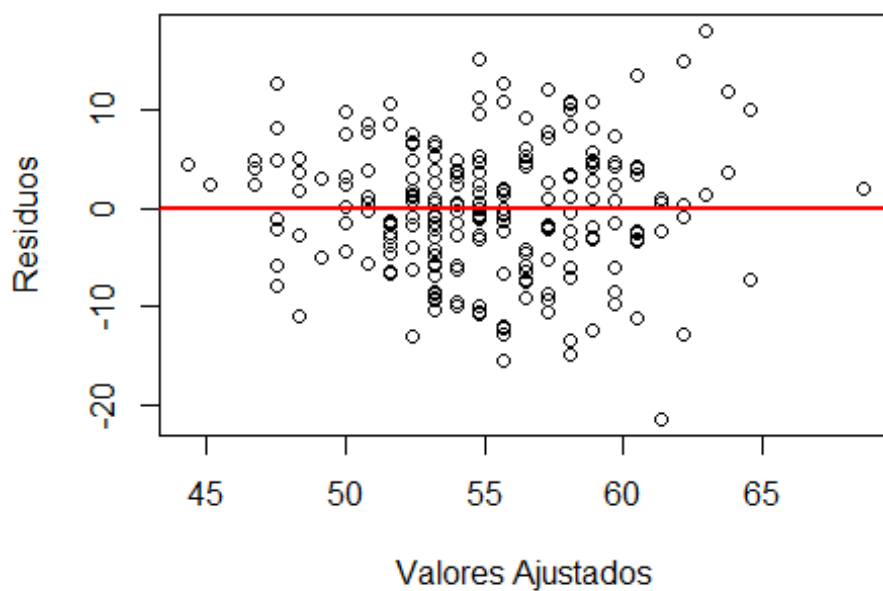
Prueba de Breusch-Pagan y White H_0 : La varianza de los errores es constante (homocedasticidad) H_1 : La varianza de los errores no es constante (heterocedasticidad)

```
# Homocedasticidad Hombres  
# Gráfico de Los residuos vs valores ajustados  
plot(fitted(modelo_hombres), residuals(modelo_hombres), main="Residuos vs  
Valores Ajustados", xlab="Valores Ajustados", ylab="Residuos")  
abline(h = 0, col = "red", lwd = 2)
```



```
# Homocedasticidad Mujeres  
# Gráfico de Los residuos vs valores ajustados  
plot(fitted(modelo_mujeres), residuals(modelo_mujeres), main="Residuos vs  
Valores Ajustados", xlab="Valores Ajustados", ylab="Residuos")  
abline(h = 0, col = "red", lwd = 2)
```


Residuos vs Valores Ajustados



```
library(lmtest)

# Hombres
bptest(modelo_hombres)

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

gqtest(modelo_hombres)

##
## Goldfeld-Quandt test
##
## data: modelo_hombres
## GQ = 0.84148, df1 = 108, df2 = 108, p-value = 0.8144
## alternative hypothesis: variance increases from segment 1 to 2

# Mujeres
bptest(modelo_mujeres)

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

```
gqtest(modelo_mujeres)

##
## Goldfeld-Quandt test
##
## data: modelo_mujeres
## GQ = 1.4265, df1 = 108, df2 = 108, p-value = 0.03313
## alternative hypothesis: variance increases from segment 1 to 2
```

Homocedasticidad:

Hombres:

Breusch-Pagan test: p-valor = 0.334, no hay evidencia de heterocedasticidad.

Goldfeld-Quandt test: p-valor = 0.8144, lo que refuerza la homocedasticidad.

Mujeres:

Breusch-Pagan test: p-valor = 0.003556, lo que sugiere heterocedasticidad.

Goldfeld-Quandt test: p-valor = 0.03313, lo que confirma una posible heterocedasticidad en los residuos. Conclusión:

En el modelo de hombres se cumple el supuesto de homocedasticidad.

En el modelo de mujeres no se cumple el supuesto de homocedasticidad (presencia de heterocedasticidad).

4. Independencia

Pruebas de hipótesis: Test de Durbin-Watson y Prueba Breusch-Godfrey para independencia H_0 : Los errores no están autocorrelacionados. H_1 Los errores están autocorrelacionados.

