

Regresión Lineal - Análisis de los errores

Luis Maximiliano López Ramírez

2024-09-04

Validez del modelo

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

```
datos <- read.csv("Estatura-peso_HyM.csv")

datos$Sexo_numeric <- ifelse(datos$Sexo == 'H', 0, 1)

dataM = subset(datos, datos$Sexo=="M")
dataH = subset(datos, datos$Sexo == "H")

modelo_sin_interaccion <- lm(Peso ~ Estatura + Sexo_numeric, data =
datos)

modelo_con_interaccion <- lm(Peso ~ Estatura * Sexo_numeric, data =
datos)

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

modelo_mujeres <- lm(Peso ~ Estatura, data = dataM)
```

1. Normalidad de los residuos

% Hipótesis de Normalidad (Anderson-Darling)

H_0 : Los residuos siguen una distribución normal.
 H_1 : Los residuos no siguen una distribución normal.

```
library(nortest)

resultado_sin_interaccion <- ad.test(modelo_sin_interaccion$residuals)
resultado_con_interaccion <- ad.test(modelo_con_interaccion$residuals)
resultado_hombres <- ad.test(modelo_hombres$residuals)
resultado_mujeres <- ad.test(modelo_mujeres$residuals)

# Mostrar Los resultados de Las pruebas
cat("Resultados de la prueba de normalidad Anderson-Darling:\n")
```

Resultados de la prueba de normalidad Anderson-Darling:

```
cat("\nModelo sin interacción:\n")
```

```
##
```

```
## Modelo sin interacción:
```

```
print(resultado_sin_interaccion)
```

```
##
```

```
## Anderson-Darling normality test
```

```
##
```

```
## data: modelo_sin_interaccion$residuals
```

```
## A = 0.79651, p-value = 0.03879
```

```
cat("\nModelo con interacción:\n")
```

```
##
```

```
## Modelo con interacción:
```

```
print(resultado_con_interaccion)
```

```
##
```

```
## Anderson-Darling normality test
```

```
##
```

```
## data: modelo_con_interaccion$residuals
```

```
## A = 0.8138, p-value = 0.03516
```

```
cat("\nModelo de hombres:\n")
```

```
##
```

```
## Modelo de hombres:
```

```
print(resultado_hombres)
```

```
##
```

```
## Anderson-Darling normality test
```

```
##
```

```
## data: modelo_hombres$residuals
```

```
## A = 0.3009, p-value = 0.5771
```

```
cat("\nModelo de mujeres:\n")
```

```
##
```

```
## Modelo de mujeres:
```

```
print(resultado_mujeres)
```

```
##
```

```
## Anderson-Darling normality test
```

```
##
```

```
## data: modelo_mujeres$residuals
```

```
## A = 0.24899, p-value = 0.7451
```

```

# Verificar normalidad en base al valor p
if (resultado_sin_interaccion$p.value > 0.03) {
  cat("\nSe tiene normalidad en el modelo sin interacción.\n")
} else {
  cat("\nNo se tiene normalidad en el modelo sin interacción.\n")
}

##
## Se tiene normalidad en el modelo sin interacción.

if (resultado_con_interaccion$p.value > 0.03) {
  cat("\nSe tiene normalidad en el modelo con interacción.\n")
} else {
  cat("\nNo se tiene normalidad en el modelo con interacción.\n")
}

##
## Se tiene normalidad en el modelo con interacción.

if (resultado_hombres$p.value > 0.03) {
  cat("\nSe tiene normalidad en el modelo de hombres.\n")
} else {
  cat("\nNo se tiene normalidad en el modelo de hombres.\n")
}

##
## Se tiene normalidad en el modelo de hombres.

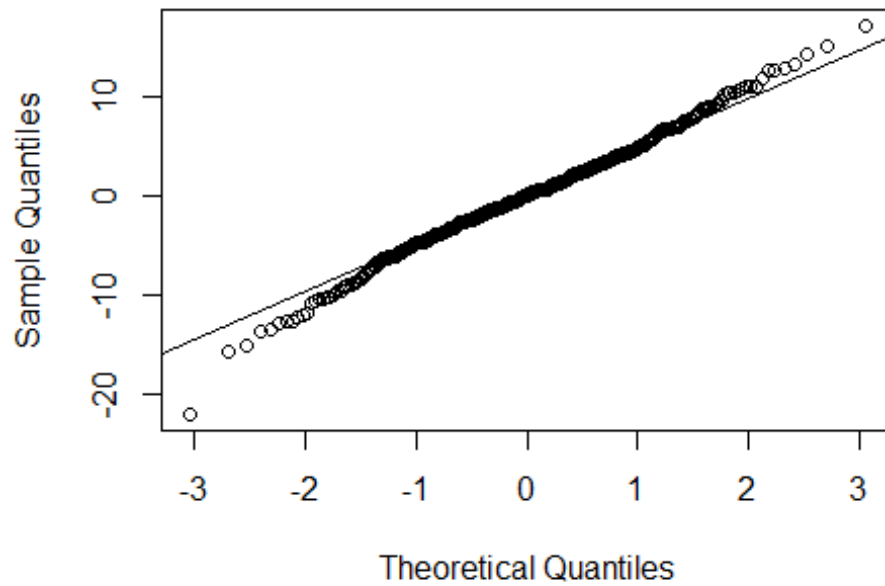
if (resultado_mujeres$p.value > 0.03) {
  cat("\nSe tiene normalidad en el modelo de mujeres.\n")
} else {
  cat("\nNo se tiene normalidad en el modelo de mujeres.\n")
}

##
## Se tiene normalidad en el modelo de mujeres.

qqnorm(modelo_sin_interaccion$residuals)
qqline(modelo_sin_interaccion$residuals)

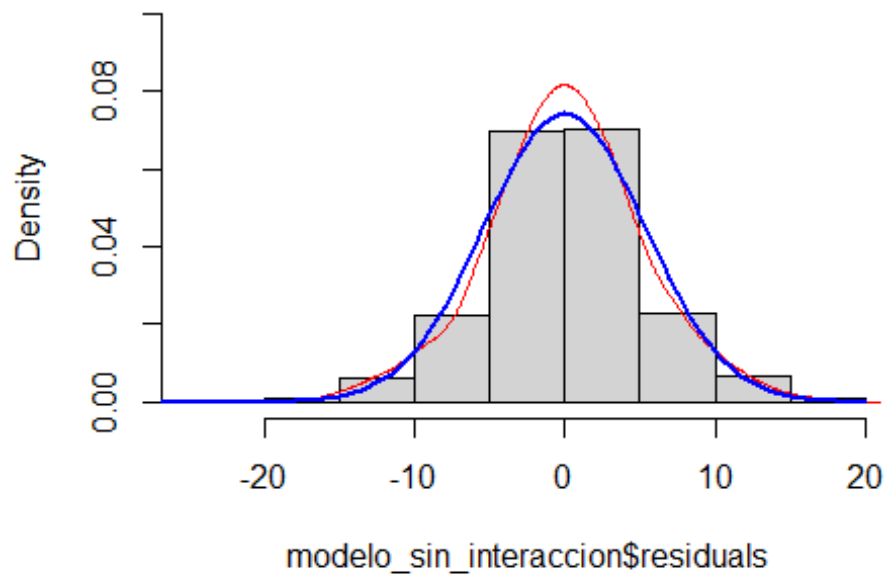
```

