Uma contribuição ao estudo da regressão linear múltipla

Jessé Montello *

1. Introdução; 2. Coeficiente de correlação parcial entre variáveis independentes; 3. Coeficiente de correlação simples entre os estimadores de mínimos quadrados de dois coeficientes de regressão.

1. Introdução

Considere-se a equação de regressão linear

$$(1) \qquad y = X\beta + \varepsilon,$$

onde y é o vetor das observações da variável dependente:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}$$

* Professor da Escola de Pós-Graduação em Economia da FGV.

X é a matriz das observações das variáveis independentes:

$$X = \begin{bmatrix} x_{11} & x_{12} & \ldots\ldots & x_{1k} \\ x_{21} & x_{22} & \ldots\ldots & x_{2k} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x_{n1} & x_{n2} & \ldots\ldots & x_{nk} \end{bmatrix}.$$

ε é o vetor coluna das perturbações aleatórias:

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

e β o vetor dos coeficientes de regressão:

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix}$$

O estimador de mínimos quadrados de β é o vetor:

$$b = \begin{bmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ b_k \end{bmatrix}$$

Com respeito ao vetor ε, o modelo clássico de regressão estabelece as seguintes propriedades:

$$E(\varepsilon \mid X) = 0, \quad Var(\varepsilon \mid X) = \sigma^2_I$$

Com base nessas propriedades demonstra-se que:

$$E(b) = O, \quad Var\ b = (X' X)^{-1} \sigma^2$$

desde que X tenha característica $k < n$.

Considere-se agora as variáveis independentes:

$$(2) \qquad x_1, x_2, \ldots, x_k$$

e designemos por $r_{ij \cdot 1,2,\ldots,i-1,i+1,\ldots,j-1,j+1,\ldots,k}$, ou simplesmente por r^*_{ij} o coeficiente de correlação parcial de amostra entre x_i e x_j, quando se elimina a influência das outras varáveis.

Esses coeficientes de correlação parcial entre as variáveis (2) formam a matriz:

$$R^* = \begin{bmatrix} 1 & r^*_{12} & \ldots & r^*_{1k} \\ r^* & 1 & \ldots & r^*_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ r^*_{k1} & r^*_{k2} & \ldots & 1 \end{bmatrix}$$

Designando-se por S a matriz: $(X'X)$, tem-se:

$$Var\ b = S^{-1} \sigma^2$$

Representando-se por S^{ij} o elemento genérico da matriz inversa S^{-1}, tem-se:

$$Var\ b = \begin{bmatrix} S^{11} & S^{12} & \dots & S^{1k} \\ S^{21} & S^{22} & \dots & S^{2k} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ S^{k1} & S^{k2} & \dots & S^{kk} \end{bmatrix} \sigma^2$$

Dessa matriz deduz-se que os coeficientes de correlação entre as componentes de b formam a matriz:

$$P = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1k} \\ \rho_{21} & \rho_{22} & \dots & \rho_{2k} \\ \dots & \dots & \dots & \dots \\ \rho_{k1} & \rho_{k2} & \dots & 1 \end{bmatrix}$$

onde $\rho_{ij} = \dfrac{S^{ij}}{\sqrt{S^{ii}}\ \sqrt{S^{jj}}}$

No presente trabalho, vamos mostrar que:

$$\rho_{ij} = \rho(b_i, b_j) = -\overset{*}{r}_{ij}, \text{ para } i, j = \overline{1, k}$$

2. Coeficiente de correlação parcial entre variáveis independentes

Consideremos as regressões lineares entre x_1 e $x_3, \dots, x_k$ e entre x_2 e $x_3, \dots, x_k$.

A matriz X das observações das variáveis $x_1, x_2, \dots, x_k$ pode ser particionada do seguinte modo:

$$X = [x_1\ x_2\ w]$$

Designando-se por N a seguinte matriz, idem potente:

$$(3) \qquad N = I - w(w'w)^{-1}w',$$

é imediato que

$x_1'\ Nx_1$ representa a soma dos quadrados dos resíduos da primeira regressão mencionada;

$x_2' \, N x_2$ representa a soma dos quadrados dos resíduos da segunda regressão mencionada.

