9,6

Trabajo 1

Estudiantes

Dairo García Vergara
Cristian Escobar Aguirre
Yanela Miranda Torres
Esteban Álvarez Granda

Equipo: 5

Docente

Veronica Guarin Escudero

Asignatura

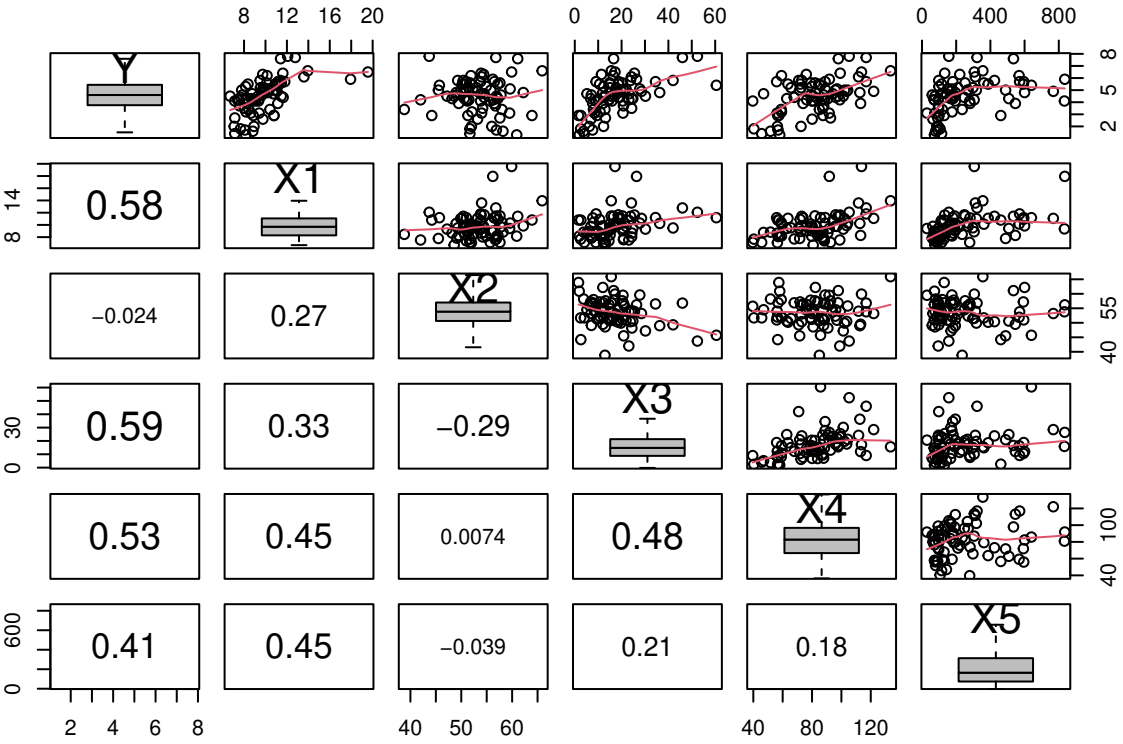
Estadística II

Sede Medellín
5 de octubre de 2023

output: pdf_document

```
##
## Please cite as:
##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

Matriz de gráficas de dispersión con boxplots y correlaciones de las variables



Con base en la gráfica, se tiene que las correlaciones entre la variable Y y las regresoras X1, X3 y X4 son mayores a 0.5, por lo tanto, se cree que pueden ser significativas para el modelo. Además, con respecto a las correlaciones entre las regresoras se puede afirmar que no hay una correlación tan alta por que ninguna supera el 0.5, por lo cual se intuye que es un modelo que no posee multicolinealidad

Modelo a estimar

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \epsilon$$

$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

Punto 1

1.1 Estimación de los parámetros del modelo

Table 1: Resultados de la Regresión

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.1496778	1.4444812	0.1036205	0.9177756
X1	0.1940306	0.0729756	2.6588421	0.0097679
X2	0.0036031	0.0279047	0.1291211	0.8976429
X3	0.0481519	0.0132931	3.6223348	0.0005580
X4	0.0133038	0.0070260	1.8935025	0.0625464
X5	0.0012473	0.0006770	1.8423822	0.0697793

1.2 Análisis de la significancia de la regresión Para ello vamos utilizar la tabla ANOVA, de donde se puede obtener el estadístico F, que nos sirve para testear la siguiente hipótesis:

H0: B1=B2=B3=B4=B5=0

H1: Bj es diferente de cero para almenos algún j desde 1 hasta 5

Análisis de varianza

```
my_Anova <- myAnova(modelo)
my_Anova
```

```
##       Sum_of_Squares DF Mean_Square F_Value     P_value
## Model        85.0784  5   17.015685 17.0219 6.99125e-11
## Error        67.9752 68    0.999636
```

Con base en la tabla ANOVA, se nota que el estadístico F, el cual se calcula como el cociente entre el minimum square regression *(MSR)* y el minimum square error *(MSE)*, tiene un valor-p bastante menor al nivel de siginificancia de 0.05. Por lo tanto, se rechaza la hipóteis nula y se conluye que por lo menos algún parámetro es difente de cero y el modelo es siginificativo.

1.3 Análisis de siginificancia de los parámetros individuales En este caso nos basamos en el estadístico t de cada parámetro para probar la siguiente hipótesis:

H0:Bj=0 para j entre 0 y 5

H1:Bj es diferente de cero con j que va desde 0 hasta 5

En la columna 4 de la tabla 1 se presentan los estadísticos t de cada parámetro. A su vez, en la columna 5 está el valor p para cada estadítico. Con base en estos, se puede decir que, con un nivel de significancia α= 0.05, los parámetros *B1 gorro* y *B3 gorro* son estadísticamente significativos por tener un valor-p asociado menor al dicho nivel de significancia. Bajo el criterio del valor p, *B4 gorro* no resulta estadísticamente significativo por tener un valor p mayor a 0.05, contrariando la intuición que se obtuvo en la gráfica de correlaciones.

Interpretación: Si se da un aumento en 1 día de la estadía promedio de todos los pacientes en el hospital, la probabiliad promedio de obtener una infección aumenta, en promedio, 0.19 pp manteniendo las demás variables constantes. Por otro lado, si el se da un aumento en una unidad en el número promedio de camas en el hospital, la probabilidad de obtener una infección aumenta, en promedio, 0.04 pp manteniendo las demás variables constantes

1.4 Cálculo e interpretación del R^2 El R cuadrado se puede obtener a partir de la identidad de sumas de cuadrados:

SST = SSR + SSE

De donde R^2 es:

R^2 = SSR/SST= SSR/(SSR+SSE) = 85.0784/(85.0784+67.9752) = 0.5558732

Verificando con código:

