Fundamentos de econometría

Formulario · Primavera 2021

Carlos Lezama

EST · 21104
ITAM

1. Fundamentos estadísticos de econometría

Definición 1.1 (Error cuadrático medio). *Definimos el **error cuadrático medio** como sigue:*

$$\mathrm{ECM}(\hat{\theta}) = \mathbb{E}_\theta\left[\left(\hat{\theta}-\theta\right)^2\right] = \mathrm{Var}(\hat{\theta}) + \mathrm{Bias}^2(\hat{\theta},\theta).$$

*Este mide la diferencia en media cuadrada entre nuestros valores estimados y el real. Asimismo, podemos definir el **error cuadrático medio** de una variable X respecto a alguna constante c como sigue:*

$$\mathbb{E}\left[(X-c)^2\right] = \sigma_{\mathrm{x}}^2 + (c-\mu_{\mathrm{x}})^2.$$

Teorema 1.1: Error cuadrático mínimo

El valor que minimiza $\mathbb{E}\left[(X-c)^2\right]$ es μ_{x}.

Corolario 1.1: Desigualdad de Markov I

Sean φ una función monótona creciente y no negativa para los reales no negativos, X una variable aleatoria, $a \geq 0$ y $\varphi(a) > 0$, entonces:

$$\mathbb{P}\left(|X| \geq a\right) \leq \frac{\mathbb{E}\left[\varphi\left(|X|\right)\right]}{\varphi(a)}.$$

Corolario 1.2: Desigualdad de Markov II

Sea X una variable aleatoria y $a > 0$, entonces:

$$\mathbb{P}\left(|X| \geq a\right) \leq \frac{\mathbb{E}\left[|X|^r\right]}{a^r}.$$

Teorema 1.2: Desigualdad de Chebyshev

Sean X una variable aleatoria con $\mathbb{E}(X) = \mu$ y $\mathrm{Var}(X) = \sigma^2$, y $k > 0$, entonces:

$$\mathbb{P}\left(\|X-\mu\|_\alpha \geq k\sigma\right) \leq \frac{1}{k^2},$$

o bien,

$$\mathbb{P}\left(\|X-\mu\|_\alpha \geq k\right) \leq \frac{\sigma^2}{k^2};$$

donde $\|\cdot\|_\alpha$ es la norma α.

Observación 1.1

Sean $X_1, X_2, \ldots, X_n$ una colección de $n \in \mathbb{N}$ variables aleatorias, podemos expresar las funciones de densidad conjunta en términos de densidades condicionales de la siguiente manera:

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &= f(x_1 \mid x_2, x_3, \ldots, x_n) \cdot f(x_2, x_3, \ldots, x_n) \\ &= f(x_1 \mid x_2, \ldots, x_n) \cdot f(x_2 \mid x_3, \ldots, x_n) \cdot \cdots \cdot f(x_{n-1} \mid x_n) \cdot f(x_n). \end{aligned}$$

Teorema 1.3: Ley de esperanzas iteradas

Sean X y Y dos variables aleatorias tales que $\mathbb{E}\left[X\right]$ está definida y ambas están en el mismo espacio de probabilidad, entonces:

$$\mathbb{E}\left[X\right] = \mathbb{E}\left[\mathbb{E}\left[X \mid Y\right]\right].$$

Proposición 1.1

Por la **ley de esperanzas iteradas**, tenemos:

$$\mathrm{Var}(X) = \mathbb{E}\left[\mathrm{Var}(X \mid Y)\right] + \mathrm{Var}\left(\mathbb{E}[Y \mid X]\right).$$

Proposición 1.2

$$\mathbb{E}[XY] = \mathbb{E}\left[X\mathbb{E}[Y \mid X]\right].$$

Proposición 1.3

Sea $\varepsilon = Y - \mathbb{E}[Y \mid X]$, entonces:

(I) $\mathbb{E}[\varepsilon \mid X] = 0$.
(II) $\mathrm{Var}(\varepsilon \mid X) = \sigma_{\mathrm{y|x}}^2$.
(III) $\mathbb{E}[\varepsilon] = 0$.
(IV) $\mathrm{Var}(\varepsilon) = 0$.
(V) $\mathrm{Cov}(h(X), \varepsilon) = 0$.

Proposición 1.4

$$\underset{h(X)}{\arg\min}\, \mathbb{E}\left[\left(Y - h(X)\right)^2\right] = \mathbb{E}\left[Y \mid X\right].$$

Definición 1.2 (Mejor predictor lineal). *Dado que $\mathbb{E}[Y \mid X]$ puede ser una función no lineal bastante complicada — excepto en el caso normal bivariado, consideramos $\mathbb{E}^*[Y \mid X]$ el **mejor predictor lineal** (BLP, por sus siglas en inglés) y lo definimos como sigue:*

$$\mathbb{E}^*\left[Y \mid X\right] = \alpha + \beta X,$$

tal que

$$\alpha = \mu_{\mathrm{y}} - \beta\mu_{\mathrm{x}} \qquad \beta = \frac{\sigma_{\mathrm{xy}}}{\sigma_x^2}.$$

Teorema 1.4

Si la esperanza condicional es lineal, esta coincide con el mejor predictor lineal.

