UAEH
Universidad Autónoma del Estado de Hidalgo

Modelos de regresión múltiple

Alma Sofía Santillán Hernández, Juan Roberto Vargas Sánchez y Roberto Ávila Pozos

Regresión lineal múltiple

Resumen:

El objetivo de este material didáctico consiste en presentar la estimación de los coeficientes de un modelo de regresión lineal múltiple en forma matricial.

Keywords: Regresión múltiple, Mínimos cuadrados ordinarios, estimación matricial.

Regresión lineal múltiple

Abstract

This didactic material aims to present the estimation of the coefficients of a multiple linear regression model in matrix form.

Keywords: Multiple linear regression model, Ordinary Least Square

El problema

- Se desea “explicar” o analizar el comportamiento de una variable aleatoria en términos de un conjunto de variables $x_1, x_2, \dots, x_k$

- El mejor modelo es:

$$y = E(y|x) + u$$

donde u es un término de error

- Usualmente se asume

$$E(y|x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

La estimación

- Dada una muestra aleatoria de tamaño n se desea estimar $\beta_0, \beta_1, \ldots, \beta_k$

- La estimación se puede llevar a cabo por el método de mínimos cuadrados ordinarios
- $\min_{\{\beta_0,\ldots,\beta_k\}} \sum_i u_i^2 = \sum_i (y_i - \beta_0 - \cdots - \beta_k x_{ki})^2 \qquad (1)$

- Para resolver el problema se deriva (1) respecto a $\beta_0, \beta_1, \ldots, \beta_k$. De esto se obtiene un sistema de $k+1$ ecuaciones, el cual debe ser resuelto para las $k+1$ variables.

En forma matricial

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

$$Y_{nx1} = X_{nx(k+1)}\, \beta_{(k+1)x1} + U_{nx1}$$

La ecuación (1) se escribe como: $\min_{\beta}(Y - X\beta)^T(Y - X\beta)$

Solución:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Varianza de los estimadores

$$\mathrm{Var}(\hat{\beta}) = Var[(X^TX)^{-1}X^TY]$$
$$= \sigma^2(X^TX)^{-1}$$

Para el caso de $k = 1$

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_i (x_i - \bar{X})^2}$$

$$\mathrm{Var}(\hat{\beta}_0) = \frac{\sigma^2 \sum_i x_i^2}{n \sum_i (x_i - \bar{X})^2}$$

Hallar la varianza de los estimadores para el caso $k = 2$

Varianza de los estimadores (solución para k=2)

- $$\mathrm{Var}\left(\hat{\beta}_1\right)=\frac{\sigma^2\sum_i\left(x_{2i}-\overline{X_2}\right)^2}{\sum_i\left(x_{1i}-\bar{X}_1\right)^2\sum_i\left(x_{2i}-\overline{X_2}\right)^2-\left(\sum_i\left(x_{1i}-\overline{X_1}\right)\ \ (x_{2i}-\overline{X_2})\right)^2}$$

- $$\mathrm{Var}\left(\hat{\beta}_2\right)=\frac{\sigma^2\sum_i\left(x_{1i}-\overline{X_1}\right)^2}{\sum_i\left(x_{1i}-\bar{X}_1\right)^2\sum_i\left(x_{2i}-\overline{X_2}\right)^2-\left(\sum_i\left(x_{1i}-\overline{X_1}\right)\ \ (x_{2i}-\overline{X_2})\right)^2}$$

- $$\mathrm{Var}\left(\hat{\beta}_0\right)=\sigma^2\left(\frac{1}{n}+\frac{\overline{X_1^2}\sum_i(x_{2i}-\overline{X_2})^2+\overline{X_2^2}\sum_i(x_{1i}-\overline{X_1})^2-2\overline{X_1}\,\overline{X_2}\sum_i(x_{1i}-\overline{X_1})(x_{2i}-\overline{X_2})}{\sum_i\left(x_{1i}-\bar{X}_1\right)^2\sum_i\left(x_{2i}-\overline{X_2}\right)^2-\left(\sum_i\left(x_{1i}-\overline{X_1}\right)\ \ (x_{2i}-\overline{X_2})\right)^2}\right)$$

Referencias bibliográficas

- Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.