```
R_sq <- summary(modelo)$r.squared
R_sq
```

```
## [1] 0.5558732
```

La interpretación de este valor es la siguiente: el 55.58% de la variabilidad total de la la probabilidad promedio de adquirir infección es explicada por el modelo de regresión estimado.

Ahora, una medida de bondad de ajuste más rigurosa es el R^2 ajustado, el cual se calcula de la siguiente forma

R^2adj = 1 - (n-1)MSE/SST = 1 - (74-1)0.9996/153.0536 = 0.5232

Con código

```
R_2adj <- summary(modelo)$adj.r.squared
R_2adj
```

```
## [1] 0.5232169
```

Como vemos el valor ajustado, al ser menor que el R^2, indica que en el modelo hay variables incluidas que no son significativas. El valor ajustado lo que hace es penalizar la inclusión de variables no relevantes.

Punto 2

Con base en la estimación del modelo, se puede notar que los 3 parámetros con el valor-p más pequeño son *B1*, *B3* y *B4*. Con ello claro, es posible plantear una prueba de significancia para ese conjunto de coeficientes. Dicha prueba parte de la siguientes hipótesis:

H0: B1=B3=B4=0

H1: Bj diferente de cero para algún j=1,3 ó 4

Para testear esta hipótesis se puede partir de la suma de cuadrados extra *(SSextra)*, que para la hipóteis planteada toma la siguente forma:

SSextra(B1, B3, B4|B0, B2, B5) = SSR(B0, B1, B2, B3, B4, B5) - SSR(B0, B2, B5)

Esta suma de cuadrados se puede expresar de forma alterna mediante las diferencia de sumas de cuadrados del error:

SSextra(B1, B3, B4|B0, B2, B5)= SSE(B0, B2, B5) - SSE(B0, B1, B2, B3, B4, B5)

Para conocer el valor de la SSextra debemos conocer las SSE para el modelo restrigido y para el modelo no restringido. Esto último lo podemos hacer mediante el conocimiento de todas las regresiones posibles:

```
Allreg <- myAllRegTable(modelo)
Allreg
```

```
##    k  R_sq adj_R_sq     SSE     Cp Variables_in_model
## 1  1 0.346    0.337 100.158 30.194                 X3
## 2  1 0.338    0.329 101.283 31.320                 X1
## 3  1 0.278    0.268 110.571 40.611                 X4
## 4  1 0.171    0.159 126.895 56.941                 X5
## 5  1 0.001   -0.013 152.965 83.021                 X2
## 6  2 0.513    0.499  74.547  6.574              X1 X3
## 7  2 0.434    0.418  86.641 18.673              X3 X5
## 8  2 0.425    0.409  87.989 20.021              X1 X4
## 9  2 0.424    0.408  88.188 20.221              X3 X4
```

```
## 10 2 0.383    0.366  94.371 26.405             X4 X5
## 11 2 0.372    0.355  96.045 28.080             X1 X2
## 12 2 0.370    0.352  96.438 28.473             X2 X3
## 13 2 0.366    0.348  97.007 29.042             X1 X5
## 14 2 0.278    0.258 110.451 42.491             X2 X4
## 15 2 0.171    0.148 126.886 58.932             X2 X5
## 16 3 0.533    0.513  71.405  5.431          X1 X3 X4
## 17 3 0.532    0.512  71.592  5.618          X1 X3 X5
## 18 3 0.513    0.492  74.532  8.559          X1 X2 X3
## 19 3 0.498    0.477  76.764 10.792          X3 X4 X5
## 20 3 0.457    0.434  83.120 17.150          X1 X4 X5
## 21 3 0.456    0.433  83.246 17.276          X2 X3 X5
## 22 3 0.447    0.423  84.662 18.693          X1 X2 X4
## 23 3 0.436    0.411  86.396 20.428          X2 X3 X4
## 24 3 0.391    0.365  93.255 27.289          X1 X2 X5
## 25 3 0.384    0.357  94.338 28.373          X2 X4 X5
## 26 4 0.556    0.530  67.992  4.017       X1 X3 X4 X5
## 27 4 0.534    0.507  71.368  7.394       X1 X2 X3 X4
## 28 4 0.532    0.505  71.559  7.585       X1 X2 X3 X5
## 29 4 0.510    0.481  75.042 11.069       X2 X3 X4 X5
## 30 4 0.470    0.439  81.092 17.121       X1 X2 X4 X5
## 31 5 0.556    0.523  67.975  6.000    X1 X2 X3 X4 X5
```

Sólo muestra datos de interés

Con base en el resultado de la tabla de todas las regresiones posibles se puede notar que la *SSE(B0, B2, B5)=126.886* y que la *SSE* del modelo completo es, como ya se había visto en la ANOVA, *67.975*. De esa forma la *SSEextra* es:

SSextra = 126.886 - 67.975 = 57.911

La *SSE(B0, B2, B5)* tiene asociados *n-3 = 74-3 = 71* grados de libertad. Por su parte, la *SSE(B0, B1, B2, B3, B4, B5)* tiene *n-6 = 74-6 = 68* grados de libertad. Con base en estos valores se puede decir que la *SSextra* tiene *71-68 = 3* grados de libertad.

Dividiendo la SSextra entre sus grados de libertad se obtiene el Cuadrado Medio Extra *(MSextra)*:

MSextra = SSextra/gl = 57.911/3 = 19.30367

El estadístico de prueba, que se distribuye como una *f*, se obtiene del cociente entre el *MSextra* y el *MSE* del modelo:

F0 = MSextra/MSE = 19.30367/0.999636 = 19.2107

La hipótesis nula sobre la significancia conjunta se rechaza si *F0* > al cuantil *1-0.05 = 95* bajo una distribución con 3 gl en el numerador y 68 gl en el denominador

Veamos el valor del percentil:

→ lo calculaste mal

```
percentil <- qf(0.05, 3, 68, lower.tail = TRUE, log.p = FALSE)

percentil
```

↳ FALSE

```
## [1] 0.1167311
```

Al ser menor al estadístico *F0*, este cae en la zona de rechazo. También podemos usar el valor p del estadístico *F0*:

