Taller 5 Estadísitca II

Autor

Simón Pedro Galeano Muñoz

Docentes

Raúl Alberto Pérez Agamez
Carlos Mario Lopera Gómez

Asignatura

Estadística II

UNIVERSIDAD NACIONAL DE COLOMBIA

Sede Medellín
Noviembre de 2021

Índice

Índice de figuras

Índice de cuadros

1. Ejercicio 1

a) Verdadero, pues como H es simétrica e idempotente se tiene que $H^{2021} = H$ y $H^T = H$, por tanto $H = (H^{2021})^T$.
b) Verdadero, ya que $gl(SSE) = n - p = 98$, entonces $n = 98 + p = 98 + 4 = 102$.
c) Verdadero, es una definición teórica.
d) Verdadero, pues el rango de la matriz $\mathbf{L}$ y el número de filas linealmente independientes son lo mismo.

2. Ejercicio 2

2.1. Descripción de la base de datos

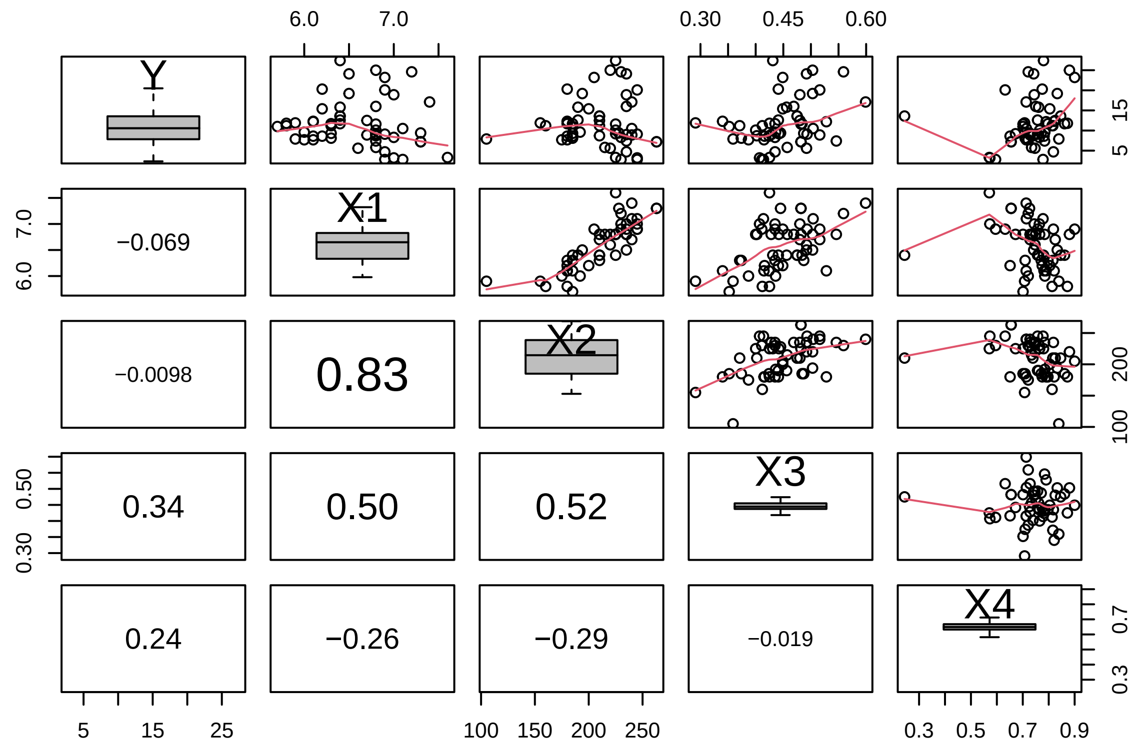

Figura 1: Matriz de gráficos

Del gráfico anterior podemos notar que existe una relación de manera lineal muy pobre entre la respuesta y las covariables, además, se puede notar redundancia entre las variables X_1 y X_2 puesto que esán relacionadas fuertemente de manera lineal y se podría explicar una con la otra. Finalmente, la variable X_2 es la que presenta una dispersión mayor respecto al resto, además, no se observaron datos atípicos.

2.2. Ajuste del modelo

Se desea ajustar el siguiente modelo de regresión lineal múltiple

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \varepsilon_i,\ \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2);\ 1 \leq i \leq 54$$

Una vez ajustado el modelo se obtuvo lo siguiente

$$\hat{y}_i = 4.148 + -3.6905 x_{1i} + 0.0095 x_{2i} + 47.9402 x_{3i} + 11.3710 x_{4i};\ 1 \leq i \leq 54$$

2.3. Análisis marginal de los coeficientes

Para el análisis marginal se realizó la siguiente prueba para $j = 1, \cdots, 4$.

$$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{cases}$$

con $T_0 = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim t_{49}$ bajo H_0

Cuadro 1: Resumen de los coeficientes

	Estimación	Error estándar	T_0	Valor P
β_0	4.1487	14.8550	0.2793	0.7812
β_1	-3.6905	2.9708	-1.2423	0.2201
β_2	0.0095	0.0463	0.2043	0.8390
β_3	47.9402	15.7091	3.0517	0.0037
β_4	11.3710	7.8685	1.4451	0.1548

Del cuadro 1, se puede observar que a nivel marginal, la única variable cuyo efecto es significativo en la respuesta, es X_3, con un nivel de significancia de $\alpha = 0.05$.