Normal Q-Q Plot



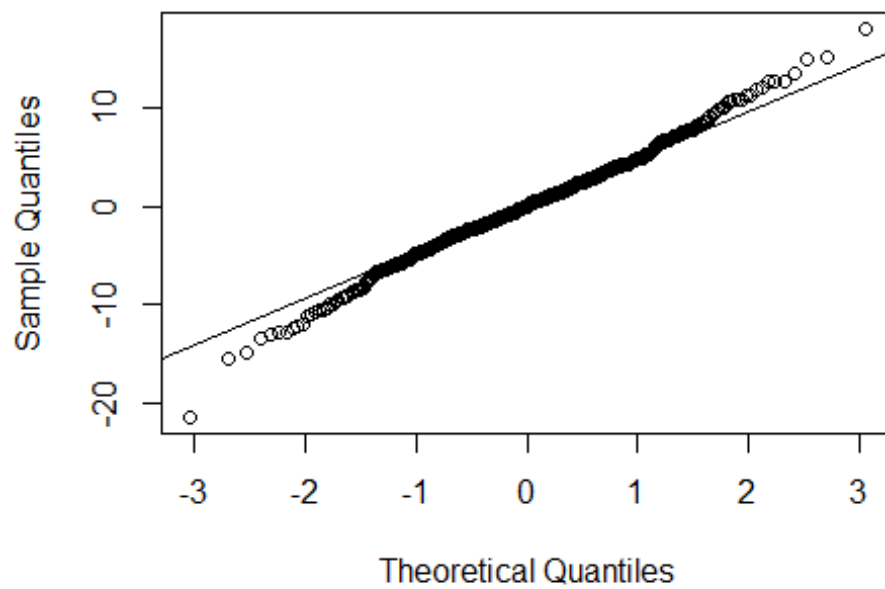
```
hist(modelo_sin_interaccion$residuals,freq=FALSE , ylim = c(0, 0.1))  
lines(density(modelo_sin_interaccion$residual),col="red")  
curve(dnorm(x,mean=mean(modelo_sin_interaccion$residuals),sd=sd(modelo_si  
n_interaccion$residuals)), from=-30, to=20, add=TRUE, col="blue",lwd=2)
```

Histogram of modelo_sin_interaccion\$residuals



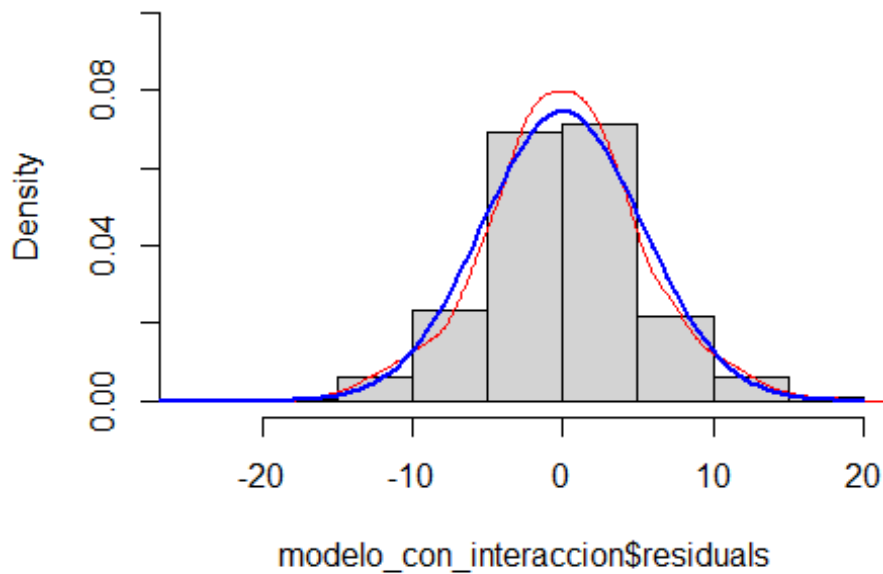
```
qqnorm(modelo_con_interaccion$residuals)  
qqline(modelo_con_interaccion$residuals)
```

Normal Q-Q Plot



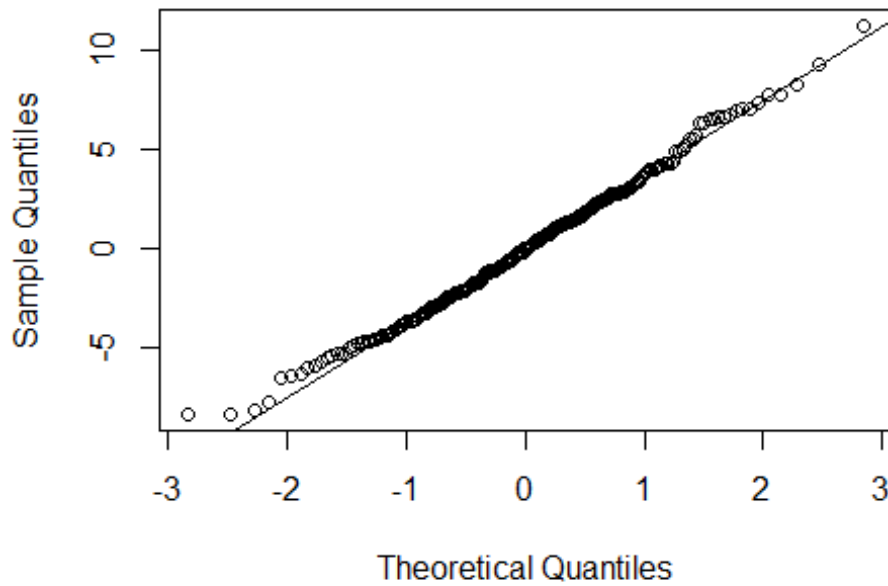
```
hist(modelo_con_interaccion$residuals,freq=FALSE, ylim = c(0, 0.1))
lines(density(modelo_con_interaccion$residual),col="red")
curve(dnorm(x,mean=mean(modelo_con_interaccion$residuals),sd=sd(modelo_con_interaccion$residuals)), from=-30, to=20, add=TRUE, col="blue",lwd=2)
```

Histogram of modelo_con_interaccion\$residuals



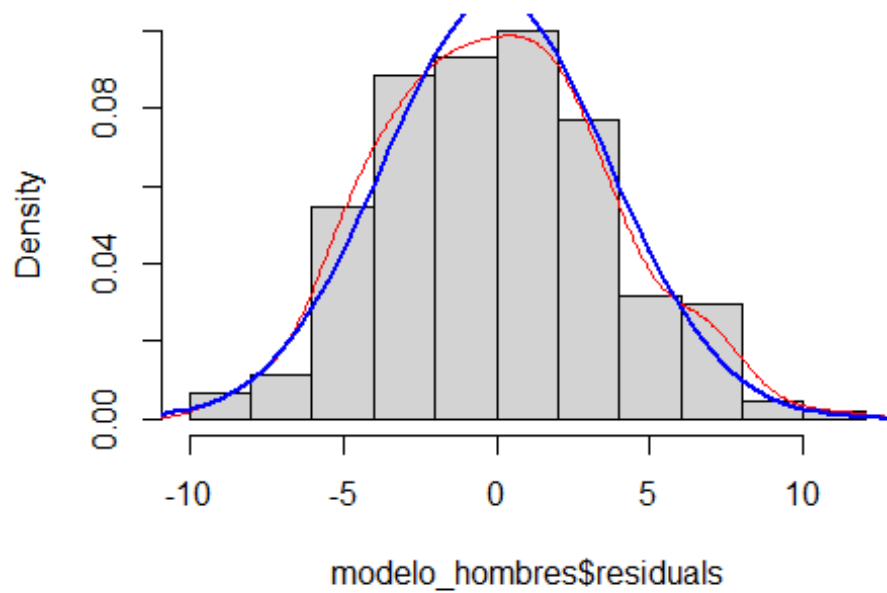
```
qqnorm(modelo_hombres$residuals)
qqline(modelo_hombres$residuals)
```