```
valor_p <- pf(19.2107, 3, 68, lower.tail = FALSE, log.p = FALSE)

valor_p
```


1 pt

```
## [1] 3.951378e-09
```

Al ser un valor menor al nivel de significancia de 0.05 se convalida el rechazo a la hipóteis nula. Lo anterior quiere decir que por lo menos una de las variables puestas a prueba mediante sus coeficientes, explican a la variable respuesta de una forma significativa.

Punto 3

¿Al aumentar en una unidad en el número de pacientes en el hospital por día durante el periodo del estudio (X4), la probabilidad promedio estimada de adquirir infección en el hospital (en porcentaje) incrementará en la misma medida,que si aumentará en una unidad el número promedio de enfermeras presentes a tiempo completo en el hospital, durante el período del estudio?

- ¿Al aumentar en una unidad la razón de número de cultivos realizados sin síntomas de infección hospitalaria por cada 100 individuos(X2), afectara significativamente la probabilidad promedio estimada de adquirir infección en el hospital (porcentaje)?

$H_0 : \beta_4 = \beta_5, \beta_2 = 0 VS H_1 : \beta_4 \neq \beta_5, \beta_2 \neq 0$ Podemos reescribir la hipotesis nula de la siguiente manera:

$$H_0 : \beta_4 - \beta_5 = 0, \beta_2 = 0$$

Notemos que a partir de lo anterior, se puede proponer una hipótesis nula que contiene m=2 ecuaciones, de la en la forma H0 : Lβ = 0, de manera que se tiene una prueba de hipótesis lineal general de la forma:

$$\begin{cases} Ho : L\underline{\beta} \\ H1 : L\underline{\beta} \end{cases}$$

En ese sentido, veamos Ho con un sistema de dos ecuaciones:

$$Ho : \begin{cases} \beta_4 - \beta_5 = 0 \\ \beta_2 = 0 \end{cases}$$

En forma matricial se puede expresar como: $H_o : \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Por lo tanto, se tiene una prueba de hipotesis lineal general con:

$$L : \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Notemos que la matriz L tiene r=2 filas linealmente independientes(ninguna de las 2 filas puede escribirse como múltiplo escalar de la otra)

Por lo tanto el modelo bajo H_0 es: $Y = \beta_0 + \beta_1 X_{i_1} + \beta_2 X_{i_2} + \beta_3 X_{i_3} + \beta_4 X_{i_4} + \beta_5 X_{i_5} + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$

RM: $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_4 (X_{i_4} + X_{i5}) + \varepsilon_i \quad \varepsilon_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right)$
$= \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_4 X_{4,5} + \varepsilon_i$

Donde

$$X_4, 5 = X_i 4 + X_i 5$$

Finalmente la forma del estadístico de prueba tiene la forma: $F_0 = \dfrac{\text{MSH}}{\text{MSE}} = \dfrac{\text{SSH}/2}{\text{MSE}} = \dfrac{[\text{SSE(RM)}^* - \text{SSE(FM)}]/2}{\text{MSE}} = \dfrac{[\text{SSE(RM)}^* - 67.9752]/2}{0.999636}$

Punto 4

20 pt

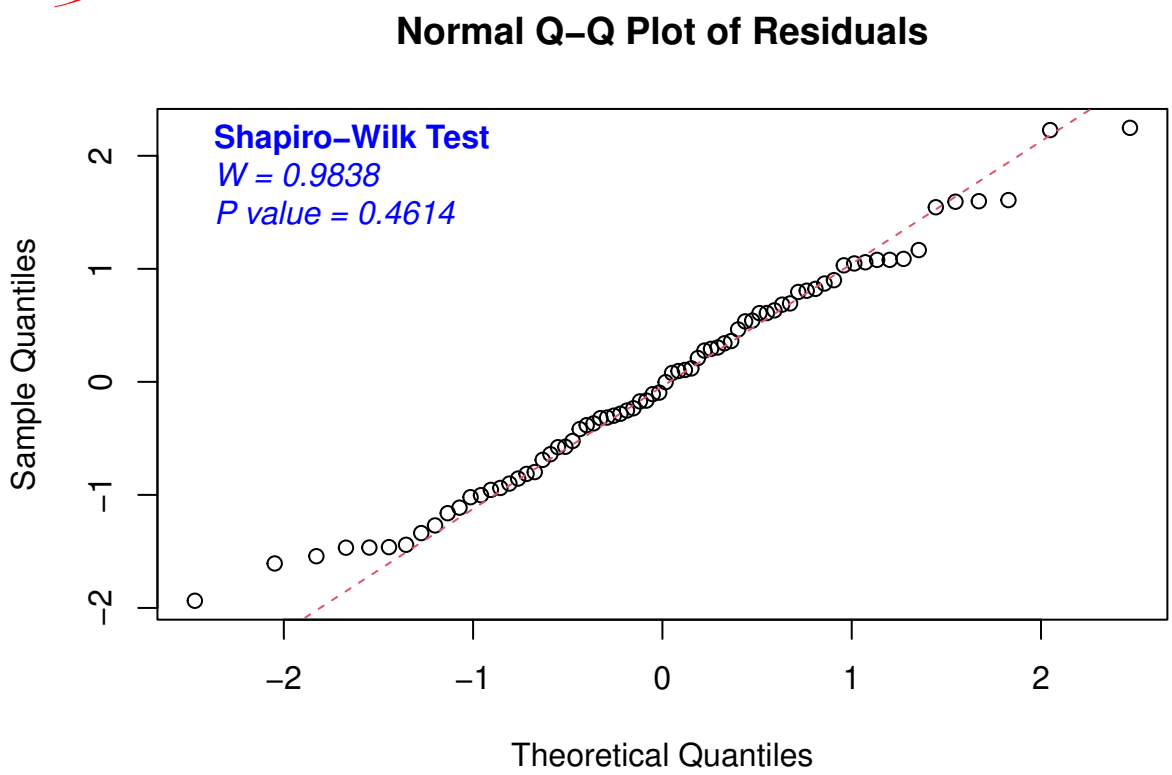
4.1 Prueba de normalidad de los errores Para analizar el supuesto de la normalidad de los errores se usan sus estimaciones, es decir, los residuales. Para la prueba de normalidad se parte de la siguentes hipótesis:

H0: Los errores se ditribuyen como una normal

H1: Los errores no se distribuyen como una normal

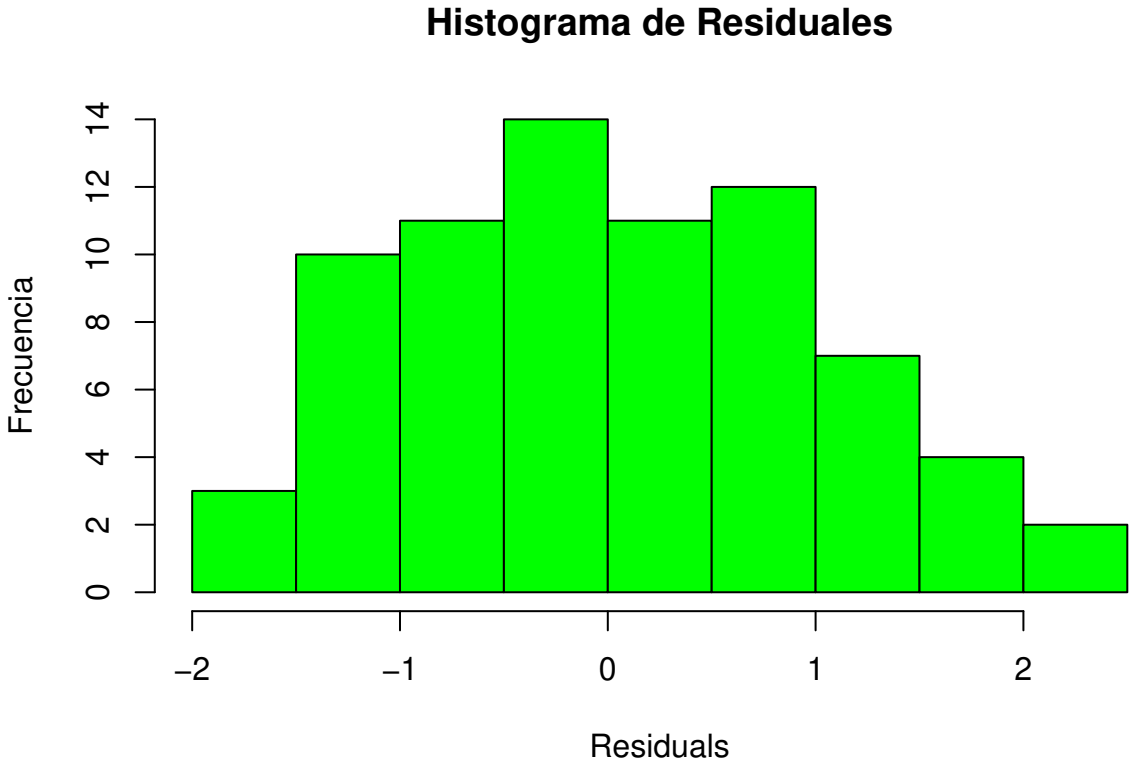
Para testear estas hipótesis se utilizan dos enfoques: *i)* El enfoque gráfico mediante la *Q-Q Plot* y un *histograma. ii)* el enfoque teórico mediante la prueba de *Shapiro-Wilk.*