```
library(lmtest)

# Hombres
dwtest(modelo_hombres)

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

bgtest(modelo_hombres)

##
## Breusch-Godfrey test for serial correlation of order up to 1
##
## data: modelo_hombres
## LM test = 0.20778, df = 1, p-value = 0.6485
```

```
# Mujeres
dwtest(modelo_mujeres)

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

bgtest(modelo_mujeres)

##
## Breusch-Godfrey test for serial correlation of order up to 1
##
## data: modelo_mujeres
## LM test = 1.4655, df = 1, p-value = 0.2261
```

Independencia de los residuos:

Hombres:

*+Durbin-Watson test: **DW = 2.0556, p-valor = 0.6599, lo que indica que no hay autocorrelación significativa.** Breusch-Godfrey test:** p-valor = 0.6485, lo que confirma la independencia de los residuos.

Mujeres:

Durbin-Watson test: DW = 1.8062, p-valor = 0.07532, lo que sugiere que no hay evidencia fuerte de autocorrelación.

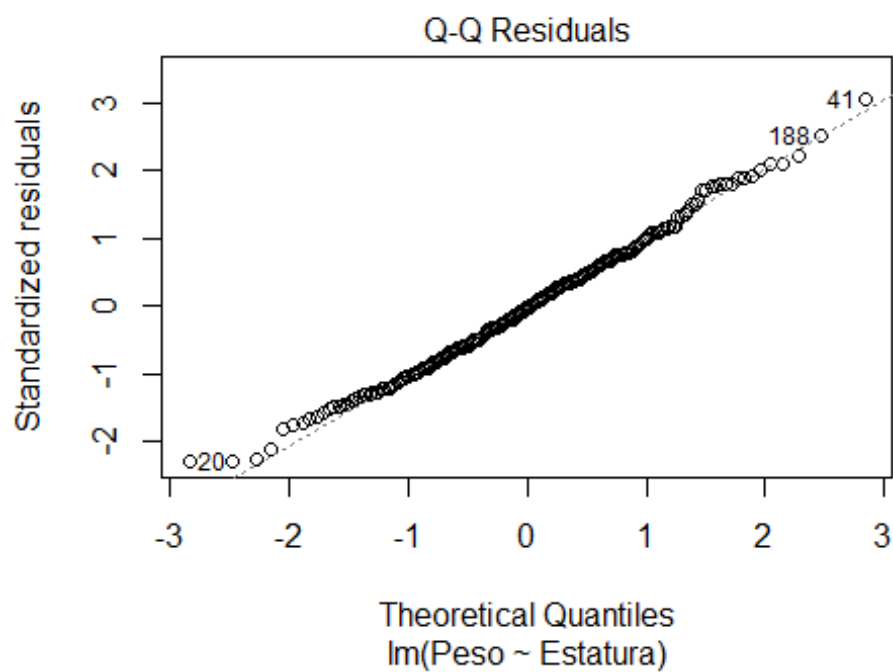
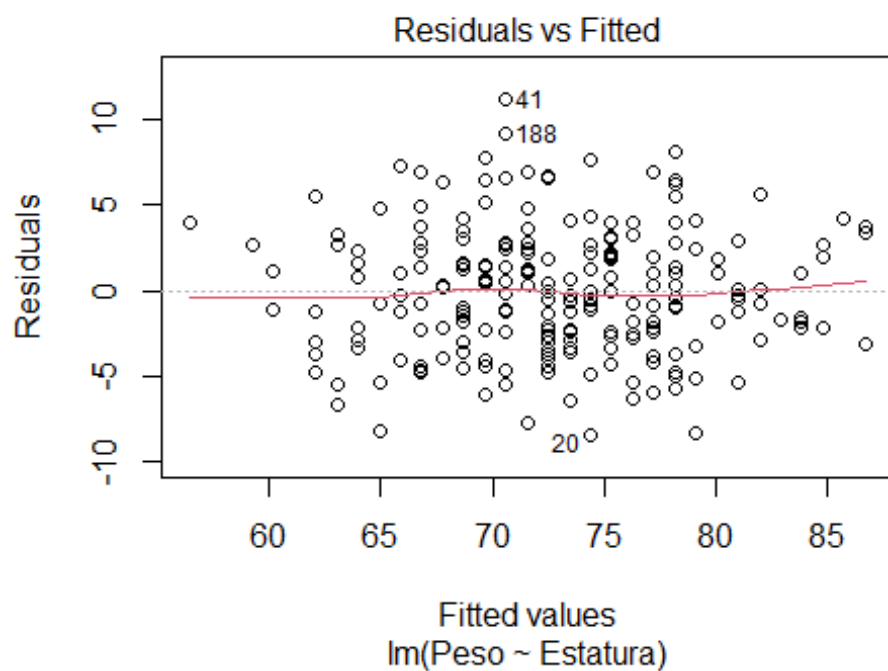
Breusch-Godfrey test: p-valor = 0.2261, lo que confirma la independencia de los residuos.

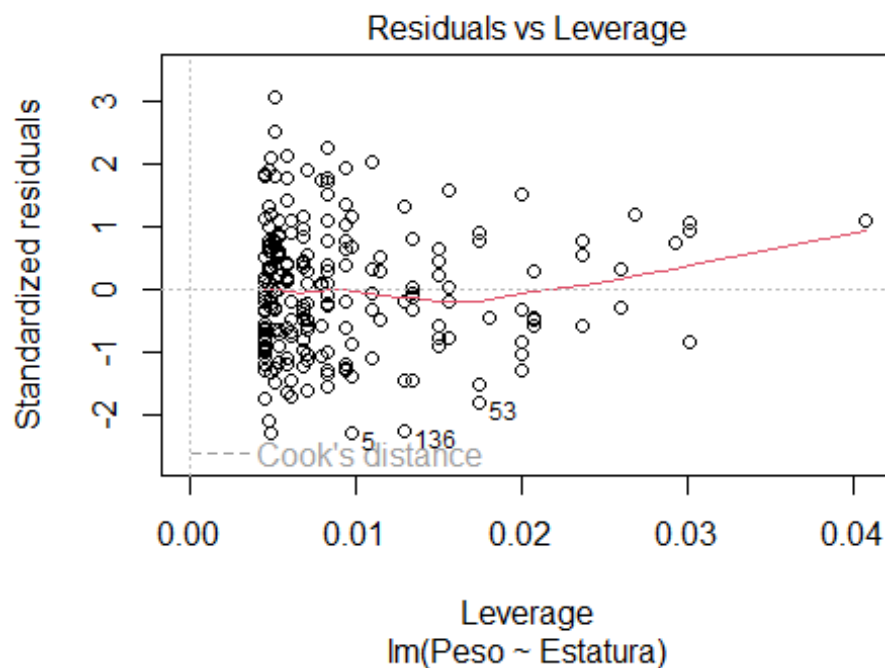
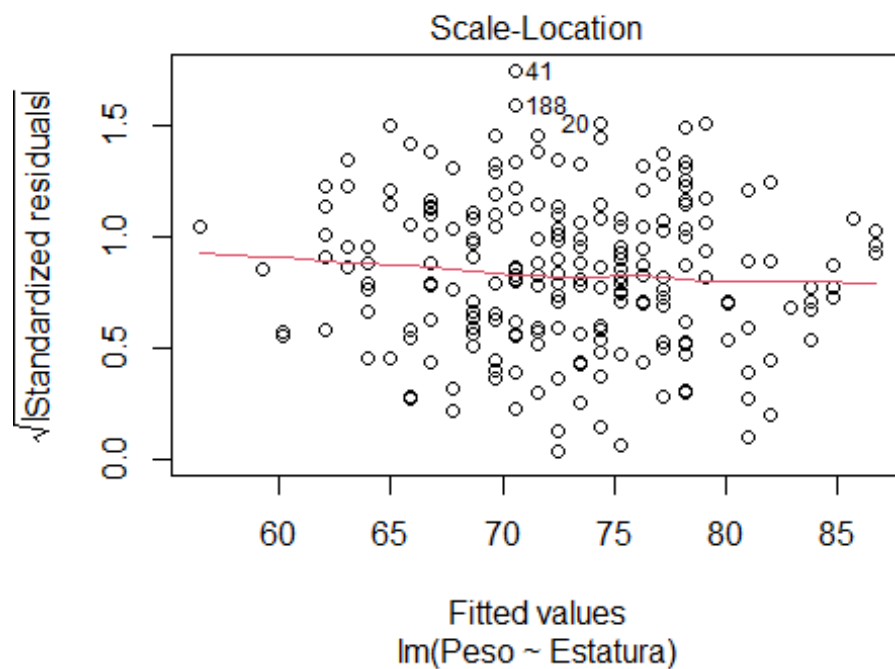
Se cumple el supuesto de independencia de los residuos para ambos modelos.

Hombres: Se cumplen todos los supuestos (normalidad, media cero, homocedasticidad, independencia).

Mujeres: Se cumplen todos los supuestos excepto el de homocedasticidad, ya que se observó heterocedasticidad significativa.