Normal Q-Q Plot



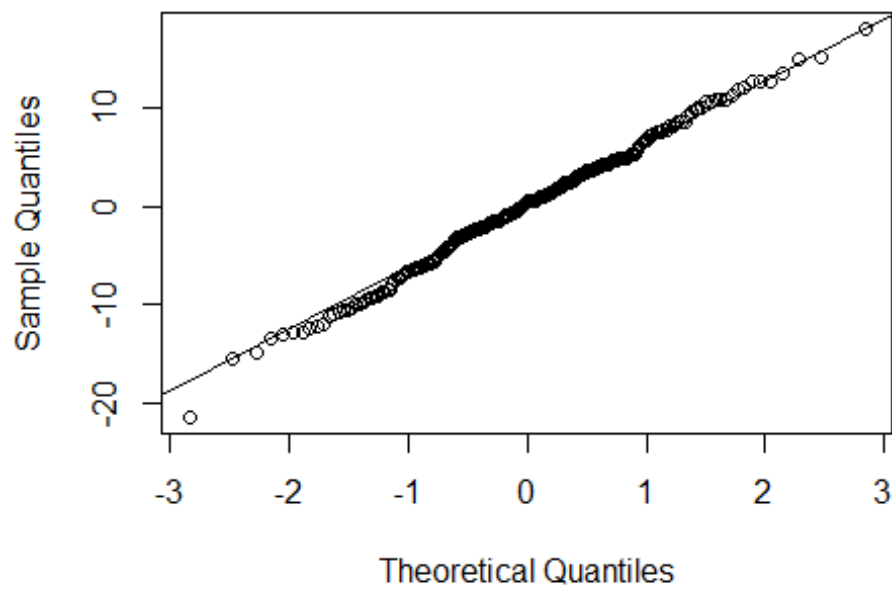
```
hist(modelo_hombres$residuals,freq=FALSE, ylim = c(0, 0.1))  
lines(density(modelo_hombres$residual),col="red")  
curve(dnorm(x,mean=mean(modelo_hombres$residuals),sd=sd(modelo_hombres$residuals)), from=-30, to=20, add=TRUE, col="blue",lwd=2)
```

Histogram of modelo_hombres\$residuals



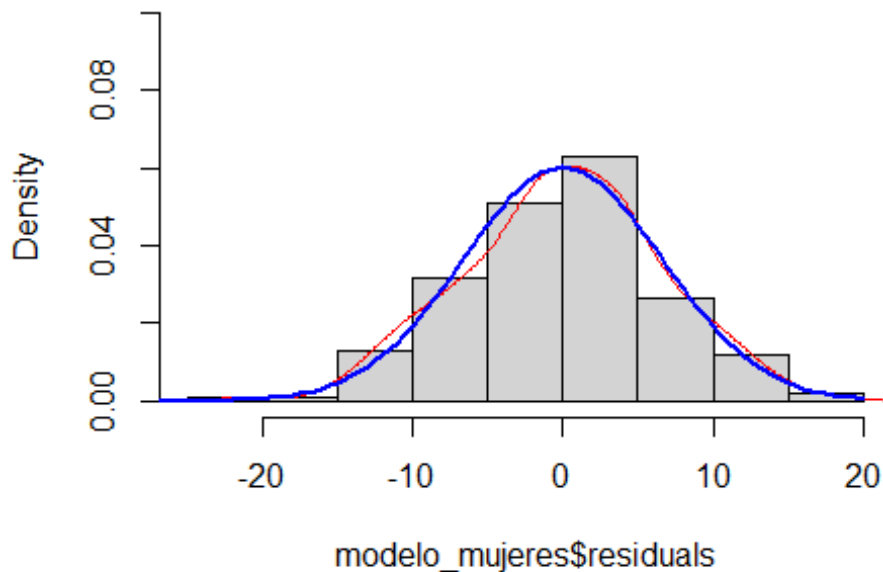
```
qqnorm(modelo_mujeres$residuals)  
qqline(modelo_mujeres$residuals)
```

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



2. Verificación de media cero

% Hipótesis de Media Diferente de Cero (t de Student)

$$\begin{cases} H_0: \mu = 0 & \text{(El promedio de los residuos es igual a cero)} \\ H_1: \mu \neq 0 & \text{(El promedio de los residuos es diferente de cero)} \end{cases}$$

```
# Realiza La prueba t para Los residuos sin interacción
resultado_t_sin_interaccion <- t.test(modelo_sin_interaccion$residuals)

resultado_t_con_interaccion <- t.test(modelo_con_interaccion$residuals)

resultado_t_hombres <- t.test(modelo_hombres$residuals)

resultado_t_mujeres <- t.test(modelo_mujeres$residuals)

# Mostrar Los resultados de Las pruebas
cat("Resultados de la prueba t de Student para los residuos:\n")

## Resultados de la prueba t de Student para los residuos:
cat("Modelo sin interacción:\n")
```

```

## Modelo sin interacción:

print(resultado_t_sin_interaccion)

##
## One Sample t-test
##
## data: modelo_sin_interaccion$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

cat("\nModelo con interacción:\n")

##
## Modelo con interacción:

print(resultado_t_con_interaccion)

##
## One Sample t-test
##
## data: modelo_con_interaccion$residuals
## t = -8.5817e-16, df = 439, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -0.5017741 0.5017741
## sample estimates:
## mean of x
## -2.190956e-16

cat("Modelo de hombres:\n")

## Modelo de hombres:

print(resultado_t_hombres)

##
## 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

```

```

cat("\nModelo de mujeres:\n")

##
## Modelo de mujeres:

print(resultado_t_mujeres)

##
## 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

# Verificar si el promedio de los residuos es significativamente
diferente de cero
if (resultado_t_sin_interaccion$p.value > 0.03) {
  cat("\nEl promedio de los residuos del modelo sin interacción no es
significativamente diferente de cero.\n")
} else {
  cat("\nEl promedio de los residuos del modelo sin interacción es
significativamente diferente de cero.\n")
}

##
## El promedio de los residuos del modelo sin interacción no es
significativamente diferente de cero.

if (resultado_t_con_interaccion$p.value > 0.03) {
  cat("\nEl promedio de los residuos del modelo con interacción no es
significativamente diferente de cero.\n")
} else {
  cat("\nEl promedio de los residuos del modelo con interacción es
significativamente diferente de cero.\n")
}

##
## El promedio de los residuos del modelo con interacción no es
significativamente diferente de cero.

if (resultado_t_hombres$p.value > 0.03) {
  cat("\nEl promedio de los residuos del modelo sin interacción no es
significativamente diferente de cero.\n")
} else {
  cat("\nEl promedio de los residuos del modelo sin interacción es
significativamente diferente de cero.\n")
}

```

```
##
## El promedio de los residuos del modelo sin interacción no es
significativamente diferente de cero.

if (resultado_t_mujeres$p.value > 0.03) {
  cat("\nEl promedio de los residuos del modelo de hombres no es
significativamente diferente de cero.\n")
} else {
  cat("\nEl promedio de los residuos del modelo de mujeres es
significativamente diferente de cero.\n")
}

##
## El promedio de los residuos del modelo de hombres no es
significativamente diferente de cero.
```