```
myQQnorm(modelo)
```




```
datos <- modelo$residuals

hist(datos,
     main = "Histograma de Residuales",
     xlab = "Residuals",
     ylab = "Frecuencia",
     col = "green",
     border = "black",
     breaks = 10)
```



El análisis gráfico resulta bastante subjetivo. Con ello, se nota una percepción débil de normalidad ya que la gráfica *Q-Q plot* presenta unas desviaciones en los extremos a la línea diagonal. Además, el *histograma* presenta una forma de campana pero más bien amplia. A pesar de que la prueba teórica de *Shapiro-Wilk* indica un *valor-p* bastante mayor a cualquier nivel de significancia aceptable para una prueba de hipótesis, se rechaza el supuesto de normalidad ya que se le da mayor ponderación a la gráfica.

4.2 Prueba de varianza constante. Para analizar la varianza se parte de las siguietes hipótesis:

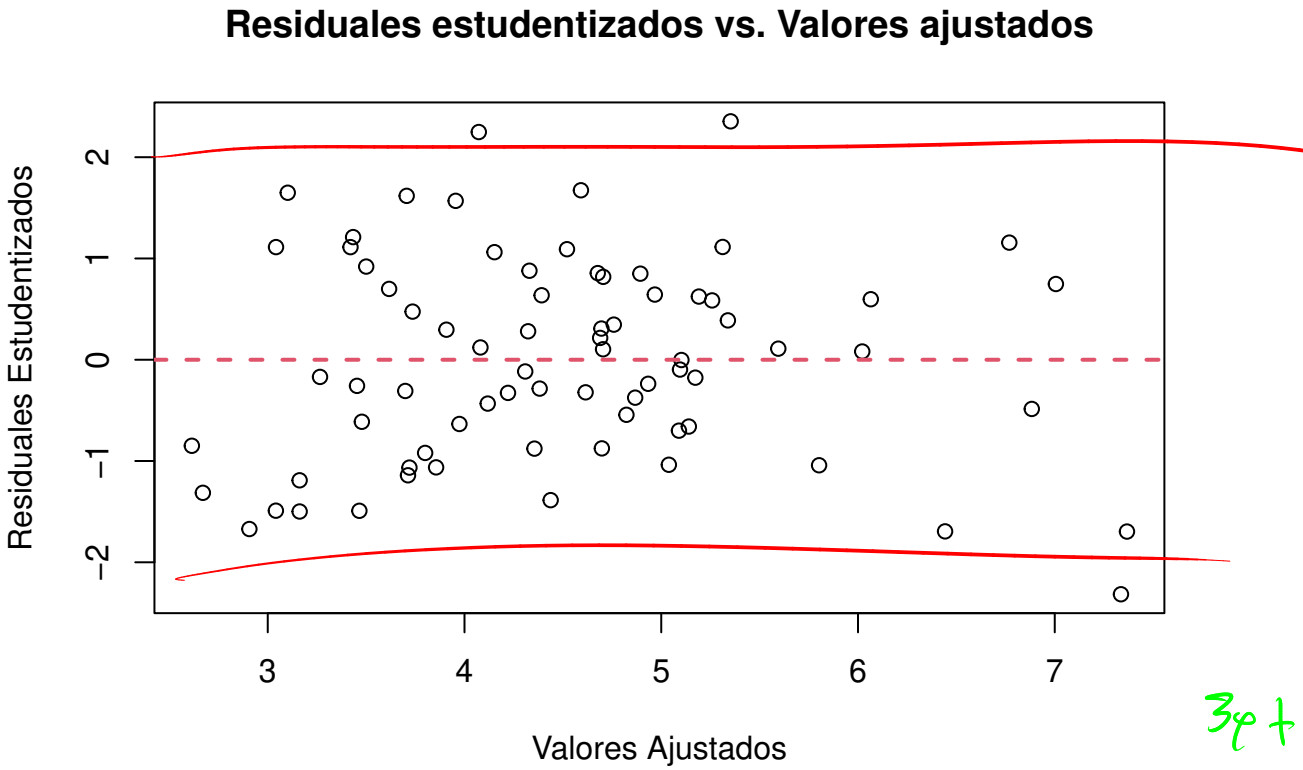
H0: Los errores tienen varianza constante

H1: Los errores no tienen varianza constante

Para testear las hipótesis se utiliza la gráfica de los residuales contra los valores ajustados de la variable independiente.

