

PRESENTACION DE UNA ECUACION DE REGRESION

POR : ADOLFO CORTES R.

Profesor Asociado U.N.

El análisis de regresión tiene aplicación en muy diversas disciplinas científicas pero con excepción de la rama econométrica, su presentación se ha descuidado muchísimo. Este artículo trata de dar normas para unificar su escritura.

1. Planteamiento de una ecuación de regresión.

$$\hat{Y} = \hat{B}_1 + \hat{B}_2 X_2 + \hat{B}_3 X_3 + \dots + \hat{B}_k X_k$$

2. La solución

Se obtiene por el método de mínimos cuadrados para calcular :

a. Las estimaciones de los coeficientes de regresión (\hat{B}_i)

b. La varianza residual

$$S^2_{Y.X_2, \dots, X_k} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n - k}$$

c. El error de estimación (EE)

$$EE = \sqrt{S^2_{Y.X_2, X_3, \dots, X_k}}$$

d. La varianza y el error de estimación de las \hat{B}_i

e. El coeficiente de determinación (r^2) y el de correlación (r).

3. Análisis de Significación.

El análisis de significación de un modelo se hace a través del test de hipótesis para cada una de las variables y para todas en su conjunto.

a. Análisis Individual, para cada variable Independiente.

Para cada B_i se hace la prueba de hipótesis :

$$H_0 : B_i = 0$$

$$H_1 : B_i \neq 0$$

En general, interesa rechazar la hipótesis nula, lo cual implica aceptar la hipótesis alternativa y lo cual significa que al cambiar X_i en una unidad, Y cambia en B_i y por lo tanto que la acción de esta variable independiente es explicativa en el comportamiento de Y .

La estadística de trabajo (t ratio) es : $\frac{\hat{B}_i}{S\hat{B}_i}$.

Cuyo valor se escribe dentro de un doble paréntesis. Para examinar el nivel α de significación se examina una tabla t con $n-k$, grados de libertad (gl) donde :

n = No. de observaciones

k = No. de variables en la ecuación.

b. Análisis Global del Modelo.

Se plantea la hipótesis de que el coeficiente de correlación múltiple es igual a cero, (o de que todo $B_i = 0$) es decir, que no hay asociación entre la variable dependiente Y y todas las variables independientes en su conjunto :

$$H_0 : \rho_{Y, X_2, X_3, \dots, X_k} = 0$$

$$H_1 : \rho_{Y, X_2, X_3, \dots, X_k} \neq 0$$

En este caso también interesaría rechazar la hipótesis nula. La estadística de prueba (F ratio) es :

$$F_{k-1, n-k} = \frac{r^2}{1-r^2} \cdot \frac{n-k}{k-1}$$

Cuyo valor también se escribe dentro de un doble paréntesis. Para examinar su significación se recurre a una tabla F con (k-1) grados de libertad en el numerador y (n - k) grados de libertad en el denominador.

4. Determinación del nivel de significación

Como en el análisis de regresión, lo que en general interesa es rechazar cada hipótesis nula al máximo nivel (uno por mil). Para ello se recorre la correspondiente tabla de probabilidad y se va colocando un asterisco sobre el coeficiente, cada vez que la estadística de trabajo cae en la zona de rechazo, en la siguiente forma:

- Si el test es significativo al 10% = *
- Si el test es significativo al 5% **
- Si el test es significativo al 1% ***
- Si el test es significativo al 1‰ = ****

Cuando el test es significativo implica que la hipótesis nula se rechaza a ese nivel de significación, cuando el test no es significativo a ninguno de los niveles anteriores, el correspondiente coeficiente no lleva estrellas.

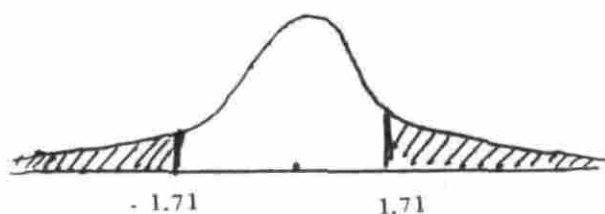
Ejemplo 1 :

Considérese que para un test, se obtuvo como valor calculado de la estadística de trabajo :

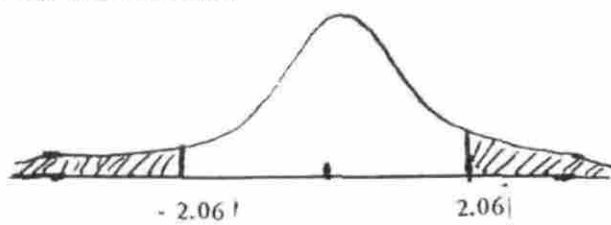
$$t_{25, \text{cal}} = 2.94$$

Utilizando una t_{25} se recorren los diversos niveles :

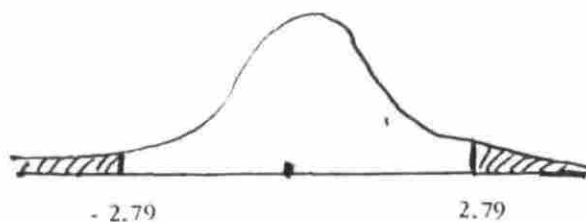
- a. Para $\alpha = 10\%$ cae en la zona de rechazo y por lo tanto llevaría una estrella o asterisco y se pasa al siguiente nivel.