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





¿Cuáles son las diferencias y similitudes de estos gráficos con respecto a los que ya habías analizado? ** Gráfico Q-Q (Normalidad de los residuos):** Ambas gráficas Q-Q muestran que los puntos siguen de cerca la línea roja diagonal, lo que sugiere que los residuos se distribuyen aproximadamente de manera normal. Ambas gráficas Q-Q

indican que el supuesto de normalidad de los residuos se cumple razonablemente bien.

Histograma de los residuos: El histograma de los residuos parece tener una distribución cercana a la normal, lo que está respaldado por las curvas de densidad superpuestas. Aunque no es completamente simétrico, la forma general del histograma y las curvas de densidad sugieren que la distribución de los residuos está lo suficientemente cerca de la normalidad. Este gráfico respalda la normalidad de los residuos.

Residuos vs Valores Ajustados (Homocedasticidad): Este gráfico muestra cómo se distribuyen los residuos en función de los valores ajustados por el modelo. La dispersión de los puntos alrededor de la línea roja (en 0) es relativamente uniforme, lo que sugiere que no hay un patrón claro en la variabilidad de los residuos. La homocedasticidad parece mantenerse, ya que no hay evidencia clara de que los residuos se “abran” o “cierren” en forma de embudo. Estas mismas observaciones existen en los gráficos anteriores.