3. Homocedasticidad e independencia

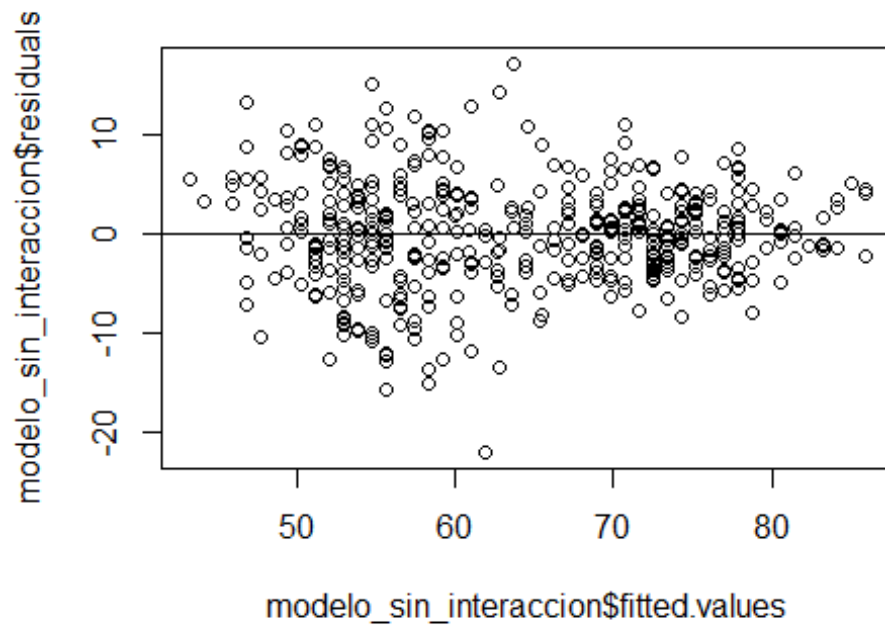
% Hipótesis de Autocorrelación (Durbin-Watson y Breusch-Godfrey)

$\begin{cases} H_0: \text{Los errores no están autocorrelacionados (independencia).} \\ H_1: \text{Los errores están autocorrelacionados.} \end{cases}$

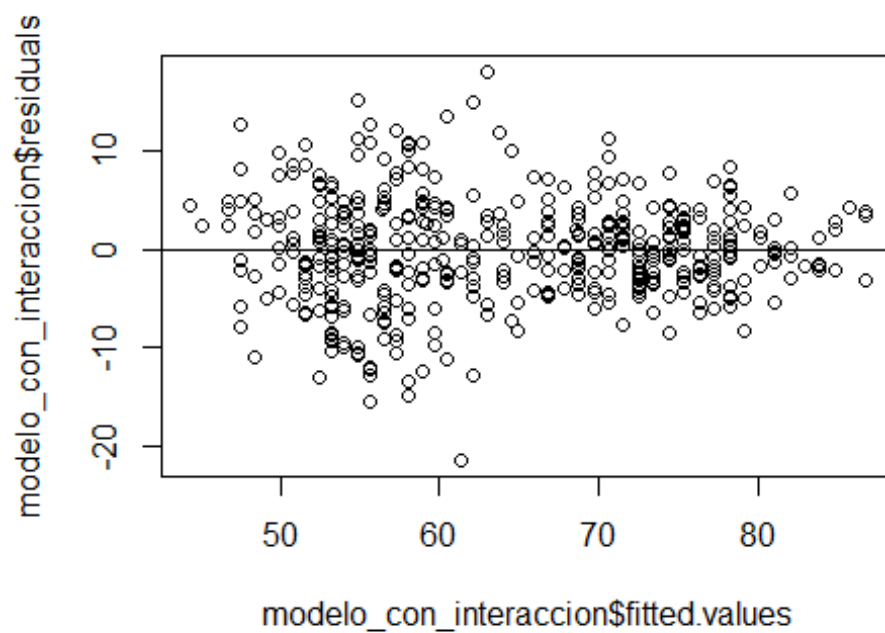
% Hipótesis de Homocedasticidad (Breusch-Pagan y Goldfeld-Quandt)

$\begin{cases} H_0: \text{La varianza de los errores es constante (homocedasticidad).} \\ H_1: \text{La varianza de los errores no es constante (heterocedasticidad).} \end{cases}$

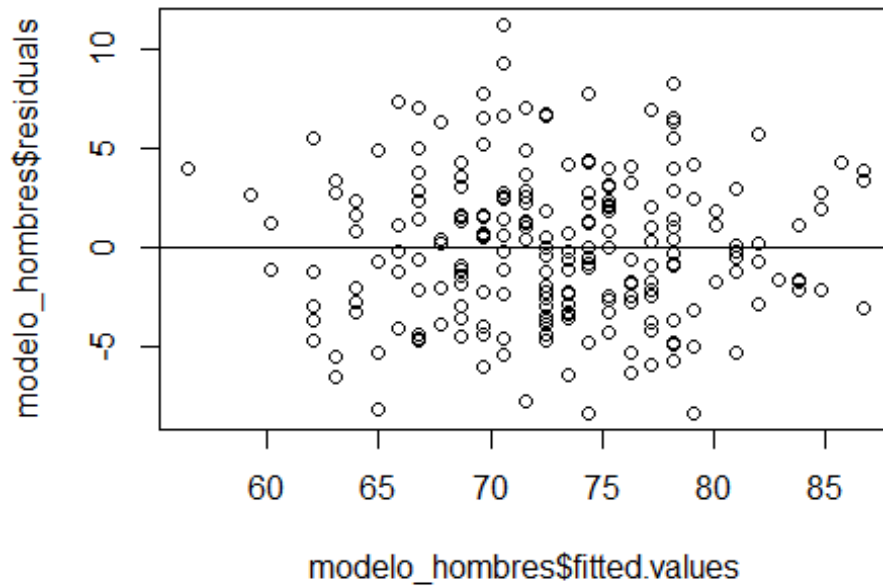
```
plot(modelo_sin_interaccion$fitted.values, modelo_sin_interaccion$residuals)
abline(h=0)
```



```
plot(modelo_con_interaccion$fitted.values, modelo_con_interaccion$residuals)
abline(h=0)
```



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



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

```
library(lmtest)
```

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

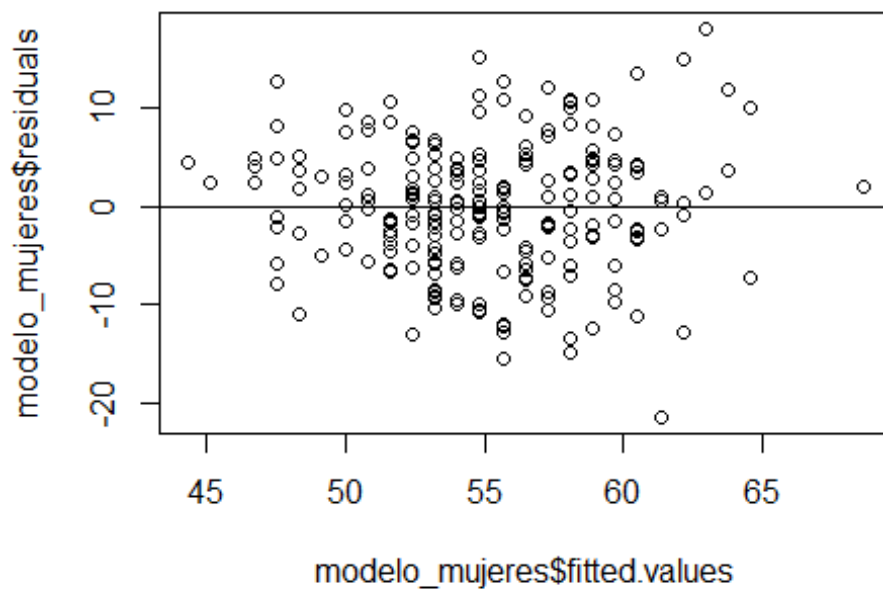
```
##
```

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

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## as.Date, as.Date.numeric
```