```
# Cálculo de residuales estudentizados y valores ajustados
res.stud <- round(rstandard(modelo), 4)
yhat <- round(modelo$fitted.values, 4)

# Gráfica de Residuales estudentizados vs. Valores ajustados
plot(yhat, res.stud, xlab = "Valores Ajustados", ylab = "Residuales Estudentizados", main = "Residuales
abline(h = 0, lty = 2, lwd = 2, col = 2)
```



Basados en la gráfica, se concluye que el supuesto de varianza constante se cumple ya que si se traza una línea por arriba y por debajo de los datos se puede evidenciar que hay ciertos puntos que salen de esta pero como no hay evidencia fuerte que dañe o viole el supuesto de varianza constante y basados en la gráfica se logra observar cómo la gran mayoría de los datos aún siguen manteniendo la forma de un rectángulo. Por tal razón, se concluye que el supuesto de varianza constante se cumple, ya que, no se perciben tendencias significativamente marcadas que conduzcan a concluir lo contrario.

Punto 4.3 Valores atípico, de balanceo e influenciales Para verificar si existen valores extremos analizamos a los residuales estudentizados, los valores de la diagonal de la matriz *hat* y los diagnósticos.

```
## Diagnósticos para identificar valores extremos

# Cálculo de errores estándar de los valores ajustados
se.yhat <- round(predict(modelo, se.fit = T)$se.fit, 4)
# Residuales crudos del modelo
residuals <- round(modelo$residuals, 4)
# Distancias de Cook
Cooks.D <- round(cooks.distance(modelo), 4)
# Valores de la diagonal de la matriz H
hii.value <- round(hatvalues(modelo), 4)
# Dffits
Dffits <- round(dffits(modelo), 4)
# Tabla de diagnósticos
data.frame(base, yhat, se.yhat, residuals, res.stud, Cooks.D, hii.value, Dffits)
```

```
##      Y    X1   X2   X3    X4  X5   yhat se.yhat residuals res.stud Cooks.D
## 1  4.5  6.70 48.6 13.0  80.8  76 3.4205  0.2389    1.0795   1.1119  0.0125
## 2  2.0  7.08 52.0 12.3  56.4  87 3.1619  0.2170   -1.1619  -1.1905  0.0117
## 3  6.1 13.59 54.0 24.2 111.7 312 6.0216  0.2625    0.0784   0.0813  0.0001
## 4  5.4 11.18 45.7 60.5  85.8 640 7.3365  0.5480   -1.9365  -2.3157  0.3837
```

```
## 5  4.4  7.70 56.9 12.2  67.9 129 3.5004  0.2084    0.8996   0.9200  0.0064
## 6  5.0 11.03 49.9 19.7 102.1 318 5.1732  0.2052   -0.1732  -0.1770  0.0002
## 7  6.4 11.62 53.9 25.5  99.2 133 5.3120  0.2125    1.0880   1.1136  0.0098
## 8  5.1  9.76 50.9 21.9  97.0 150 4.7589  0.1773    0.3411   0.3467  0.0007
## 9  2.7  7.14 57.6 13.1  92.6  92 3.7201  0.2863   -1.0201  -1.0649  0.0169
## 10 5.9 17.94 56.2 26.4  91.8 835 7.3671  0.5019   -1.4671  -1.6966  0.1616
## 11 5.6 11.48 57.6 20.3  82.0 252 4.9674  0.1819    0.6326   0.6434  0.0024
## 12 5.5  8.37 50.7 15.1  84.8 115 3.9551  0.1735    1.5449   1.5690  0.0127
## 13 3.9 11.15 56.5  7.7  73.9 281 4.2211  0.1944   -0.3211  -0.3274  0.0007
## 14 3.7  7.58 56.7 20.8  88.0  97 4.1180  0.2590   -0.4180  -0.4329  0.0022
## 15 5.5 11.08 50.2 18.6  63.6 387 4.7049  0.2368    0.7951   0.8186  0.0066
## 16 4.1  9.35 53.8 15.9  80.9 833 5.0386  0.4248   -0.9386  -1.0370  0.0395
## 17 4.7 10.72 53.8 23.2  94.1 113 4.9335  0.1844   -0.2335  -0.2376  0.0003
## 18 6.6 13.95 65.9 15.6 133.5 356 6.0651  0.4447    0.5349   0.5973  0.0147
## 19 2.9  8.86 51.3  9.5  87.5 100 3.7999  0.2082   -0.8999  -0.9202  0.0064
## 20 3.2  8.19 52.1 10.8  59.2 176 3.4537  0.1820   -0.2537  -0.2580  0.0004
## 21 5.1 10.30 59.6 27.8  88.9 175 5.1025  0.2697   -0.0025  -0.0026  0.0000
## 22 4.3  9.42 50.6 24.8  62.8 508 4.8230  0.2726   -0.5230  -0.5437  0.0040
## 23 2.6  9.76 53.2  6.9  80.1  64 3.7128  0.2202   -1.1128  -1.1410  0.0111
## 24 5.2  9.53 51.5 15.0  65.7 298 4.1524  0.1643    1.0476   1.0622  0.0052
## 25 3.4 10.42 58.0  8.0  59.0 119 3.6990  0.2433   -0.2990  -0.3083  0.0010
## 26 7.8 12.07 43.7 52.4 105.3 157 6.7690  0.4525    1.0310   1.1564  0.0574
## 27 5.0  9.78 52.3 17.6  95.9 270 4.6958  0.1552    0.3042   0.3080  0.0004
## 28 3.1  8.63 54.0  8.4  56.2  76 3.2657  0.2121   -0.1657  -0.1696  0.0002
## 29 2.0  8.93 56.0  6.2  72.5  95 3.4657  0.1844   -1.4657  -1.4916  0.0131
## 30 5.7 11.80 53.8  9.1 116.9 571 5.3387  0.3680    0.3613   0.3886  0.0039
## 31 7.6 11.41 61.1 16.6  97.9 535 5.3528  0.2971    2.2472   2.3540  0.0895
## 32 4.1  9.05 51.2 20.5  79.8 195 4.3821  0.1397   -0.2821  -0.2850  0.0003
## 33 5.8  9.50 49.3 42.0  70.9  98 5.2585  0.3812    0.5415   0.5859  0.0097
## 34 5.0  7.78 45.5 20.9  71.6 489 4.3920  0.2966    0.6080   0.6367  0.0065
## 35 4.2  7.39 51.0 14.6  88.4  72 3.7362  0.2170    0.4638   0.4752  0.0019
## 36 6.5 19.56 59.9 17.2 113.7 306 6.8833  0.6144   -0.3833  -0.4859  0.0239
## 37 4.3  8.30 57.2  6.8  83.8 167 3.6168  0.2117    0.6832   0.6992  0.0038
## 38 4.8  9.84 62.2 12.0  82.3 600 4.7042  0.3725    0.0958   0.1033  0.0003
## 39 2.9 10.79 44.2  2.6  56.6 461 3.8557  0.4371   -0.9557  -1.0629  0.0445
## 40 1.6  8.82 58.2  3.8  51.7  80 3.0413  0.2538   -1.4413  -1.4904  0.0255
## 41 5.3  8.15 54.9 12.3  79.8  99 3.7062  0.1724    1.5938   1.6183  0.0134
## 42 4.2  7.53 42.0 23.1  98.9  95 4.3086  0.3481   -0.1086  -0.1159  0.0003
## 43 3.1  9.41 59.5 20.6  91.7  29 4.4379  0.2616   -1.3379  -1.3865  0.0235
## 44 4.1  7.13 55.7  9.0  39.6 279 3.0420  0.3078    1.0580   1.1122  0.0216
## 45 4.4 11.65 54.5 18.6  96.1 248 5.0900  0.1698   -0.6900  -0.7003  0.0024
## 46 3.4  8.45 38.8 12.9  85.0 235 3.9741  0.4254   -0.5741  -0.6345  0.0148
## 47 6.3  9.74 54.4 11.4  76.1 221 4.0726  0.1352    2.2274   2.2485  0.0157
## 48 3.5  8.03 54.2 24.3  87.3  97 4.3555  0.2231   -0.8555  -0.8778  0.0067
## 49 4.2  9.00 56.3 14.6  76.4  72 3.9080  0.1758    0.2920   0.2966  0.0005
## 50 5.2  9.84 53.0 17.7  72.6 210 4.3300  0.1403    0.8700   0.8789  0.0026
## 51 4.9  9.89 50.5 17.7 103.6 167 4.6895  0.2174    0.2105   0.2157  0.0004
## 52 5.8 11.41 50.4 23.8  73.0 424 5.1912  0.2175    0.6088   0.6238  0.0032
## 53 4.7  8.77 54.5  5.2  47.0 143 3.4017  0.2458    1.5983   1.6492  0.0291
## 54 4.5  9.31 47.2 30.2 101.3 170 5.1401  0.2416   -0.6401  -0.6597  0.0045
## 55 4.2  8.88 51.5 10.1  86.9 305 4.0811  0.1910    0.1189   0.1212  0.0001
## 56 4.5 11.46 56.9 15.6  97.7 191 4.8675  0.1933   -0.3675  -0.3746  0.0009
## 57 2.9 10.80 63.9  1.6  57.4 130 3.4783  0.3297   -0.5783  -0.6127  0.0076
## 58 4.9 11.07 53.2 28.5 122.0 768 6.4426  0.4137   -1.5426  -1.6948  0.0989
```