Scale-Location: Verifica si la varianza de los residuos es constante. Si el gráfico no muestra un patrón claro de apertura o cierre, se cumple la homocedasticidad.

Residuals vs Leverage: Identifica puntos influyentes. Si algunos puntos tienen un leverage alto y grandes residuos, podrían estar afectando de manera desproporcionada el ajuste del modelo.

Estos gráficos de diagnóstico sugieren que los supuestos clave del modelo lineal se cumplen

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

No cambian las conclusiones con respecto a las obtenidas con anterioridad.

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

El mejor modelo fue el de hombres, ya que fue el único en cumplir con los 4 supuestos del modelo de regresión lineal: normalidad, media cero, homocedasticidad e independencia.

- Los residuos eran normales.
- La media de los residuos era cero.
- Se observó homocedasticidad (varianza constante de los errores).
- No hubo autocorrelación de los residuos.
- No se detectaron puntos influyentes importantes que pudieran desestabilizar el modelo.

Intervalos en regresión

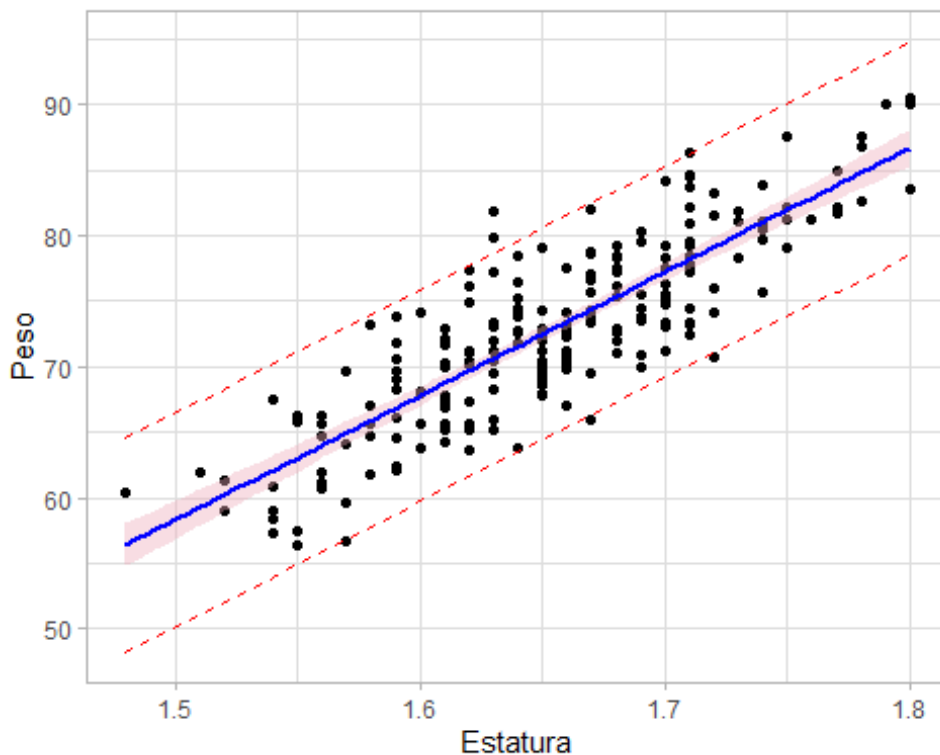
```
Ip=predict(object=modelo_hombres,interval="prediction",level=0.97)
```

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

datos=cbind(dataH,Ip)
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.3.3

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

Mejor modelo