$x_1' \, N x_2$ representa a soma dos produtos desses resíduos.

Resulta então, que o coeficiente de correlação parcial entre x_1 e x_2 será:

$$r_{12}^* = \frac{x_1' \, N x_2}{\sqrt{(x_1' \, N x_1)\,(x_2' \, N x_2)}} \tag{4}$$

3. Coeficiente de correlação simples entre os estimadores de mínimos quadrados de dois coeficientes de regressão

Usando a notação da introdução, o coeficiente de correlação entre b_1 e b_2 é:

$$\rho_{12} = \rho\,(b_1,\, b_2) = \frac{S^{12}}{\sqrt{S^{11}\, S^{22}}} \,. \tag{5}$$

onde S^{ij} são os elementos da matriz inversa S^{-1} da matriz $S = X'X$.

Ora, sendo

$$X = [x_1 \; x_2 \; w] \,,$$

tem-se:

$$S = X' X = \left[\begin{array}{cc|c} x_1' \, x_1 & x_1' \, x_2 & x_1' \, w \\ x_2' \, x_1 & x_2' \, x_2 & x_2' \, w \\ \hline w' \, x_1 & w' \, x_2 & w' \, w \end{array}\right]$$

particionada da maneira indicada. Fazendo-se:

$$P_1 = \begin{bmatrix} x_1' \, x_1 & x_1' \, x_2 \\ x_2' \, x_1 & x_2' \, x_2 \end{bmatrix}, \quad Q_1 = w' \, w$$

$$R_1 = \begin{bmatrix} x_1' \, w \\ x_2' \, w \end{bmatrix}$$

tem-se

$$S = X' X = \begin{bmatrix} P_1 & R_1 \\ R_1' & Q_1 \end{bmatrix}$$

Por uma conhecida fórmula para inversão de matrizes particionadas, a submatriz da matriz S^{-1}, correspondente à submatriz P_1 da matriz S, será: [1]

$$P_2 = (P_1 - R_1 Q_1^{-1} R_1')^{-1}$$

Notando-se que:

$$P_1 - R_1 Q_1^{-1} R_1' = \begin{bmatrix} x_1' x_1, & x_1' x_2 \\ x_2' x_1, & x_2' x_2 \end{bmatrix} - \begin{bmatrix} x_1' w \\ x_2' w \end{bmatrix} (w' w)^{-1} \left[(x_1' w)^1, (x_2' w)^1\right]$$

$$= \begin{bmatrix} x_1' x_1, & x_1' x_2 \\ x_2' x_1, & x_2' x_2 \end{bmatrix} - \begin{bmatrix} x_1' w (w' w)^{-1} w' x_1, & x_1' w (ww)^{-1} w_2' \\ x_2' w (w' w)^{-1} w' x_1, & x_2' w (w')^{-1} w' x_2 \end{bmatrix}$$

ou

$$(6) \qquad P_1 - R_1 Q_1^{-1} R_1' = \begin{bmatrix} x_1' N x_1 & x_1' N x_2 \\ x_2' N x_1 & x_2' N x_2 \end{bmatrix}$$

tendo em vista a definição da matriz N dada pela relação (3).

Segue-se então que:

$$(P_1 - R_1 Q_1^{-1} R_1')^{-1} = \begin{bmatrix} x_1' N x_1 & x_1' N x_2 \\ x_2' N x_1 & x_2' N x_2 \end{bmatrix}^{-1}$$

$$= \frac{1}{\Delta} \begin{bmatrix} x_2' N x_2 & -x_2' N x_1 \\ -x_1' N x_2 & x_1' N x_1 \end{bmatrix}$$

tendo em vista que os elementos da matriz (6) são escalares e

$\Delta = (x_1' N x_1)(x_2' N x_2) - (x_1' N x_2)^2$ é o determinante dessa matriz.