2.4. Análisis de varianza

Se desea verificar la significancia de la regresión mediante la tabla ANOVA, esto para contrastar el siguiente juego de hipótesis.

$$\begin{cases} H_0 : \beta_1 = \cdots = \beta_4 = 0 \\ H_1 : \text{Al menos un } \beta_j \neq 0 \end{cases}$$

Se obtuvo lo siguiente realizando el análisis de varianza

Cuadro 2: Tabla ANOVA para el modelo

	Suma de cuadrados	gl	Cuadrado Medio	F_0	Valor P
Regresión	409.934	4	102.4834	3.50058	0.0136397
Error	1434.532	49	29.2762		

Donde $F_0 = \frac{MSR}{MSE} \sim F_{4,49}$ bajo H_0

De la tabla ANOVA se rechaza la hipótesis nula a un nivel de significancia del 5 %, que al menos uno de los coeficientes del modelo es significativo.

2.5. Prueba lineal generalizada

Se tiene que los parámetros β_1, β_2 y β_4 no son significativos de manera marginal, por tanto se desea probar la siguiente hipótesis.

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

donde

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

Se define el modelo reducido como sigue

$$y_i = \beta_3 x_{3i} + \varepsilon_i,\ \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2);\ 1 \leq i \leq 54$$

y se toma al modelo completo como el que se definió anteriormente.

Se calculó el estadístico $F_0 = \frac{MSH}{MSE} \sim F_{3,49}$ bajo H_0 como se muestra a continuación

$$F_0 = \frac{SSE(\beta_0, \beta_3) - SSE(\beta_0, \beta_1, \beta_3, \beta_3, \beta_4)/3}{MSE} = 2.2574$$

con lo que se obtiene un Valor P de 0.09341, por lo que no se rechaza la hipótesis nula.

2.6. Modelo reducido

Se ajustó el modelo reducido y se obtuvo que su estimación fue

$$\hat{y}_i = 35.34 x_{3i};\ 1 \leq i \leq 54$$

Se presenta la gfráfica de dicho modelo

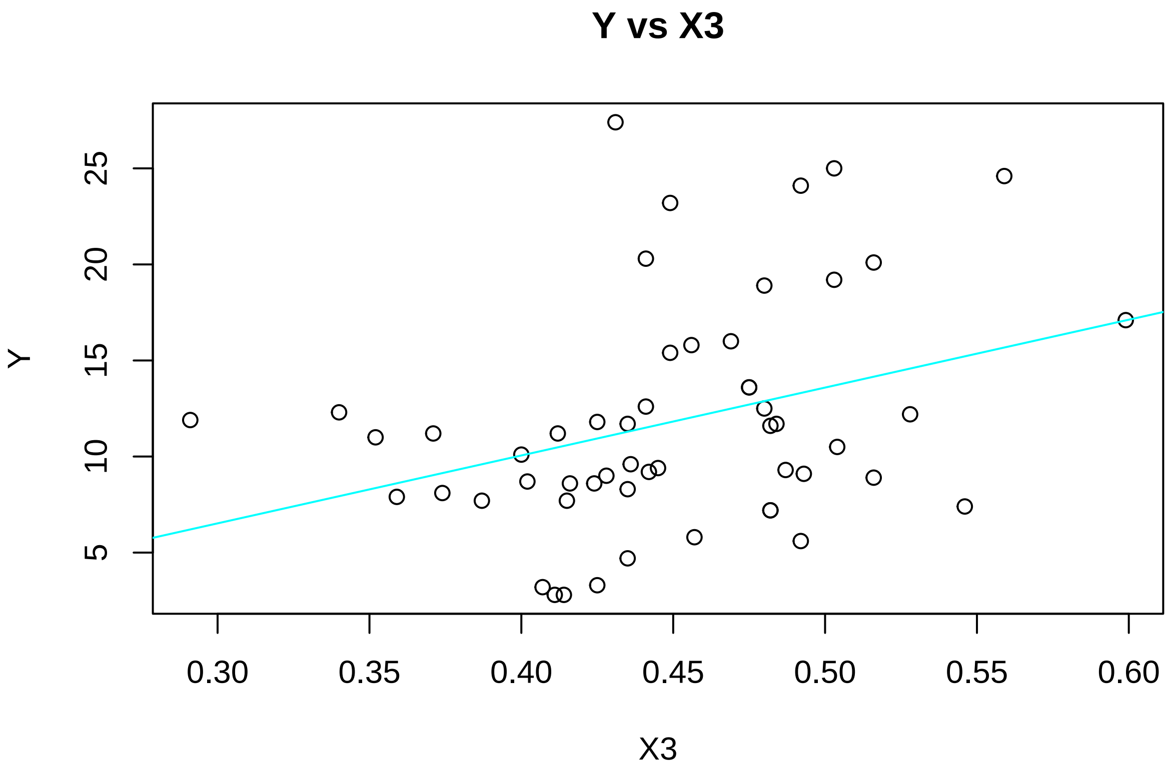

Figura 2: Modelo reducido