Puntos de datos: Los puntos negros representan los datos observados de **Peso** en función de **Estatura** para los hombres. Es decir, son las observaciones reales del conjunto de datos.

Línea de regresión: La línea azul en el centro es la línea de **regresión ajustada**. Esta representa la mejor aproximación lineal entre la variable predictora (Estatura) y la variable dependiente (Peso). Esta línea predice el peso promedio de los hombres para una determinada estatura.

Intervalo de confianza de la regresión: El área sombreada en rosa alrededor de la línea azul es el **intervalo de confianza** para la regresión al 97% de confianza. Indica la incertidumbre de la estimación de la línea de regresión. Si se repitiera el muestreo, esperaríamos que el valor promedio de peso para una altura dada cayera dentro de esta región el 97% de las veces.

Intervalos de predicción: Las líneas rojas discontinuas representan los **intervalos de predicción** al 97%. Este intervalo tiene en cuenta la variabilidad tanto en la estimación del valor promedio como en la variabilidad individual de los datos. Los intervalos de predicción son más amplios que los intervalos de confianza, lo que refleja la mayor incertidumbre cuando intentamos predecir un nuevo valor de Peso dado una estatura específica. Es decir, predicen el rango en el que caerían **nuevas observaciones individuales** del Peso, no solo el valor promedio.

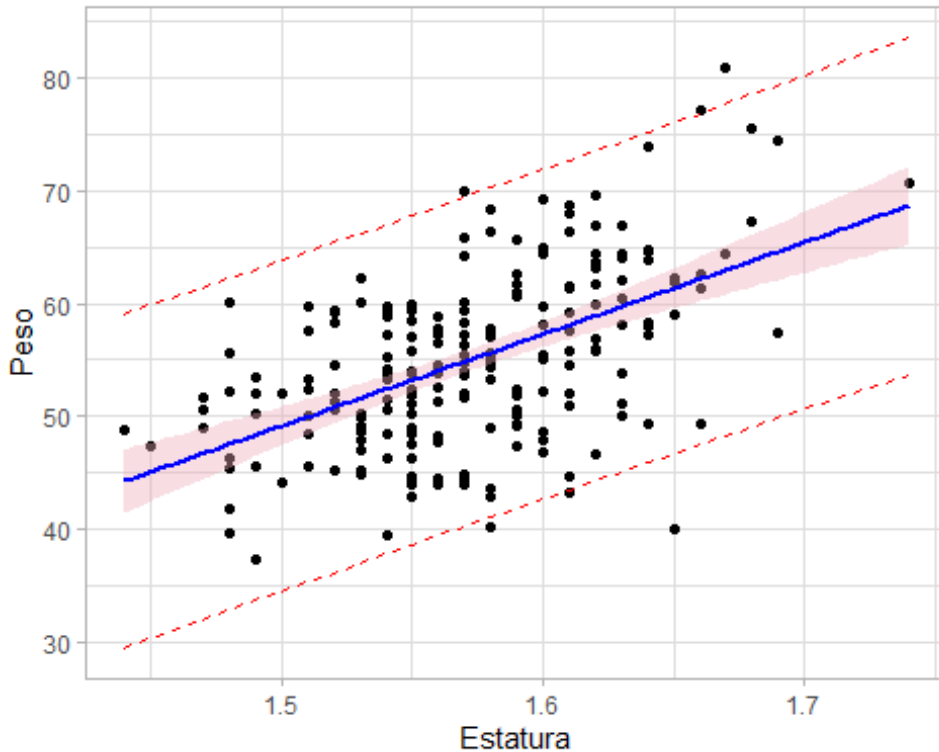