Por conseguinte, tem-se:

$$S^{11} = \frac{x_2' N x_2}{\Delta}, \quad S^{22} = \frac{x_1' N x_1}{\Delta}$$

e

$$S^{12} = -\frac{x_1' N x_2}{\Delta} = -\frac{x_2' N x_1}{\Delta}$$

[1] Veja Theil, Henri. *Principles of econometrics*. John Wiley and Sons, p. 18.

Donde, tendo em vista (6), podemos escrever:

$$\rho_{12} = \rho(b_1, b_2) = -\frac{x_1' X x_2}{\sqrt{(x_1' X x_1)(x_2' X x_2)}}$$

Comparando esta relação com a (15), tem-se:

$$\rho_{12} = -\overset{*}{r}_{12}$$

Por alteração na ordem das variáveis independentes na equação de regressão (1), obtém-se que, em geral, tem-se:

$$\rho_{ij} = \rho(b_i, b_j) = -\overset{*}{r}_{ij},$$

que é a relação que desejávamos demonstrar. Essa relação permite enunciar o seguinte teorema:

Teorema: O coeficiente de correlação linear entre os estimadores de mínimos quadrados b_i e b_j *dos coeficientes de regressão* β_i *e* β_j ***é igual,*** *com sinal contrário, ao coeficiente de correlação parcial entre as variáveis* x_i *e* x_j, *a que se referem esses coeficientes.*

No caso particular em que se tem uma equação de regressão com três variáveis:

$$y_\alpha = \beta_0 + \beta_1 x_{\alpha_1} + \beta_2 x_{\alpha_2} + \varepsilon_\alpha, \ \alpha = \overline{1, n}$$

tem-se que:

$$\rho_{12}(b_1, b_2) = -\overset{*}{r}_{12} = -r_{12} = -\frac{\sum_{\alpha=1}^{n}(x_{\alpha_1} - \bar{x}_1)(x_{\alpha 2} - \bar{x}_2)}{\sqrt{\left[\sum_{\alpha=1}^{n}(x_{\alpha_1} - \bar{x}_1)^2\right]\left[\sum_{\alpha=1}^{n}(x_{\alpha_2} - x_2)^2\right]}}$$

tendo em vista que nesse caso o coeficiente de correlação parcial entre as variáveis x_1 e x_2 coincide com o coeficiente de correlação simples.

A propriedade enunciada vem mostrar a importância que tem a matriz $(X'X)^{-1}$ para a determinação dos coeficientes de correlação parcial entre as variáveis independentes (regressoras), que figuram numa equação de regressão. Pode-se empregar essa propriedade para calcular coeficientes parciais, usando os programas para computação eletrônica de regressão.

DEMOGRAFIA Y ECONOMIA

Redactores

Raúl Benítez Zenteno, Gerardo M. Bueno, Gustavo Cabrera Acevedo, Eliseo Mendoza Berrueto, Leopoldo Solís M., Rodolfo Stavenhagen, Claudio Stern, Luis Unikel S., Víctor L. Urquidi.

Secretario de redacción: Raúl de la Peña

Vol. VII, Núm. 1 (19) 1973

ARTICULOS

Víctor L. Urquidi y Adalberto García Rocha

La construcción de vivienda y el empleo en México

Luis Unikel, Crescencio Ruíz Chiapetto y Omar Lazcano

Factores de rechazo en la migración rural en México, 1950-1960

Larissa Lomnitz

Supervivencia en una barriada de la ciudad de México

Neide Lopes Patarra

Transición demográfica: —¿ Resumen histórico o teoría de población?—

Sofía Méndez Villarreal

La capacidad del sector industrial para generar ocupación

INFORMES

Planificación familiar: Tesis del Gobierno de México

Mensaje del episcopado al pueblo de México sobre la paternidad responsable

RESEÑA DE LIBROS

NOTAS BREVES

DEMOGRAFIA Y ECONOMIA se publica tres veces al año.

Redacción y administración:

El Colegio de México, Guanajuato 125, México 7, D. F.

Precio del ejemplar: México, $ 25.00; Extranjero, Dls. 2.50

Suscripción anual: México, $ 60.00; Extranjero, Dls. 6.00