```
# Pruebas de autocorrelación de errores
cat("Pruebas de autocorrelación de errores:\n")

## Pruebas de autocorrelación de errores:

# Modelo sin interacción
dw_sin_interaccion <- dwtest(modelo_sin_interaccion)
bg_sin_interaccion <- bgtest(modelo_sin_interaccion)

cat("\nModelo sin interacción")

##
## Modelo sin interacción

# Verificar autocorrelación en base al valor p
if (dw_sin_interaccion$p.value > 0.03) {
  cat("\nNo se rechaza H0: Los errores no están autocorrelacionados
(Durbin-Watson).\nTiene independencia.\n")
} else {
  cat("\nSe rechaza H0: Los errores están autocorrelacionados (Durbin-
Watson).\nNo tiene independencia.\n")
}

##
## No se rechaza H0: Los errores no están autocorrelacionados (Durbin-
Watson).
## Tiene independencia.
```

```

if (bg_sin_interaccion$p.value > 0.03) {
  cat("No se rechaza H0: Los errores no están autocorrelacionados
(Breusch-Godfrey).\nTiene independencia.\n")
} else {
  cat("Se rechaza H0: Los errores están autocorrelacionados (Breusch-
Godfrey).\nNo tiene independencia.\n")
}

## No se rechaza H0: Los errores no están autocorrelacionados (Breusch-
Godfrey).
## Tiene independencia.

# Modelo con interacción
dw_con_interaccion <- dwtest(modelo_con_interaccion)
bg_con_interaccion <- bgtest(modelo_con_interaccion)

cat("\nModelo con interacción")

##
## Modelo con interacción

# Verificar autocorrelación en base al valor p
if (dw_con_interaccion$p.value > 0.03) {
  cat("\nNo se rechaza H0: Los errores no están autocorrelacionados
(Durbin-Watson).\nTiene independencia.\n")
} else {
  cat("\nSe rechaza H0: Los errores están autocorrelacionados (Durbin-
Watson).\nNo tiene independencia.\n")
}

##
## No se rechaza H0: Los errores no están autocorrelacionados (Durbin-
Watson).
## Tiene independencia.

if (bg_con_interaccion$p.value > 0.03) {
  cat("No se rechaza H0: Los errores no están autocorrelacionados
(Breusch-Godfrey).\nTiene independencia.\n")
} else {
  cat("Se rechaza H0: Los errores están autocorrelacionados (Breusch-
Godfrey).\nNo tiene independencia.\n")
}

## No se rechaza H0: Los errores no están autocorrelacionados (Breusch-
Godfrey).
## Tiene independencia.

# Modelo sin interacción
dw_hombres <- dwtest(modelo_hombres)
bg_hombres <- bgtest(modelo_hombres)

cat("\nModelo de hombres")

```



```
##
## Modelo de hombres

# Verificar autocorrelación en base al valor p
if (dw_hombres$p.value > 0.03) {
  cat("\nNo se rechaza H0: Los errores no están autocorrelacionados
(Durbin-Watson).\nTiene independencia.\n")
} else {
  cat("\nSe rechaza H0: Los errores están autocorrelacionados (Durbin-
Watson).\nNo tiene independencia.\n")
}

##
## No se rechaza H0: Los errores no están autocorrelacionados (Durbin-
Watson).
## Tiene independencia.

if (bg_hombres$p.value > 0.03) {
  cat("No se rechaza H0: Los errores no están autocorrelacionados
(Breusch-Godfrey).\nTiene independencia.\n")
} else {
  cat("Se rechaza H0: Los errores están autocorrelacionados (Breusch-
Godfrey).\nNo tiene independencia.\n")
}

## No se rechaza H0: Los errores no están autocorrelacionados (Breusch-
Godfrey).
## Tiene independencia.

# Modelo con interacción
dw_mujeres <- dwtest(modelo_mujeres)
bg_mujeres <- bgtest(modelo_mujeres)

cat("\nModelo de mujeres")

##
## Modelo de mujeres

# Verificar autocorrelación en base al valor p
if (dw_mujeres$p.value > 0.03) {
  cat("\nNo se rechaza H0: Los errores no están autocorrelacionados
(Durbin-Watson).\nTiene independencia.\n")
} else {
  cat("\nSe rechaza H0: Los errores están autocorrelacionados (Durbin-
Watson).\nNo tiene independencia.\n")
}

##
## No se rechaza H0: Los errores no están autocorrelacionados (Durbin-
Watson).
## Tiene independencia.
```

```

if (bg_mujeres$p.value > 0.03) {
  cat("No se rechaza H0: Los errores no están autocorrelacionados (Breusch-Godfrey).\nTiene independencia.\n")
} else {
  cat("Se rechaza H0: Los errores están autocorrelacionados (Breusch-Godfrey).\nNo tiene independencia.\n")
}

## No se rechaza H0: Los errores no están autocorrelacionados (Breusch-Godfrey).
## Tiene independencia.

# Pruebas de homocedasticidad
cat("\nPruebas de homocedasticidad:\n")

##
## Pruebas de homocedasticidad:

# Modelo sin interacción
bp_sin_interaccion <- bptest(modelo_sin_interaccion)
gq_sin_interaccion <- gqtest(modelo_sin_interaccion)

cat("\nModelo sin interacción")

##
## Modelo sin interacción

# Verificar homocedasticidad en base al valor p
if (bp_sin_interaccion$p.value > 0.03) {
  cat("\nNo se rechaza H0: La varianza de los errores es constante (Breusch-Pagan).\nTiene homocedasticidad.\n")
} else {
  cat("\nSe rechaza H0: La varianza de los errores no es constante (Breusch-Pagan).\nNo tiene homocedasticidad.\n")
}

##
## Se rechaza H0: La varianza de los errores no es constante (Breusch-Pagan).
## No tiene homocedasticidad.

if (gq_sin_interaccion$p.value > 0.03) {
  cat("No se rechaza H0: La varianza de los errores es constante (Goldfeld-Quandt).\nTiene homocedasticidad.\n")
} else {
  cat("Se rechaza H0: La varianza de los errores no es constante (Goldfeld-Quandt).\nNo tiene homocedasticidad.\n")
}

## Se rechaza H0: La varianza de los errores no es constante (Goldfeld-Quandt).
## No tiene homocedasticidad.

```

```

# Modelo con interacción
bp_con_interaccion <- bptest(modelo_con_interaccion)
gq_con_interaccion <- gqtest(modelo_con_interaccion)

cat("\nModelo con interacción")

##
## Modelo con interacción

# Verificar homocedasticidad en base al valor p
if (bp_con_interaccion$p.value > 0.03) {
  cat("\nNo se rechaza H0: La varianza de los errores es constante
(Breusch-Pagan).\nTiene homocedasticidad.\n")
} else {
  cat("\nSe rechaza H0: La varianza de los errores no es constante
(Breusch-Pagan).\nNo tiene homocedasticidad.\n")
}

##
## Se rechaza H0: La varianza de los errores no es constante (Breusch-
Pagan).
## No tiene homocedasticidad.

if (gq_con_interaccion$p.value > 0.03) {
  cat("No se rechaza H0: La varianza de los errores es constante
(Goldfeld-Quandt).\nTiene homocedasticidad.\n")
} else {
  cat("Se rechaza H0: La varianza de los errores no es constante
(Goldfeld-Quandt).\nNo tiene homocedasticidad.\n")
}

