

Trabajo 1

Estudiantes

Integrante 1

Integrante 2

Integrante 3

Integrante 4

Docente

Raúl Alberto Pérez Agamez

Asignatura

Estadística II



Sede Medellín
mayo de 2022

Índice

1. Pregunta 1	3
2. Pregunta 2	3
3. Pregunta 3	4
4. Pregunta 4	5
4.1. Supuestos del modelo	5
4.1.1. Normalidad de los residuales	5
4.2. Verificación de las observaciones	6
4.2.1. Datos atípicos	6
4.2.2. Con los hii	7

Índice de figuras

1.	Normalidad de los residuales	5
2.	Identificación de datos atípicos	6

Índice de cuadros

1.	Tabla ANOVA para el modelo	3
2.	Resumen de los coeficientes	3
3.	Resumen tabla todas regresiones	4

1. Pregunta 1

Con la base de datos brindada, se ajusta el siguiente modelo de regresión.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i, \quad \varepsilon_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2); \quad 1 \leq i \leq 50$$

Luego, el modelo ajustado fue

$$\hat{Y}_i = 18.6908 + 0.5975X_{1i} - 0.0442X_{2i} + 0.2373X_{3i} - 0.1654X_{4i}$$

Con respecto a la significancia de la regresión, se presenta la tabla ANOVA.

Cuadro 1: Tabla ANOVA para el modelo

	Sumas de cuadrados	gl	Cuadrado medio	F_0	P-valor
Regresión	3640.11	4	910.0279	17.4078	1.06986e-08
Error	2352.47	45	52.2771		

De la tabla ANOVA, se ve que la regresión es significativa.

Ahora, respecto a los parámetros individuales.

Cuadro 2: Resumen de los coeficientes

	$\hat{\beta}_j$	$SE(\hat{\beta}_j)$	T_{0j}	P-valor
β_0	18.6908	8.6708	2.1556	0.0365
β_1	0.5975	0.0743	8.0360	0.0000
β_2	-0.0442	0.0679	-0.6506	0.5186
β_3	0.2373	0.0667	3.5584	0.0009
β_4	-0.1654	0.0872	-1.8963	0.0643

Se deja como tarea al lector la interpretación de los coeficientes de regresión.

El modelo tiene un $R^2 = 0.6074$, lo que significa que el modelo ...

2. Pregunta 2

Las covariables con valores p más altas fueron (inserte covariables), luego a través de la tabla de todas las regresiones posibles se pretende hacer la siguiente prueba de hipótesis usando sumas de cuadrados extra.

$$\begin{cases} H_0 : \beta_2 = \beta_4 = 0 \\ H_1 : \text{Al menos uno de los coeficientes es diferente de cero} \end{cases}$$

Cuadro 3: Resumen tabla todas regresiones

	SSE	Covariables en el modelo
Modelo completo	2352.47	X1 X2 X3 X4
Modelo reducido	2551.28	X1 X3

Se construye el estadístico de prueba como

$$\begin{aligned} F_0 &= \frac{SSE(\beta_0, \beta_1, \beta_3 | \beta_2, \beta_4) - SSE(\beta_0, \dots, \beta_4)}{MSE(\beta_0, \dots, \beta_4)} \stackrel{H_0}{\sim} F_{2,45} \\ &= \frac{(2551.28 - 2352.47)/2}{52.27711} \\ &= 1.901501 \end{aligned}$$

Luego, comparando F_0 con $F_{0.975,2,45} = 4.0085$, se puede ver que $F_0 < F_{0.975,2,45}$, por tanto no se rechaza H_0 y las variables pueden ser descartadas del modelo.

3. Pregunta 3

Se plantea la siguiente prueba de hipótesis.

$$\begin{cases} H_0 : \beta_2 = \beta_4, \beta_1 = 0 \\ H_1 : \text{Al menos una de las igualdades no se cumple} \end{cases}$$

Reescribiendo la hipótesis matricialmente

$$\begin{cases} H_0 : \mathbf{L}\underline{\beta} = \underline{\mathbf{0}} \\ H_1 : \mathbf{L}\underline{\beta} \neq \underline{\mathbf{0}} \end{cases}$$

con \mathbf{L} dada por

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

El modelo reducido está dado por

$$Y_i = \beta_0 + \beta_2 X_{2i}^* + \beta_3 X_{3i} + \varepsilon_i, \quad \varepsilon_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2); \quad 1 \leq i \leq 50$$

donde $X_{2i}^* = X_{2i} + X_{4i}$.