Definición 1.3 (Independencia en distribución). *Decimos que, en un conjunto finito de n variables aleatorias $\{X_1, \ldots, X_n\}$, $X_1, \ldots, X_n$ son **mutuamente independientes** si:*

$$F_{\mathrm{x}_1,\ldots,\mathrm{x}_n}(x_1, \ldots, x_n) = \prod_{k=1}^{n} F_{\mathrm{x}_k}(x_k), \quad \forall x_1, \ldots, x_n.$$

Definición 1.4 (Independencia en media). *Decimos que una variable aleatoria Y es **independiente en media** respecto a otra X si:*

$$\mathbb{E}[Y \mid X] = \mathbb{E}[Y].$$

Definición 1.5 (Independencia en covarianza). *Decimos que dos variables aleatorias X y Y son **independientes en covarianza** si*

$$\mathrm{Cov}(X, Y) = 0.$$

Teorema 1.5

Independencia en distribución implica independencia en medias, y esta, a su vez, implica independencia en covarianzas; pero no en el orden inverso.

Proposición 1.5

Si Y es independiente en media respecto a X, entonces:

$$\mathbb{E}[X^r Y] = \mathbb{E}[X^r]\mathbb{E}[Y], \quad \forall r \in \mathbb{R}.$$

Proposición 1.6

Si Y es independiente en distribución respecto a X, entonces:

$$\mathbb{E}[h(X)Y^s] = \mathbb{E}[h(X)]\mathbb{E}[Y^s]$$

y, por lo tanto, también se cumple

$$\mathbb{E}[X^r Y^s] = \mathbb{E}[X^r]\mathbb{E}[Y^s]$$

para cualquier función $h(X)$ y todo r y s.

Proposición 1.7

Sean $\varepsilon = Y - \mathbb{E}[Y \mid X]$ y $U = Y - \mathbb{E}^*[Y \mid X]$, entonces ε es independiente en media respecto a X y U no está correlacionado con X.

Teorema 1.6: Desigualdad Cauchy-Schwartz

Sea $\rho = \dfrac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}}$, entonces $0 \leq \rho^2 \leq 1$.

Proposición 1.8

Sea $X \sim \mathcal{N}(\mu, \sigma^2)$, entonces, para toda p entera no negativa, los momentos centrados son los siguientes:

$$\mathbb{E}[(X-\mu)^p] = \begin{cases} 0, & \text{si } p \text{ es impar;} \\ \sigma^p (p-1)!!, & \text{si } p \text{ es par.} \end{cases}$$

Nota: $n!! = \displaystyle\prod_{k=0}^{\lceil \frac{n}{2} \rceil - 1} (n-2k)$.

$$\text{i.e.} \quad (n-1)!! = (n-1)(n-3)(n-5)\cdots 3 \cdot 1,$$

para toda n par.

Proposición 1.9

En el caso normal, la ausencia de correlación equivale a independencia estocástica.

Proposición 1.10

Sea $g(Y \mid X) \sim \mathcal{N}(\mathbb{E}[Y \mid X], \operatorname{Var}(Y \mid X))$, entonces:

$$\mathbb{E}[Y \mid X] = \mu_{\mathrm{y}} + \rho \frac{\sigma_{\mathrm{y}}}{\sigma_{\mathrm{x}}}(X - \mu_{\mathrm{x}}),$$

y

$$\operatorname{Var}(Y \mid X) = \sigma_{\mathrm{y}}^2 (1-\rho)^2.$$

Nótese que la esperanza condicional es lineal y la varianza es constante.

Observación 1.2

En el caso normal, la función de regresión coincide con el mejor predictor lineal tal que, si $Y = \mathbb{E}[Y \mid X] + \varepsilon$ y $\mathbb{E}[\varepsilon \mid X] = 0$, $Y = \alpha + \beta X + \varepsilon$ con $\alpha = \mu_{\mathrm{y}} - \beta \mu_{\mathrm{x}}$ y $\beta = \dfrac{\sigma_{\mathrm{xy}}}{\sigma_{\mathrm{x}}^2}$.