```
## 59 4.6  7.84 49.1  7.1  87.9  60 3.4339  0.2690    1.1661   1.2110  0.0191
## 60 4.8 10.24 49.0 36.3 112.6 195 5.8022  0.2731   -1.0022  -1.0421  0.0146
## 61 6.2 10.15 51.9 16.4  59.2 568 4.5918  0.2760    1.6082   1.6735  0.0385
## 62 5.7 11.20 56.5 34.5  88.9 180 5.5949  0.2771    0.1051   0.1094  0.0002
## 63 1.4  7.14 51.7  4.1  45.7 115 2.6702  0.2553   -1.2702  -1.3140  0.0201
## 64 4.6  9.68 57.8 16.7  79.0 186 4.3233  0.1708    0.2767   0.2809  0.0004
## 65 5.0 10.33 55.8 21.2 104.3 266 5.0953  0.1900   -0.0953  -0.0970  0.0001
## 66 1.8  7.67 51.7  2.5  40.4 106 2.6142  0.2846   -0.8142  -0.8495  0.0106
## 67 5.5 10.90 57.2 10.6  71.9 593 4.6773  0.2735    0.8227   0.8555  0.0099
## 68 1.7  8.09 56.9  7.6  56.9  92 3.1621  0.2187   -1.4621  -1.4987  0.0188
## 69 4.3  8.67 48.2 24.4  90.8 182 4.6155  0.1954   -0.3155  -0.3218  0.0007
## 70 5.6 10.12 51.7 14.9  79.1 362 4.5209  0.1491    1.0791   1.0915  0.0045
## 71 3.9  8.28 49.5 12.0 113.1 546 4.6981  0.4101   -0.7981  -0.8753  0.0258
## 72 5.7 11.18 51.0 18.8  55.9 595 4.8938  0.3159    0.8062   0.8499  0.0133
## 73 1.3  8.16 60.9  1.9  58.0  73 2.9066  0.2758   -1.6066  -1.6717  0.0384
## 74 7.7 12.78 56.8 46.0 116.9 322 7.0059  0.3715    0.6941   0.7478  0.0149
##    hii.value  Dffits
## 1     0.0571  0.2741
## 2     0.0471 -0.2655
## 3     0.0689  0.0220
## 4     0.3004 -1.5693
## 5     0.0435  0.1959
## 6     0.0421 -0.0368
## 7     0.0452  0.2427
## 8     0.0315  0.0621
## 9     0.0820 -0.3186
## 10    0.2520 -0.9988
## 11    0.0331  0.1185
## 12    0.0301  0.2796
## 13    0.0378 -0.0645
## 14    0.0671 -0.1154
## 15    0.0561  0.1991
## 16    0.1805 -0.4869
## 17    0.0340 -0.0443
## 18    0.1978  0.2952
## 19    0.0434 -0.1957
## 20    0.0331 -0.0474
## 21    0.0728 -0.0007
## 22    0.0743 -0.1533
## 23    0.0485 -0.2582
## 24    0.0270  0.1771
## 25    0.0592 -0.0768
## 26    0.2048  0.5884
## 27    0.0241  0.0481
## 28    0.0450 -0.0365
## 29    0.0340 -0.2825
## 30    0.1355  0.1529
## 31    0.0883  0.7589
## 32    0.0195 -0.0399
## 33    0.1454  0.2404
## 34    0.0880  0.1969
## 35    0.0471  0.1051
## 36    0.3776 -0.3763
## 37    0.0448  0.1509
```

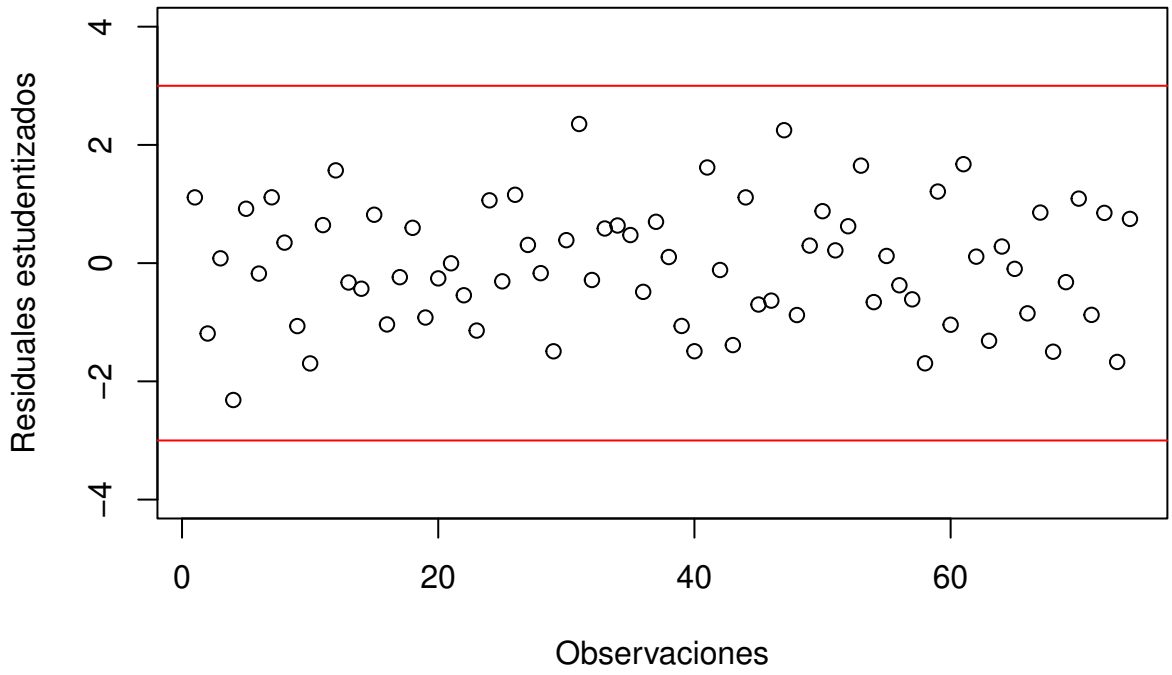
```
## 38    0.1388  0.0412
## 39    0.1912 -0.5172
## 40    0.0644 -0.3947
## 41    0.0297  0.2868
## 42    0.1212 -0.0427
## 43    0.0684 -0.3784
## 44    0.0948  0.3605
## 45    0.0288 -0.1202
## 46    0.1810 -0.2970
## 47    0.0183  0.3167
## 48    0.0498 -0.2006
## 49    0.0309  0.0526
## 50    0.0197  0.1243
## 51    0.0473  0.0477
## 52    0.0473  0.1384
## 53    0.0604  0.4237
## 54    0.0584 -0.1636
## 55    0.0365  0.0234
## 56    0.0374 -0.0733
## 57    0.1088 -0.2130
## 58    0.1712 -0.7813
## 59    0.0724  0.3395
## 60    0.0746 -0.2961
## 61    0.0762  0.4873
## 62    0.0768  0.0313
## 63    0.0652 -0.3490
## 64    0.0292  0.0484
## 65    0.0361 -0.0186
## 66    0.0810 -0.2517
## 67    0.0748  0.2428
## 68    0.0479 -0.3392
## 69    0.0382 -0.0637
## 70    0.0222  0.1648
## 71    0.1683 -0.3930
## 72    0.0998  0.2824
## 73    0.0761 -0.4863
## 74    0.1381  0.2983
```

Sólo datos importantes.

Analizando la columna de los residuales estudentizados se nota que ninguno supera el valor de 3 en valor absoluto. Por lo tanto, se puede conluir que no hay observaciones atípicas desde el enfoque de las Y. ✓ 3 o +

Gráficamente también podemos comprobar que ningún residual estudentizado supera en valor absoluto el valor de 3:

```
plot(res.stud, xlab = "Observaciones", ylab = "Residuales estudentizados", ylim = c(-4, 4))
 abline(h = 3, col = "red", lty = 1)
 abline(h = -3, col = "red", lty = 1)
```



En el caso de los puntos de balanceo se analizan los valores de la diagonal de la matriz *hat*. El criterio para afirmar la existencia de un punto de balanceo es:

$hii > 2p/n$

Para los datos de este caso 2p/n = 2(6)/74 = 0.1621.

Tabla de puntos de balanceo

Observación	hii
10	0.2522
16	0.1805
18	0.1978
26	0.2848
36	0.3776
39	0.1912
46	0.1810
58	0.1712
71	0.1683

Como se pude observar, nuestros datos arrojan 9 puntos de balanceo ya que sus valores correspondietes en la matriz *hat* superan 0.1621. Los puntos de balanceo, si bien no necesariamente alteran los parámetros estimados, sí afectan valores de resumen de la regresión que son útiles para hacer conclusiones.

Para determinar si existen observaciones influenciales se recurre a los siguientes dos criterios:

- Distancia de Cook: Este criterio indica que una observación es influencial si $D\_\{i\} > 1$
- Diagnóstico DFFITS: Indica que una observación es influencial si $|DFFITS\_\{j(i)\}| > 2\sqrt{p/n}$. Para nuestro caso $2\sqrt{p/n}$ es 0.5694.

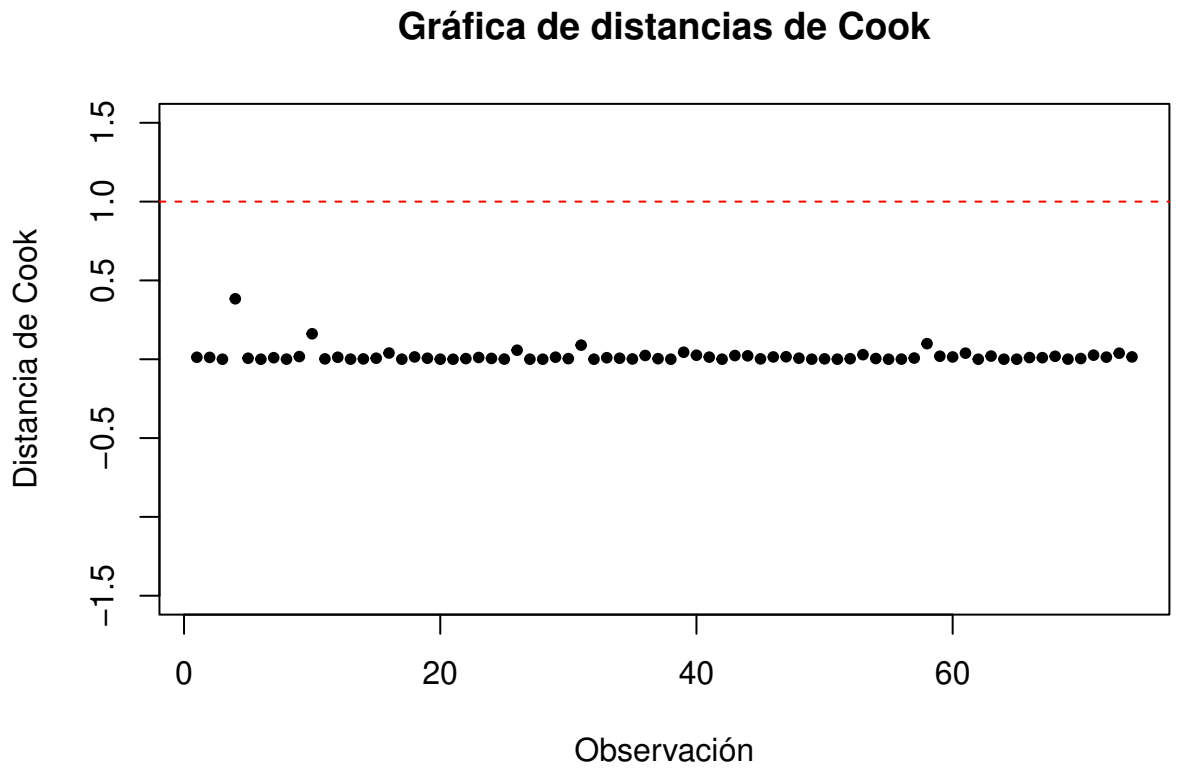
Según lo anterior se puede concluir que, basándonos en el criterio de Cook no hay observaciones influenciales. Si se tiene en cuenta el diagnóstico de DFFITS hay 5 observaciones influenciales que se muestran a continuación.

Tabla de observaciones influenciales

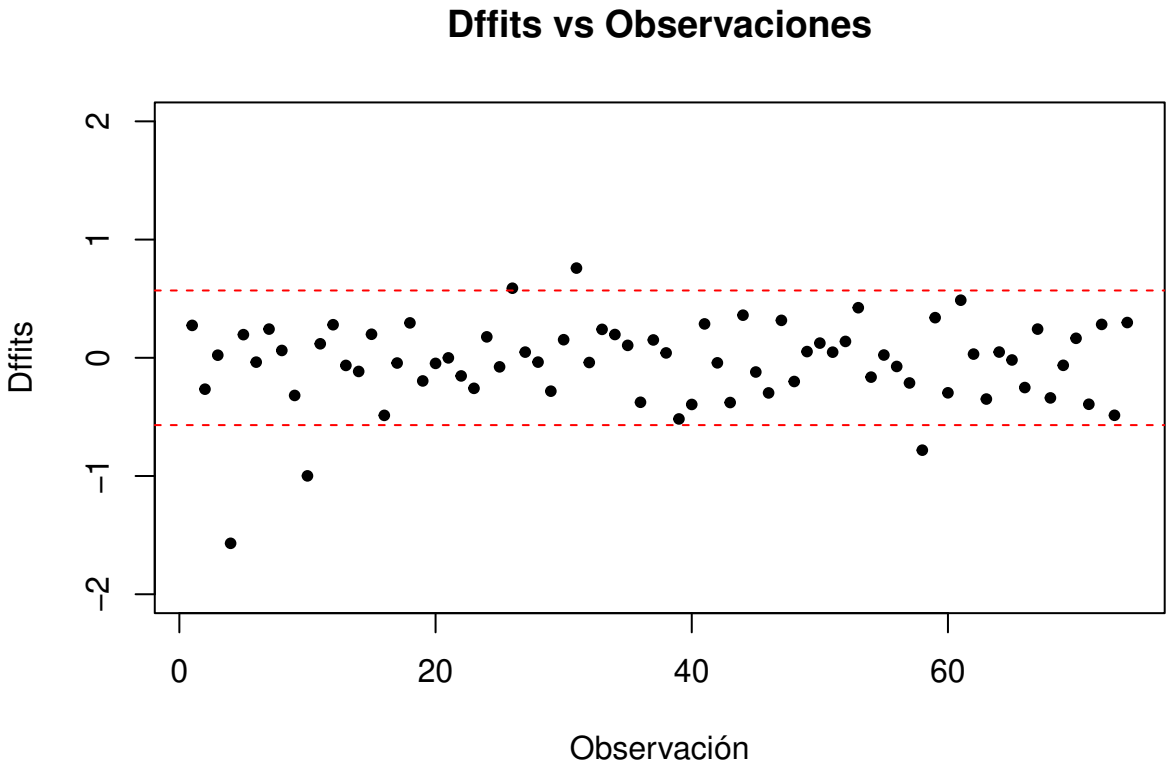
Observación	DFFITS
4	-1.5693
10	-0.9988
26	0.5884
31	0.7589
58	-0.7813

Gráficas sobre los criterios de observaciones influenciales.