## Se rechaza H0: La varianza de los errores no es constante (Goldfeld-
Quandt).
## No tiene homocedasticidad.

# Modelo sin interacción
bp_hombres <- bptest(modelo_hombres)
gq_hombres <- gqtest(modelo_hombres)

cat("\nModelo de hombres")

##
## Modelo de hombres

# Verificar homocedasticidad en base al valor p
if (bp_hombres$p.value > 0.03) {
  cat("\nNo se rechaza H0: La varianza de los errores es constante
(Breusch-Pagan).\nTiene homocedasticidad.\n")
} else {
  cat("\nSe rechaza H0: La varianza de los errores no es constante
(Breusch-Pagan).\nNo tiene homocedasticidad.\n")
}

```

```

##
## No se rechaza  $H_0$ : La varianza de los errores es constante (Breusch-
Pagan).
## Tiene homocedasticidad.

if (gq_hombres$p.value > 0.03) {
  cat("No se rechaza  $H_0$ : La varianza de los errores es constante
(Goldfeld-Quandt).\nTiene homocedasticidad.\n")
} else {
  cat("Se rechaza  $H_0$ : La varianza de los errores no es constante
(Goldfeld-Quandt).\nNo tiene homocedasticidad.\n")
}

## No se rechaza  $H_0$ : La varianza de los errores es constante (Goldfeld-
Quandt).
## Tiene homocedasticidad.

# Modelo con interacción
bp_mujeres <- bptest(modelo_mujeres)
gq_mujeres <- gqtest(modelo_mujeres)

cat("\nModelo de mujeres")

##
## Modelo de mujeres

# Verificar homocedasticidad en base al valor p
if (bp_mujeres$p.value > 0.03) {
  cat("\nNo se rechaza  $H_0$ : La varianza de los errores es constante
(Breusch-Pagan).\nTiene homocedasticidad.\n")
} else {
  cat("\nSe rechaza  $H_0$ : La varianza de los errores no es constante
(Breusch-Pagan).\nNo tiene homocedasticidad.\n")
}

##
## Se rechaza  $H_0$ : La varianza de los errores no es constante (Breusch-
Pagan).
## No tiene homocedasticidad.

if (gq_mujeres$p.value > 0.03) {
  cat("No se rechaza  $H_0$ : La varianza de los errores es constante
(Goldfeld-Quandt).\nTiene homocedasticidad.\n")
} else {
  cat("Se rechaza  $H_0$ : La varianza de los errores no es constante
(Goldfeld-Quandt).\nNo tiene homocedasticidad.\n")
}

## No se rechaza  $H_0$ : La varianza de los errores es constante (Goldfeld-
Quandt).
## Tiene homocedasticidad.

```

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

Con un nivel de significancia del 0.03 los cuatros modelos cumplen con los supuestos de normalidad, media NO diferente de 0 e independencia, sin embargo, los únicos modelos que cumplen con el supuesto de homocedasticidad son el modelo de hombres y del modelo de mujeres en solamente una de las dos pruebas de homocedasticidad.

Esto quiere decir que el único modelo válido es el que está separado en hombres y mujeres sin embargo, el coeficiente de determinación del modelo de mujeres es menor a 0.3 por lo que es muy malo sin embargo, es lo que hay y se tomaría dicho modelo.

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

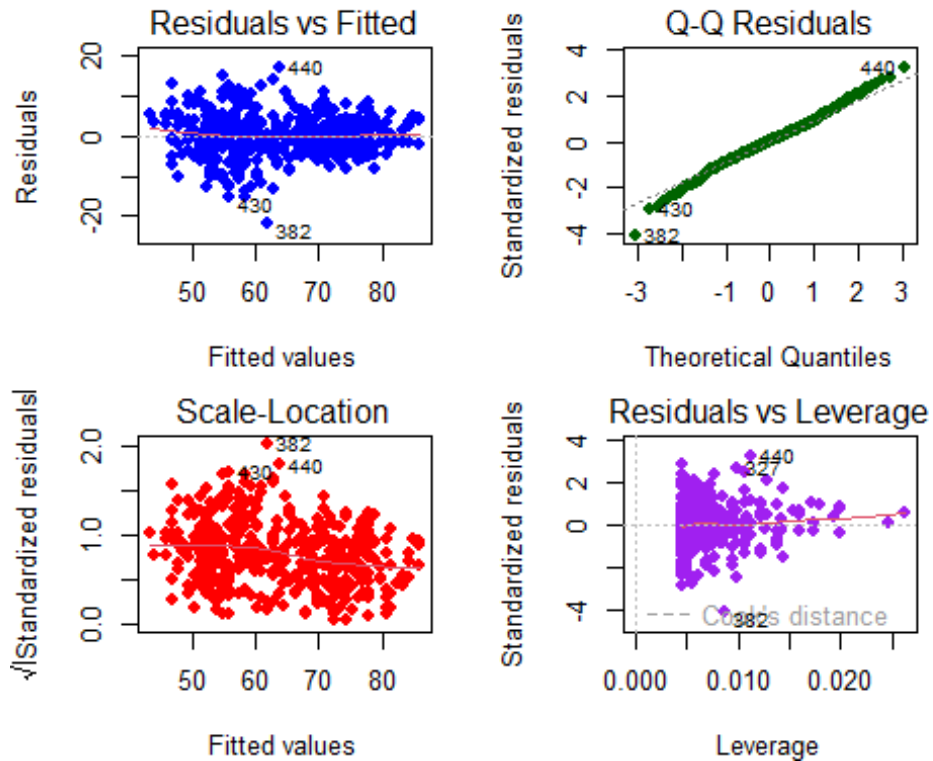
```
par(mfrow = c(2, 2), mar = c(4, 4, 2, 2))

# Gráfico 1: Residuos vs Ajustados
plot(modelo_sin_interaccion, which = 1, col = "blue", pch = 16)

# Gráfico 2: Normal Q-Q
plot(modelo_sin_interaccion, which = 2, col = "darkgreen", pch = 16)

# Gráfico 3: Scale-Location
plot(modelo_sin_interaccion, which = 3, col = "red", pch = 16)

# Gráfico 4: Residuos vs Leverage
plot(modelo_sin_interaccion, which = 5, col = "purple", pch = 16)
```



```
# Restablecer los parámetros gráficos a su configuración original
par(mfrow = c(1, 1))