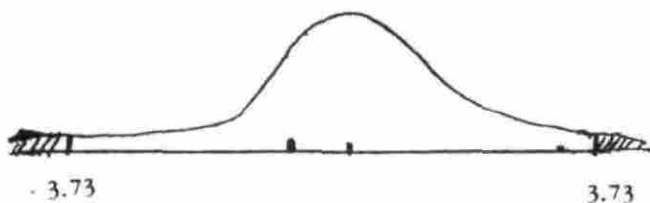
- b. Para $\alpha = 5\%$, el t observado también cae en la zona de rechazo, y por lo tanto llevaría otro asterisco.



- c. Como el test fue significativo, se pasa al siguiente nivel $\alpha = 1\%$ y también cae en la zona de rechazo, llevando otra estrella.



- d. Se pasa al último nivel $\alpha = 1$ por mil, cayendo ahora en la zona de aceptación y terminando el proceso.



conclusión : El test fue significativo al 1 por ciento (***) . Véase en el ejemplo 2 el primer coeficiente de la ecuación. En resumen, el proceso termina cuando la estadística de trabajo cae en la zona de aceptación.

5. Presentación de una Ecuación de Regresión

La presentación correcta de una ecuación de regresión según las normas que se recomiendan en el presente trabajo sería :

a. Ecuación : $\hat{Y} = \hat{B}_1^* + \hat{B}_2^{**} X_2 + \hat{B}_3^{***} X_3 + \dots + \hat{B}_k X_k$

b. Error de los Coeficientes :

$$\left(\begin{array}{c} Y \\ () \\ () \\ () \end{array} \right)$$

$$t \text{ ratio } \left(\left(\begin{array}{c} \\ \\ \\ \end{array} \right) \right) \left(\left(\begin{array}{c} \\ \\ \\ \end{array} \right) \right) \left(\left(\begin{array}{c} \\ \\ \\ \end{array} \right) \right) \left(\left(\begin{array}{c} \\ \\ \\ \end{array} \right) \right)$$

c. $gl =$ $r^2 =$ $EE =$ $d =$
 ((F ratio))

Lo anterior significa que :

1. La ecuación debe escribirse utilizando los valores numéricos de las estimaciones de los B_i , acompañadas de los asteriscos correspondientes al máximo nivel de significación que alcanzó el test y de los nombres o abreviaturas de las correspondientes variables independientes. Es recomendable utilizar para las variables su nombre completo o iniciales que le sean de fácil identificación al lector. Por ejemplo : V por ventas, GP por gastos de publicidad, Y por ingreso, P por precio, etc. en lugar de los símbolos $Y, X_2, X_3, X_4, \dots, X_k$, que son prácticos para la enseñanza.
2. Exactamente debajo de cada \hat{B}_i , dentro de un paréntesis su error y debajo el t ratio, dentro de un doble paréntesis.
3. Luego en una línea :
 - a. Los grados de libertad $gl = n - k$
 - b. El r^2 con sus estrellas y debajo el F ratio en un doble paréntesis.
 - c. El error de estimación (EE) y
 - d. El Durbin-Watson, representado por la letra d.

Ejemplo 2 :

$$\hat{V} = - 1.265^{***} + 0.2299^{****} \text{ G.P.} + 0.0162^{****} \text{ Y} - 0.0062 \text{ P}$$

(0.43)	(0.003)	(0.001)	(0.005)
((2.94))	((76.6))	((16.2))	((1.24))

$gl = 22$ $r^2 = 0.9995^{****}$ $EE = 1.22$ $d = 1.43^{**}$
((14.659))

NOTA :

En la ecuación anterior se puede concluir lo siguiente :

- a. Que los gastos en publicidad y el ingreso son significativos al uno por mil y por lo tanto explicativos del volumen de ventas.
- b. Los precios no fueron significativos a ningún nivel y por lo tanto se deberían eliminar como determinantes del nivel de las ventas.
- c. El coeficiente de correlación múltiple fue altamente significativo o sea que el modelo en su conjunto es aceptable para explicar el comportamiento de las ventas.
- d. El Durbín-Watson fue significativo al 5% por lo tanto, el modelo presenta problemas de autocorrelación.