

## $R^2$ y $R^2$ ajustada

- $R^2$ : Se le llama *coeficiente de determinación* y se interpreta como la proporción de la variación muestral de  $y$  que es explicada por el modelo de regresión, es decir, qué tan bien se ajusta el modelo a los datos observados [1].

Recordando el modelo de regresión lineal

$$Y = X\beta + u = \hat{Y} + u \quad (1)$$

para una sola observación se tiene que

$$y_i = \hat{y}_i + u_i = \mathbf{x}_i\beta + u_i \quad (2)$$

haciendo el desarrollo algebraico (ver [2]) se llega a la expresión de  $R^2$  como:

$$R^2 = 1 - \frac{SRC}{STC} = 1 - \frac{\mathbf{u}'\mathbf{u}}{\mathbf{y}'\mathbf{M}^0\mathbf{y}} = 1 - \frac{\sum_{i=1}^n u_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (3)$$

donde  $SRC$  es la suma de los residuos al cuadrado,  $STC$  es la suma total de cuadrados y  $\mathbf{M}^0$  es la *matriz de centrado* o *matriz de desviaciones respecto a la media*.

Durante el desarrollo de la ecuación (3) se utilizan propiedades que requieren que  $\mathbf{X}$  tenga una columna de unos (intercepto) ya que de lo contrario al calcular el  $R^2$ , el resultado es impredecible. Nunca será mayor y puede ser mucho menor que la misma cifra calculada para la regresión con un término constante incluido. Incluso puede ser negativo [2].

Es importante mencionar que el  $R^2$  mide la bondad de ajuste dentro de la muestra, en el sentido de conocer la cercanía entre un valor  $Y$  estimado y su valor real en la muestra dada. No hay garantía de que pronosticará bien las observaciones fuera de la muestra.

- $R^2$  ajustada ( $\bar{R}^2$ ): Se le llama *coeficiente de determinación ajustado* y es una versión modificada del  $R^2$  y se desarrolla a partir de los problemas que presenta el  $R^2$  al analizar la bondad del ajuste. Uno de ellos es que el  $R^2$  siempre aumenta al agregar variables a la regresión. El  $R^2$  ajustada penaliza la inclusión de variables en el modelo y se calcula como:

$$\bar{R}^2 = 1 - \frac{SRC/(n-k)}{STC/(n-1)} = 1 - \frac{\mathbf{u}'\mathbf{u}/(n-k)}{\mathbf{y}'\mathbf{M}^0\mathbf{y}/(n-1)} = 1 - \frac{\sum_{i=1}^n u_i^2/(n-k)}{\sum_{i=1}^n (y_i - \bar{y})^2/(n-1)}, \quad (4)$$

donde  $n$  es el número de observaciones y  $k$  es el número de parámetros estimados.

## Criterios de información

- Criterio de Información de Akaike ( $CIA$ ): Es una medida utilizada para comparar modelos estadísticos, también utiliza la idea de una penalización por añadir mas variables, se define como [3]:

$$CIA = e^{2k/n} \frac{SRC}{n} = e^{2k/n} \frac{\mathbf{u}'\mathbf{u}}{n}, \quad (5)$$

la expresión logaritmica muchas veces es más útil

$$\ln(CIA) = \frac{2k}{n} + \ln\left(\frac{SRC}{n}\right) = \frac{2k}{n} + \ln\left(\frac{\mathbf{u}'\mathbf{u}}{n}\right). \quad (6)$$

también se puede expresar en términos del valor máximo de la función de verosimilitud ( $\hat{L}$ ) [2]:

$$CIA = -2\ln(\hat{L}) + 2k. \quad (7)$$

Al comparar dos o más modelos, se preferirá el que tenga el menor valor CIA. Una ventaja del CIA es que resulta útil no sólo para el desempeño de la predicción dentro de la muestra, sino también para el de la predicción fuera de la muestra de un modelo de regresión. [3]

- Criterio de información de Schwarz (*CIS*): Similar al criterio de Akaike, pero impone una penalización mayor con un factor  $(k/n)\ln(n)$  en lugar de  $2k/n$  [3]:

$$\ln(CIS) = \frac{k}{n} \ln(n) + \ln\left(\frac{SRC}{n}\right) = \frac{k}{n} \ln(n) + \ln\left(\frac{\mathbf{u}'\mathbf{u}}{n}\right), \quad (8)$$

o en términos del valor máximo de la función de verosimilitud [2]:

$$CIS = -2\ln(\hat{L}) + k\ln(n). \quad (9)$$

Al igual que en CIA, mientras más pequeño sea el valor de CIS, mejor será el modelo. De nuevo, al igual que en CIA, CIS sirve para comparar el desempeño del pronóstico dentro de la muestra y fuera de la muestra de un modelo.

- Criterio de información de Hannan-Quinn (*CIHQ*): El criterio de información de Hannan-Quinn también penaliza la adición de variables, pero con un factor intermedio entre los dos anteriores, es decir, que el factor crece más rápido que  $2k/n$  pero más lento que  $(k/n)\ln(n)$  y se define de la siguiente manera [4]:

$$CIHQ = -2\ln(\hat{L}) + 2k\ln(\ln(n)) \quad (10)$$

donde  $\hat{L}$  es el valor máximo de la función de verosimilitud,  $k$  es el número de parámetros estimados y  $n$  es el número de observaciones.

Al igual que en los casos anteriores, se prefiere el modelo con el valor más bajo de CIHQ.

## Referencias

- [1] Jeffrey M. Wooldridge. *Introductory econometrics: a modern approach*. 4. ed. Mason, Ohio: South Western cengage learning, 2009. ISBN: 978-0-324-66054-8.
- [2] William Greene. *Econometric analysis*. 7. ed. Pearson series in economics. Boston Munich: Prentice Hall, 2012. ISBN: 978-0-13-139538-1.

- [3] Damodar N. Gujarati. *Econometría*. Quinta edición. México, D.F.: McGraw-Hill, 2015.
- [4] E. J. Hannan y B. G. Quinn. “The Determination of the Order of an Autoregression”. En: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 41.2 (1 de ene. de 1979), págs. 190-195. ISSN: 1369-7412, 1467-9868. DOI: [10.1111/j.2517-6161.1979.tb01072.x](https://doi.org/10.1111/j.2517-6161.1979.tb01072.x). URL: <https://academic.oup.com/jrssb/article/41/2/190/7027596>.