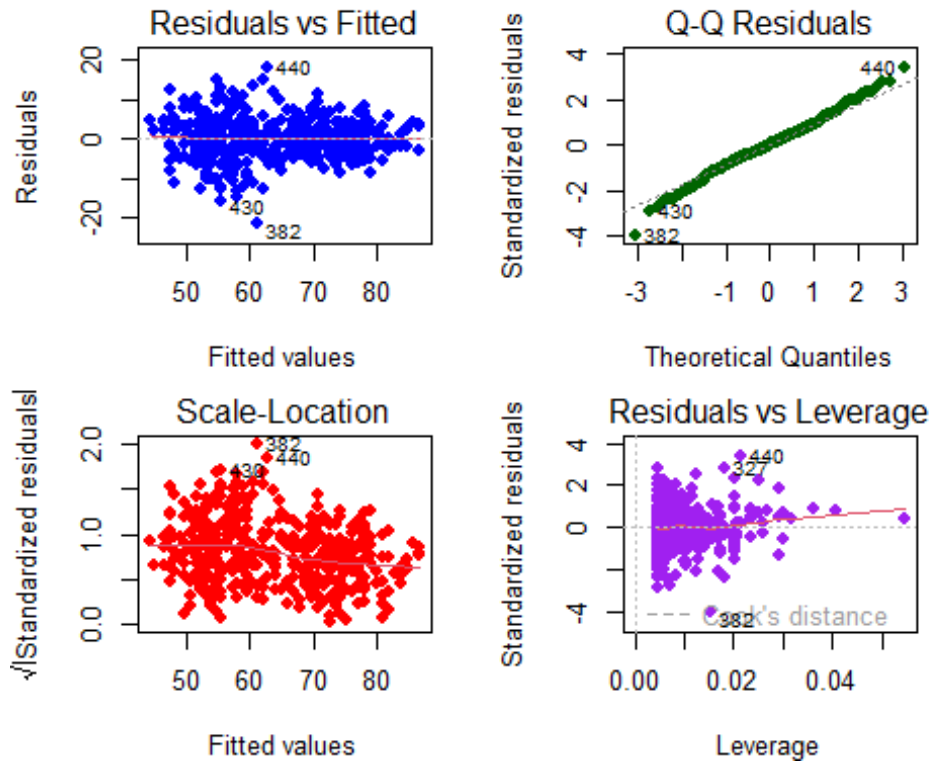
par(mfrow = c(2, 2), mar = c(4, 4, 2, 2))

# Gráfico 1: Residuos vs Ajustados
plot(modelo_con_interaccion, which = 1, col = "blue", pch = 16)

# Gráfico 2: Normal Q-Q
plot(modelo_con_interaccion, which = 2, col = "darkgreen", pch = 16)

# Gráfico 3: Scale-Location
plot(modelo_con_interaccion, which = 3, col = "red", pch = 16)

# Gráfico 4: Residuos vs Leverage
plot(modelo_con_interaccion, which = 5, col = "purple", pch = 16)
```



```
# Restablecer los parámetros gráficos a su configuración original
par(mfrow = c(1, 1))
```

```
par(mfrow = c(2, 2), mar = c(4, 4, 2, 2))
```

```
# Gráfico 1: Residuos vs Ajustados
```

```
plot(modelo_hombres, which = 1, col = "blue", pch = 16)
```

```
# Gráfico 2: Normal Q-Q
```

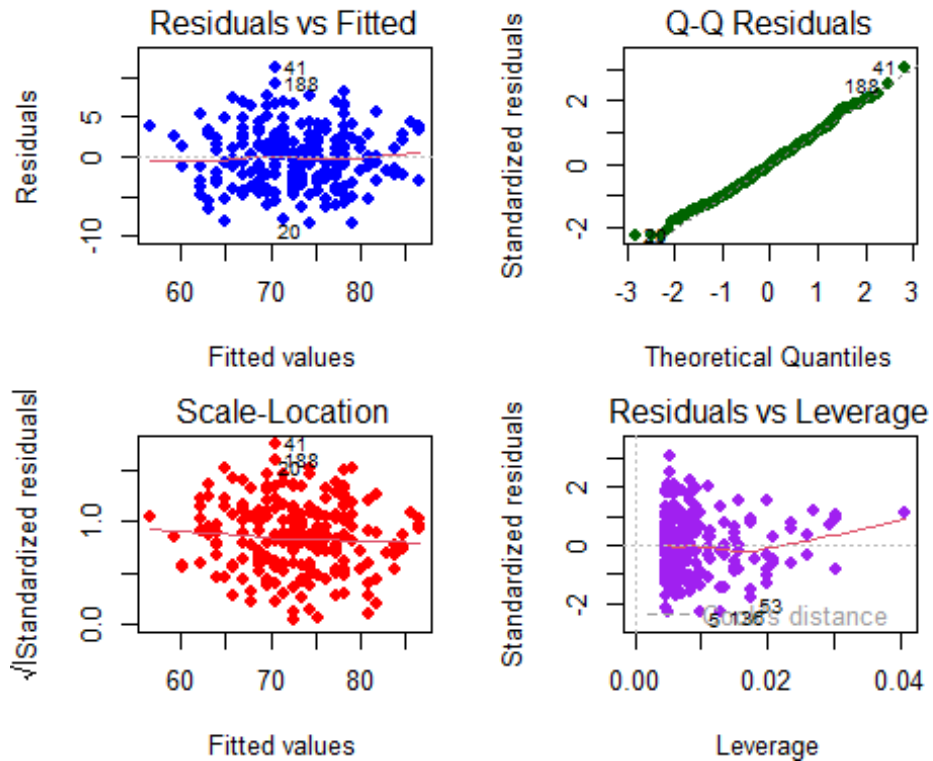
```
plot(modelo_hombres, which = 2, col = "darkgreen", pch = 16)
```

```
# Gráfico 3: Scale-Location
```

```
plot(modelo_hombres, which = 3, col = "red", pch = 16)
```

```
# Gráfico 4: Residuos vs Leverage
```

```
plot(modelo_hombres, which = 5, col = "purple", pch = 16)
```



```
# Restablecer los parámetros gráficos a su configuración original
par(mfrow = c(1, 1))
```

```
par(mfrow = c(2, 2), mar = c(4, 4, 2, 2))
```

```
# Gráfico 1: Residuos vs Ajustados
```

```
plot(modelo_mujeres, which = 1, col = "blue", pch = 16)
```

```
# Gráfico 2: Normal Q-Q
```

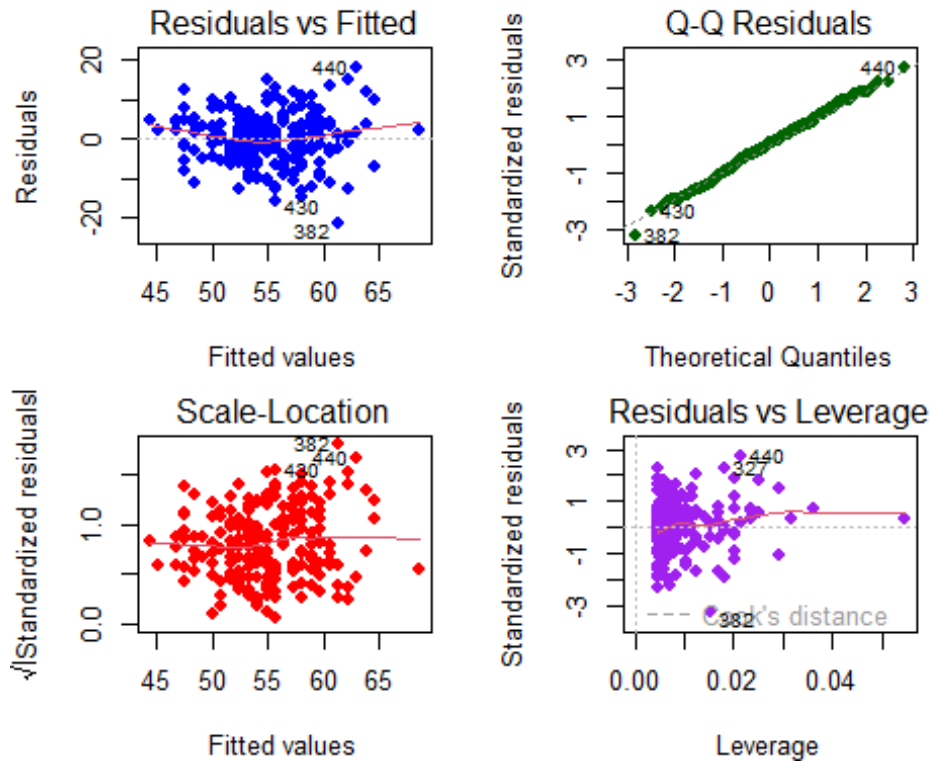
```
plot(modelo_mujeres, which = 2, col = "darkgreen", pch = 16)
```

