Daniel Freire

La base de datos data\_rls\_uti.xlsx posee información sobre utilidad y ventas, con 40 registros.

En primer lugar se realiza una regresión lineal de las utilidad contra las ventas, en donde la utilidad es la variable dependiente y ventas es la variable independiente, el resultado de esta regresión es:

Call:

lm(formula = Utilidad ~ Ventas, data = data)

Residuals:

Min 1Q Median 3Q Max

-676.35 -302.04 42.59 303.67 612.49

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 137.08270 282.69543 0.485 0.631

Ventas 0.43994 0.01859 23.663 <2e-16 \*\*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 367.4 on 38 degrees of freedom

Multiple R-squared: 0.9364, Adjusted R-squared: 0.9348

F-statistic: 559.9 on 1 and 38 DF, p-value: < 2.2e-16

Se puede observar que el p-valor (Pr(>|t|) ) para el intercepto es mayor a 0,05 por lo que se acepta la hipótesis nula (), es decir el intercepto es no significativo estadísticamente a un nivel de confianza del 95%; y por el contrario la variable ventas posee un p-valor menor a 0,05 por lo que se rechaza la hipótesis nula (), por lo cual es estadísticamente significativa a un nivel de confianza del 95%.

Ya que el intercepto es no significativo se procede a centrar los datos, esto lo hacemos restando de cada variable su media, con estas nuevas variables realizamos una nueva regresión lineal en donde utilidad1 es la variable centrada de utilidad y ventas 1 es la variable centrada de ventas, y obtenemos el siguiente resultado:

Call:

lm(formula = utilidad1 ~ ventas1)

Residuals:

Min 1Q Median 3Q Max

-676.35 -302.04 42.59 303.67 612.49

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -6.442e-13 5.809e+01 0.00 1

ventas1 4.399e-01 1.859e-02 23.66 <2e-16 \*\*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 367.4 on 38 degrees of freedom

Multiple R-squared: 0.9364, Adjusted R-squared: 0.9348

F-statistic: 559.9 on 1 and 38 DF, p-value: < 2.2e-16

Se observa que el intercepto continúa siendo no significativo pero su valor es prácticamente cero por lo que nuestro modelo lineal sería de la forma:

Utilidades1= 0.4399\*ventas1

Si realizamos el ANOVA se obtiene el siguiente resultado:

Df Sum Sq Mean Sq F value Pr(>F)

ventas1 1 75578286 75578286 559.9 <2e-16 \*\*\*

Residuals 38 5129142 134977

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Como se puede ver el p-valor (Pr(>F)) es menor a 0.05 se rechaza la hipótesis nula de no dependencia y se concluye que la variable utilidad1 depende linealmente de la variable ventas1.

Los intervalos de confianza al 95% para los betas son:

2.5 % 97.5 %

(Intercept) -117.5968432 117.5968432

ventas1 0.4022981 0.4775722

Finalmente se analiza los residuos:

En primer lugar se verifica que los residuos están distribuidos normalmente con media cero, para lo cual se obtiene la media que es -7.771561e-16 (prácticamente cero) y mediante el histograma y el grafico cuantil-cuantil se observa que están distribuidos normalmente

 

Y mediante el gráfico de residuos vs pronosticados se puede observar que la varianza de los residuos es constante que al no presentar patrones se puede concluir que son independientes



Por lo tanto el modelo cumple las hipótesis H y N con y es adecuado para modelar las utilidades mediante las ventas de manera lineal.