```
plot(Cooks.D, xlab="Observación", ylab = "Distancia de Cook",
     main = "Gráfica de distancias de Cook", ylim=c(-1.5, 1.5), pch = 20)
abline(h = 1, col="red", lty = "dashed")
```




```
plot(Dffits, xlab="Observación", ylab = "Dffits",
     main = "Dffits vs Observaciones", ylim=c(-2, 2), pch = 20)
abline(h = 0.5694, col="red", lty = "dashed")
abline(h = -0.5694, col="red", lty = "dashed")
```



Conclusión Analizando los datos del modelo se evidenció que no hay relación gráfica entre las variables regresoras. Además, se observa que el modelo es signicativo, esto debido a las pruebas de significancia tanto individual como del modelo, donde hallamos que hay 2 parámetros significativos ($/beta_1$ y $/beta_3$) para un nivel de significancia de 0.05. Al analizar los supuestos de los errores, se encontró que el supuesto de varianza constante se cumple, pero el supuesto de normalidad no, lo cual nos lleva a afirmar que el modelo no es valido. Además, el conjunto de datos presenta observaciones influenciales, las cuales, al afectar el valor de los parámtros estimados, pueden desviar los efectos parciales de la tendencia estructural.

Para identificar los valores que pueden alterar el modelo, se emplearon los criterios para encontrar las observaciones extremas, en los cuales se obtuvieron los siguientes resultados: - Ninguna de las observaciones es atípica - Las observaciones 10, 16, 18, 26, 36, 39, 46, 58 y 71 son puntos de balanceo: Entonces, estas observaciones representan puntos en el conjunto de datos de las variables independientes que están claramente distantes del resto de la muestra y podrían tener un impacto en los coeficientes estimados, influyendo potencialmente en el coeficiente de determinación (R^2) y los errores estándar de esos coeficientes.

- Las observaciones 4, 10, 26, 31 y 58 son influenciales: Por lo tanto, estas observaciones ejercen una influencia significativa en los coeficientes de regresión ajustados. En otras palabras, estas observaciones tiran del modelo en su dirección.

Claramente, se ha identificado la existencia de observaciones extremas que requerirán un análisis previo antes de utilizar el modelo, con el fin de volver a valorar su idoneidad como herramienta predictiva o estimadora de los valores de respuesta

3pt