Modelo de Regresión Lineal

Supuestos del modelo

Oscar Gálvez-Soriano

University of Houston
Department of Economics

October, 2021

Modelo de Regresión Lineal

Considere el siguiente modelo de regresión lineal, donde y_i es la variable dependiente y $\boldsymbol{X_i} = (x_{i1}, \; ... \; , \; x_{ik})$ son las variables explicativas.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

$$y = \boldsymbol{X\beta} + \varepsilon$$

Supuestos del Modelo

- *S1. Lineal en Parámetros.*

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

- *S2. Identificabilidad.*

La matriz $\boldsymbol{X}$ tiene rango completo (no tiene multicolinelidad perfecta y hay al menos tantas observaciones como variables explicativas.

Supuestos del Modelo

- *S3. Ortogonalidad.*

La matriz $\boldsymbol{X}$ es no estocástica y el término de error tiene media cero.

$$E\left(\varepsilon_i\right) = 0$$

- *S4. Errores esféricos* (homoscedasticidad y no correlación lineal).

$$Var\left(\varepsilon_i\right) = \sigma^2$$

$$Cov\left(\varepsilon_i, \varepsilon_j\right) = 0 \quad i \neq j$$

- S5. *Normalidad de los errores.*

Propiedades para muestras finitas

Sea b el estimador MCO de β,

$$b = \left(X'X\right)^{-1} X'y$$

- El estimador b es insesgado, $E\left(b\right) = \beta$.

$$b = \left(X'X\right)^{-1} X'y = \left(X'X\right)^{-1} X'\left(X\beta + \varepsilon\right) =$$

$$b = \left(X'X\right)^{-1} X'X\beta + \left(X'X\right)^{-1} X'\varepsilon = \beta + \left(X'X\right)^{-1} X'\varepsilon$$

$$E\left(b\right) = \beta + \left(X'X\right)^{-1} X'E\left(\varepsilon\right) = \beta$$

Propiedades asintóticas

No necesitamos asumir normalidad si n es grande. Tenemos que $plim\, b = \beta$. El estimador b es consistente.

$$b = \beta + \left(X^{'}X\right)^{-1} X^{'}\varepsilon = \beta + \left(\sum_{i=1}^{n} x_i^{'} x_i\right)^{-1} \sum_{i=1}^{n} x_i^{'}\varepsilon_i$$

$$b = \beta + \left(\frac{1}{n}\sum_{i=1}^{n} x_i^{'} x_i\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i^{'}\varepsilon_i = \beta + \left[E\left(x_i^{'}x_i\right)\right]^{-1} E\left(x_i^{'}\varepsilon_i\right)$$

Este último paso es cierto por la Ley de los Grandes Números. También tenemos que $E\left(x_i^{'}x_i\right)$ es positiva definida por el S2. Finalmente, $E\left(x_i^{'}\varepsilon_i\right) = 0$ por el S3. Entonces:

$$plim\, b = \beta$$

Ortogonalidad

Hasta ahora hemos asumido que $\boldsymbol{X}$ es no-estocástica. Sin embargo, es más razonable asumir que tanto $\boldsymbol{X}$ como y provienen del mismo proceso generador de datos (DGP). ¿Qué podemos decir ahora de la Ortogonalidad?

- *Supuesto Fuerte.*

$$E\left(\varepsilon_i|x_i\right) = 0 \;\Rightarrow\; E\left(\varepsilon_i\right) = 0$$

Tabién tenemos que,

$$E\left(\varepsilon_i|x_i\right) = 0 \;\Rightarrow\; E\left(x_i^{'}\varepsilon_i\right) = 0 \;\Rightarrow\; Cov(x_i, \varepsilon_i) = 0$$

- *Supuesto Débil.*

$$E\left(x_i^{'}\varepsilon_i\right) = 0$$

Regresores endógenos

- Las propiedades de insesgamiento y consistencia dependen del supuesto de ortogonalidad.
- Vamos a decir que una variable es endógena si está correlacionada con el término de error. Y diremos que es exógena si no lo está.
- Algunas razones para la endogeneidad son:

1. Variables omitidas.
2. Error de medición.
3. Simultaneaidad (reverse causality).

Si una o más variables explicativas son endógenas, el supuesto crucial de $E\left(x_i'\varepsilon_i\right) = 0$ no se cumple.

Práctica en STATA

Más escolaridad resulta en mejores salarios? Una investigadora obtiene datos de la Current Population Survey (CPS) de 1976, la cual es representativa para los hogares de EUA. Ella obtiene una muestra de 526 trabajadores.

$$ln(\omega_i) = \alpha + \beta \cdot educ_i + \varepsilon_i$$

$$\hat{\beta} = \frac{cov(x,y)}{var(x)} = \frac{cov(educ,\omega)}{var(educ)}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x} = \bar{\omega} - \hat{\beta} \cdot \overline{educ}$$