Finalmente

$$F_0 = \frac{SSE(MR) - SSE(FM)/2}{MSE(FM)} \stackrel{H_0}{\sim} F_{2,45}$$

4. Pregunta 4

4.1. Supuestos del modelo

4.1.1. Normalidad de los residuales

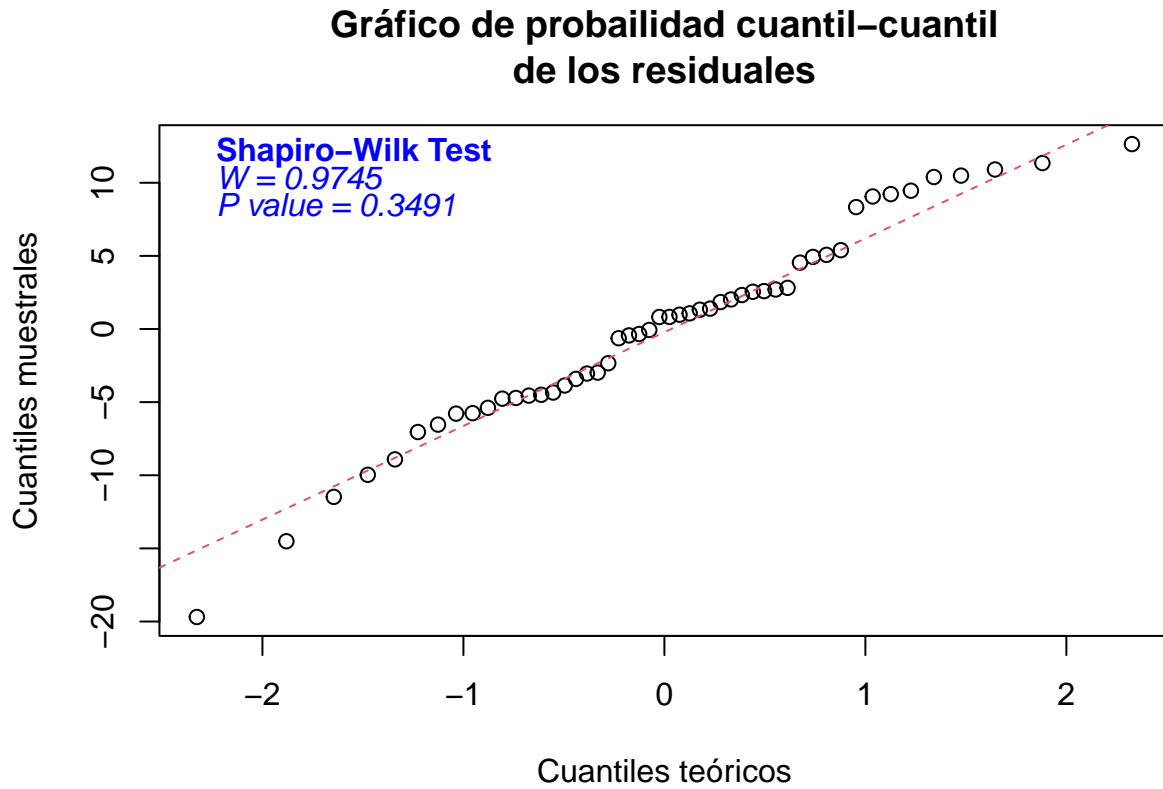


Figura 1: Normalidad de los residuales

4.2. Verificación de las observaciones