```
# Gráfico 3: Scale-Location
```

```
plot(modelo_mujeres, which = 3, col = "red", pch = 16)
```

```
# Gráfico 4: Residuos vs Leverage
```

```
plot(modelo_mujeres, which = 5, col = "purple", pch = 16)
```

```
# Restablecer los parámetros gráficos a su configuración original
par(mfrow = c(1, 1))
```

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

Las gráficas generadas son iguales a las que ya se habían generado previamente tanto en el de Residual vs Fitted como el de Q-Q Residuals

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

No, ya que son las mismas gráficas que ya se habían obtenido previamente.

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

Inicialmente el mejor modelo fue el de sin interacción ya que tenía un coeficiente de determinación muy alto además de tener todos sus coeficientes significativos sin embargo, a la hora de revisar los supuestos no se cumplió el supuesto de homocedasticidad siendo el modelo separado en hombres y mujeres el único que cumple con todos los supuestos sin embargo, hay que tomar en cuenta que el modelo de mujeres tiene un coeficiente de determinación muy bajo pero se compensa con el coeficiente de determinación de los hombres por lo que se tomaría dicho modelo ya que es el único válido.

Intervalos de confianza

1. 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.

```
modelo_hombres <- lm(Peso ~ Estatura, data = dataH)

modelo_mujeres <- lm(Peso ~ Estatura, data = dataM)

# Predicciones con intervalos de predicción para hombres
pred_hombres <- predict(modelo_hombres, interval = "prediction", level =
0.97)

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

dataH_pred <- cbind(dataH, pred_hombres)

# Predicciones con intervalos de predicción para mujeres
pred_mujeres <- predict(modelo_mujeres, interval = "prediction", level =
0.97)

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

dataM_pred <- cbind(dataM, pred_mujeres)

library(ggplot2)

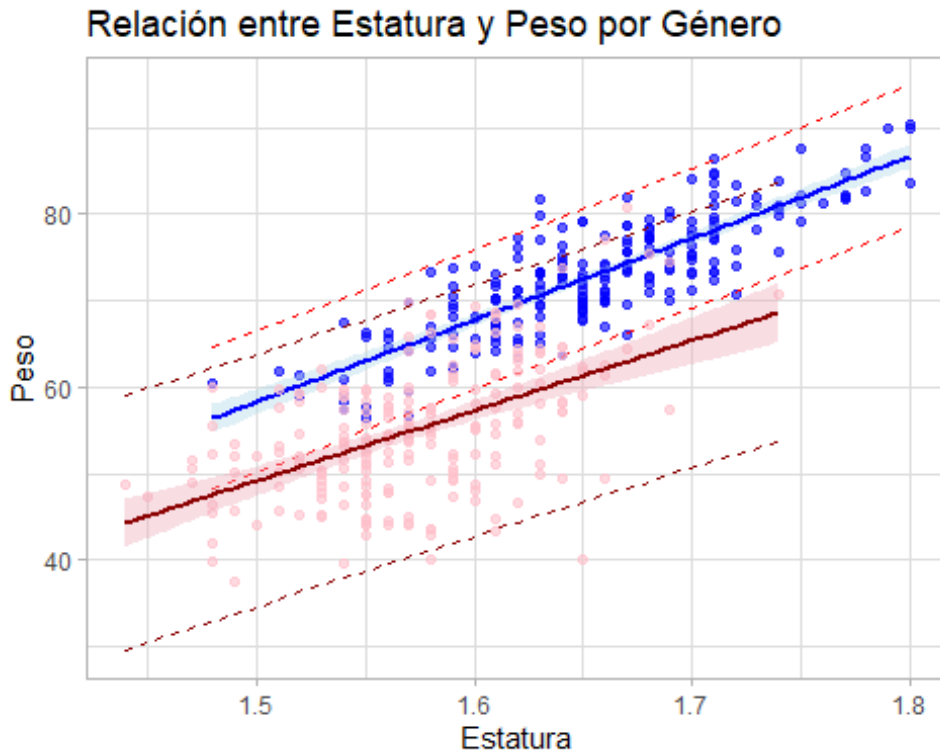
## Warning: package 'ggplot2' was built under R version 4.3.2

# Crear el gráfico combinando ambos conjuntos de datos
ggplot() +
  # Gráfico para hombres
  geom_point(data = dataH_pred, aes(x = Estatura, y = Peso), color =
"blue", alpha = 0.6) +
  geom_line(data = dataH_pred, aes(x = Estatura, y = lwr), color = "red",
linetype = "dashed") +
  geom_line(data = dataH_pred, aes(x = Estatura, y = upr), color = "red",
linetype = "dashed") +
  geom_smooth(data = dataH_pred, aes(x = Estatura, y = Peso), method =
lm, formula = y ~ x, se = TRUE, level = 0.97, col = "blue", fill =
"lightblue") +

  # Gráfico para mujeres
  geom_point(data = dataM_pred, aes(x = Estatura, y = Peso), color =
"pink", alpha = 0.6) +
  geom_line(data = dataM_pred, aes(x = Estatura, y = lwr), color =
"darkred", linetype = "dashed") +
  geom_line(data = dataM_pred, aes(x = Estatura, y = upr), color =
"darkred", linetype = "dashed") +
  geom_smooth(data = dataM_pred, aes(x = Estatura, y = Peso), method =
```

```
lm, formula = y ~ x, se = TRUE, level = 0.97, col = "darkred", fill =
"pink2") +

theme_light() +
labs(title = "Relación entre Estatura y Peso por Género", x =
"Estatura", y = "Peso")
```



Interpreta y comenta los resultados obtenidos

Color Azul (color = "blue"):

- Representación: Puntos y línea de regresión para hombres.
- Elemento: Los puntos azules representan los datos de hombres (geom_point). La línea de regresión azul (geom_smooth) muestra la relación ajustada entre la estatura y el peso para los hombres, junto con la banda de confianza en color lightblue.

Color Rojo (color = "red"):

- Representación: Líneas de los límites inferior y superior de la banda de predicción para hombres.
- Elemento: Las líneas rojas discontinuas (geom_line con linetype = "dashed") representan los intervalos de predicción para los hombres (líneas lwr y upr).

Color Light Blue (fill = "lightblue"):

- Representación: Banda de confianza alrededor de la línea de regresión para hombres.
- Elemento: La banda de confianza sombreada alrededor de la línea de regresión de los hombres muestra el intervalo de confianza del 97%.

Color Rosa (color = "pink"):

- Representación: Puntos para las mujeres.
- Elemento: Los puntos rosados representan los datos de mujeres (geom_point).

Color Dark Red (color = "darkred"):

Representación: Líneas de los límites inferior y superior de la banda de predicción para mujeres. Elemento: Las líneas lwr y upr para las mujeres se muestran en rojo oscuro con una línea discontinua.

Color Pink2 (fill = "pink2"):

- Representación: Banda de confianza alrededor de la línea de regresión para mujeres.
- Elemento: La banda de confianza sombreada alrededor de la línea de regresión de las mujeres, mostrando el intervalo de confianza del 97%.

Color Dark Red (col = "darkred"):

- Representación: Línea de regresión ajustada para las mujeres.
- Elemento: La línea de regresión para las mujeres es de color rojo oscuro (geom_smooth), indicando la relación ajustada entre estatura y peso.