

Trabajo de Regresión múltiple

Desarrollamos una ecuación de regresión para estimar la producción de gasolina como una función de las propiedades de destilación de cierto tipo de petróleo crudo. Se identificaron cuatro variables de predicción: la gravedad del petróleo crudo, 0API (X1); la presión de vapor del petróleo crudo, psi (X2); el punto de 10% **ASTM** para el petróleo crudo, 0F (X3) y el punto final **ASTM** para la gasolina, 0F (X4). Los primeros dos miden la gravedad y la presión de vapor del petróleo crudo. El punto de 10% **ASTM** es la temperatura para la cual se ha evapo-rado cierta cantidad de líquido, y el punto final para la gasolina es la temperatura para la cual se ha evaporado todo el líquido. La variable respuesta fue la cantidad de gasolina producida expresada como un porcentaje respecto al total de petróleo cru-do. El objetivo radicó en determinar una ecuación de regresión para la producción de gasolina como una función lineal de las propiedades de destilación de cierto tipo de petróleo crudo X1, X2, X3 y el punto final deseado para la gasolina X4. Los datos seleccionados aleatoriamente se muestran en la tabla.

Y	X1	X2	X3	X4
6,9	38,4	6,1	220	235
14,4	40,3	4,8	231	307
7,4	40	6,1	217	212
8,5	31,8	0,2	316	365
8	40,8	3,5	210	218
2,8	41,3	1,8	267	235
5	38,1	1,2	274	285
12,2	50,8	8,6	190	205
10	32,2	5,2	236	267
15,2	38,4	6,1	220	300
26,8	40,3	4,8	231	367
14	32,2	2,4	284	351
14,7	31,8	0,2	316	379
6,4	41,3	1,8	267	275
17,6	38,1	1,2	274	365
22,3	50,8	8,6	190	375
24,8	32,2	5,2	236	360
26	38,4	6,1	220	365
34,9	40,3	4,8	231	395
18,2	40	6,1	217	272
23,2	32,2	2,4	284	424
18	31,8	0,2	316	428
13,1	40,8	3,5	210	273
16,1	41,3	1,8	267	358
32,1	38,1	1,2	274	444
34,7	50,8	8,6	190	345
31,7	32,2	5,2	236	402
33,6	38,4	6,1	220	410
30,4	40	6,1	217	340

26,6	40,8	3,5	210	347
27,8	41,3	1,8	267	416
45,7	50,8	8,6	190	407

Donde la variable Y corresponde a la producción de gasolina que esta en función de las otras cuatro variables que son:

- X1: la gravedad del petróleo crudo.
- X2: la presión de vapor del petróleo crudo.
- X3: el punto de 10% ASTM del petróleo crudo.
- X4: el punto final ASTM para la gasolina.

Los datos obtenidos en la tabla son los siguientes:

Regression

TABLES ENCODING NOTES

model summary

210125

20010101

ANÁLISIS DE REGRESIÓN

Podemos obtener la ecuación de regresión la cual es:

$$Y = B_1 \cdot X_1 + B_2 \cdot X_2 + B_3 \cdot X_3 + B_4 \cdot X_4$$

Mediante los mínimos cuadrados obtenemos la ecuación:

$$Y = 0.03637 \cdot X_1 + 0.16 \cdot X_2 - 0.171 \cdot X_3 + 0.147 \cdot X_4 + 9.674$$

Fuente de variación	Grados libertad	Suma de los cuadrados	Cuadrados medios	Estadística F
Regresión	$K = m - 1$	$B'X'Y - Y_i/n$	$SCR/(m-1)$	$\frac{SCR/(m-1)}{SCE/(n-m)}$
Error	$n - m$	$Y'Y - B'X'Y$	$SCE/(n-m)$	
Total	$n - 1$	$Y'Y - Y_i/n$		

Donde X' es una matriz de orden $m \times n$

Teniendo presente también que el coeficiente de correlación múltiple es:

$$R^2 = \frac{SCR}{Y'Y} = 1 - \frac{SCE}{Y'Y}$$

STC STC'

Así, nuestra hipótesis nula: $H: B_1 = B_2 = B_3 = B_4 = 0$

Podemos observar mediante las tablas que nuestro f observado sí pertenece a la región crítica por lo que nuestra hipótesis nula será rechazada, por lo que todos los coeficientes de regresión (excepto el término constante) son iguales a cero.

Tenemos que las variables rechazadas son aquellas que muestran un p -value mayor que el 0.5.

Podemos decir que nuestros coeficientes de correlación múltiple son relativamente altos. Existe una pequeña duda de que la regresión entre la producción de gasolina y las cuatro variables de predicción sea estadísticamente significativas

En este ejercicio las desviaciones estándares de los estimadores de mínimos cuadrados para X_0 y B_2 son grandes, lo que sugiere que X_2 , y posiblemente otras variables de predicción, puedan no tener un gran efecto sobre la producción de gasolina.