La gráfica muestra una **relación lineal positiva** entre la **estatura** y el **peso** de los hombres: a mayor estatura, mayor es el peso estimado. La línea de regresión azul es la mejor estimación del peso promedio para una estatura dada. Las líneas rojas discontinuas muestran el rango donde esperaríamos que caigan nuevos puntos de datos, y es más amplio que el área sombreada debido a la incertidumbre adicional en predicciones futuras.

Modelo 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()
```

En

comparación con el modelo de los hombres:

Mayor dispersión: En el gráfico de las mujeres, los puntos están más dispersos alrededor de la línea de regresión en comparación con los hombres. Esto puede indicar una mayor variabilidad no explicada por el modelo.

Pendiente menor: La relación entre la estatura y el peso parece ser menos fuerte en el modelo de las mujeres, lo que sugiere que otros factores, además de la estatura, podrían estar influyendo en el peso.

Amplios intervalos de predicción: Los intervalos de predicción son bastante amplios, lo que refleja la mayor incertidumbre en predecir el peso de una mujer basándose solo en su estatura.

La gráfica muestra una relación positiva entre la estatura y el peso de las mujeres, pero con mayor variabilidad en los datos y mayor incertidumbre en las predicciones. Esto refuerza los resultados previos que indicaban que el modelo de regresión para las mujeres no ajusta tan bien como el de los hombres, debido a la mayor dispersión y a la posible presencia de heterocedasticidad.