4.2.1. Datos atípicos

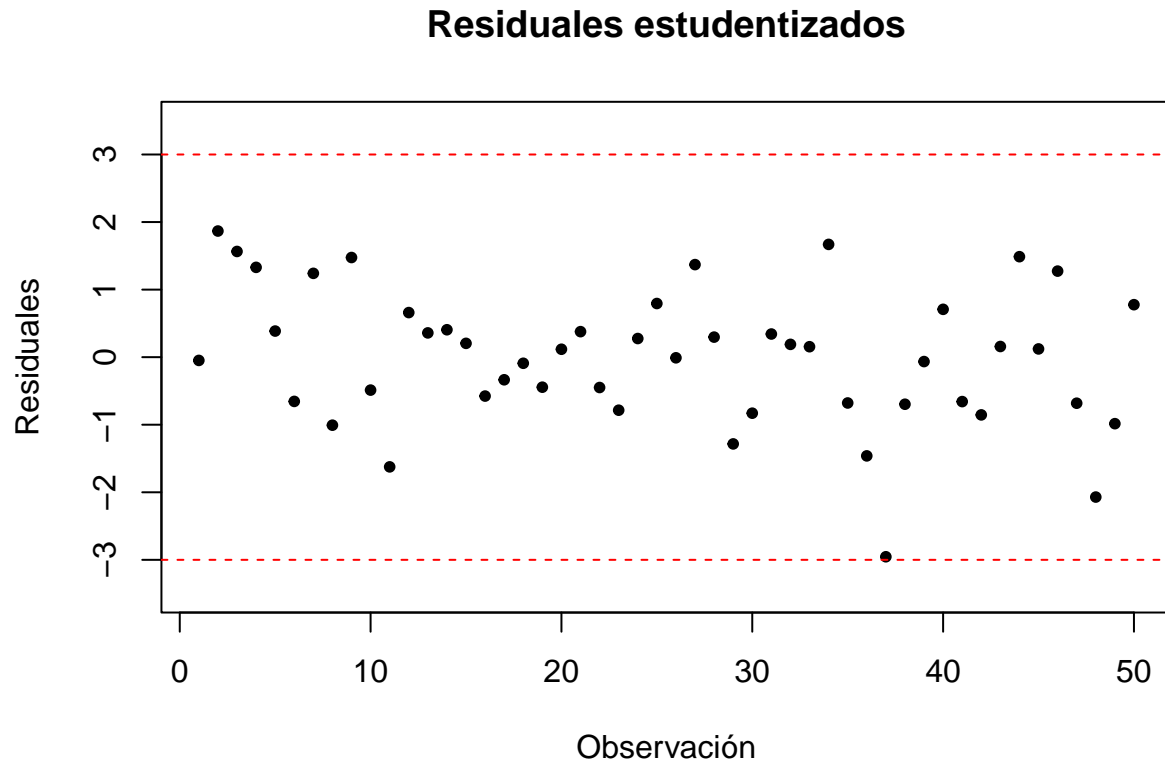
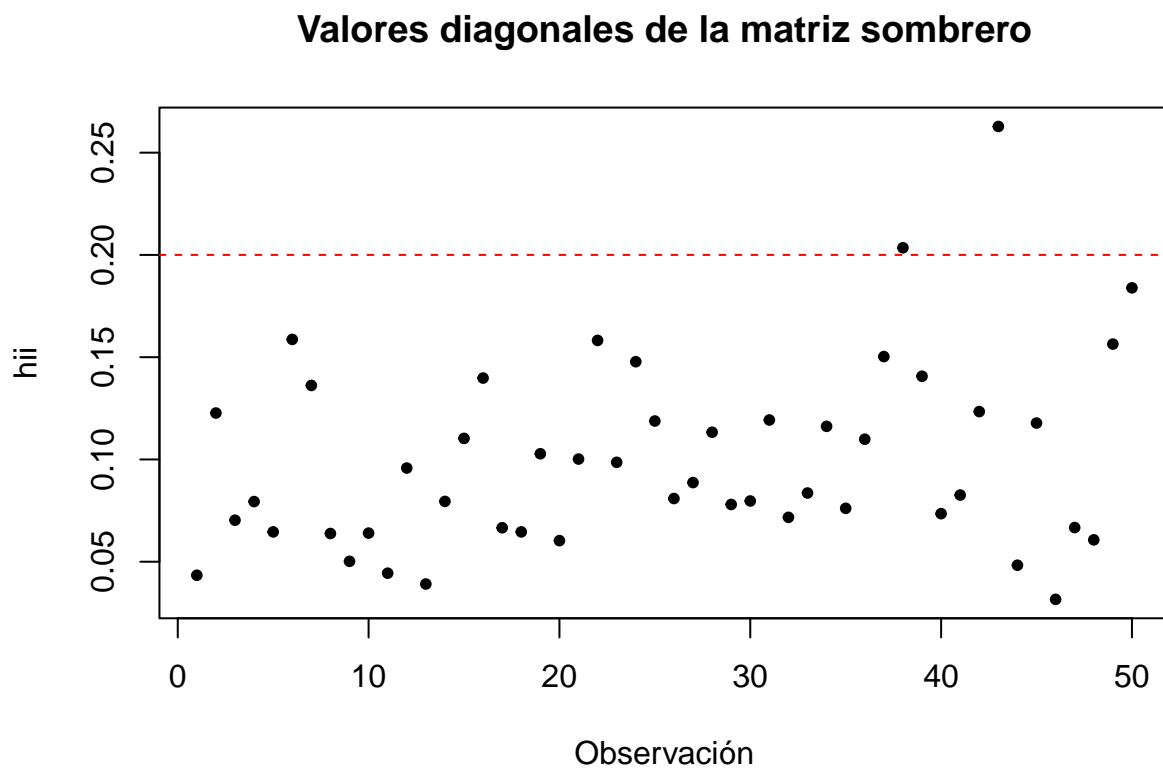


Figura 2: Identificación de datos atípicos

4.2.2. Con los h_{ii}



Las dos observaciones que son influenciales son la 38 y 43, pues son mayores que 0.2.