Proposición 1.11

La distribución condicional de Y dado X es normal.

$$Y \mid X \sim \mathcal{N}(\alpha + \beta X, \sigma^2).$$

2. Estimación

Teorema 2.1

Sea X una muestra aleatoria finita con n observaciones, entonces, $x_1, x_2, \ldots, x_n$ son independientes e idénticamente distribuídas y, por lo tanto, la densidad conjunta de dicha muestra aleatoria es:

$$F_{\mathrm{x}_1,\ldots,\mathrm{x}_n}(x_1,\ldots,x_n) = \prod_{k=1}^{n} F_{\mathrm{x}_k}(x_k).$$

Definición 2.1 (Estadístico muestral). *Sea $T_n = h(X)$ una función escalar de una muestra aleatoria, entonces, este es un **estadístico muestral**.*

Observación 2.1

Todo estadístico muestral T_n es una variable aleatoria porque su valor es determinado por el resultado de un experimento. Asimismo, a la distribución de probabilidad de T_n se le conoce como *distribución muestral*, y está completamente determinada por $h(\cdot)$, $f(\mathrm{x})$ y n.

Proposición 2.1

La media muestral satisface las siguientes propiedades:

- Si $X \sim \mathrm{Ber}(p)$, entonces, $n\bar{X} \sim \mathrm{Bin}(n,p)$.
- Si $X \sim \mathcal{N}(\mu, \sigma^2)$, entonces, $\bar{X} \sim \mathcal{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$.
- Si $X \sim \mathrm{Exp}(\lambda)$, entonces, $k\lambda\bar{X} \sim \chi^2(k)$ con $k = 2n$.

Teorema 2.2: de la media muestral

Dada un muestra aleatoria de tamaño n y cualquier población con $\mathbb{E}[X] = \mu$ y $\mathrm{Var}(X) = \sigma^2$, la media muestral $\bar{X}$ tiene esperanza μ y varianza $\dfrac{\sigma^2}{n}$.

Definición 2.2 (Momento muestral centrado).

$$M_r = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^r.$$

Definición 2.3 (Momento muestral no centrado).

$$M_r' = \frac{1}{n}\sum_{i=1}^{n} X_i^r.$$

Definición 2.4 (Momento muestral centrado en la media poblacional).

$$M_r^* = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu\right)^r.$$

*Nótese que este **no** es un estimador muestral, pues, se requiere un parámetro poblacional. Sin embargo, les será útil para algunas demostraciones.*

Observación 2.2

El *teorema de la media muestral* no puede aplicarse a momentos centrados alrededor de la media muestral.

Proposición 2.2

Sean X una muestra aleatoria finita con n observaciones y $\bar{X}$ la media muestral, entonces:

$$\mathrm{Cov}(X_i, \bar{X}) = \mathrm{Cov}(\bar{X}, \bar{X}) = \frac{\mathrm{Var}(X)}{n}, \quad \forall X_i \in X.$$

Proposición 2.3

$M_2 \leq M_2^*$ en cualquier muestra, donde M_2 es la varianza muestral. Asimismo, $\mathbb{E}[M_2] \approx \mu_2$ y $\mathrm{Var}(M_2) \approx \mathrm{Var}(M_2^*)$ si n es grande.

Distribuciones derivadas de la normal

- Sean $Z_i \sim \mathcal{N}(0,1)$ variables aleatorias independientes, entonces:

$$\sum_{i=1}^{k} Z_i^2 \sim \chi^2(k).$$

- Sean $Z \sim \mathcal{N}(0,1)$ y $W \sim \chi^2(k)$ dos variables aleatorias independientes, entonces:

$$\frac{Z}{\sqrt{\left(\frac{W}{k}\right)}} \sim t(k).$$

Proposición 2.4: Población normal estándar

Dadas una variable aleatoria $X \sim \mathcal{N}(0,1)$ y una muestra aleatoria $X_1, X_2, \ldots, X_n$ de tamaño n con media muestral $\bar{X}$ y varianza muestral S^2, se cumple:

(I) $\sqrt{n}\bar{X} \sim \mathcal{N}(0,1)$.
(II) $nS^2 \sim \chi^2(n-1)$.
(III) $\bar{X}$ y S^2 son independientes.
(IV) $\dfrac{\sqrt{(n-1)}\bar{X}}{S} \sim t(n-1)$

Proposición 2.5: Población normal general

Dadas una variable aleatoria $X \sim \mathcal{N}(\mu, \sigma^2)$ y una muestra aleatoria $X_1, X_2, \ldots, X_n$ de tamaño n con media muestral $\bar{X}$ y varianza muestral S^2, se cumple:

(I) $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.
(II) $\dfrac{nS^2}{\sigma^2} \sim \chi^2(n-1)$.
(III) $\bar{X}$ y S^2 son independientes.
(IV) $\dfrac{\sqrt{(n-1)}(\bar{X}-\mu)}{S} \sim t(n-1)$

Teorema 2.3

Dada una muestra aleatoria de tamaño n, $\{(X_i, Y_i)\}_{i=1}^{n}$ son observaciones independientes e idénticamente distribuídas tales que $(X_i, Y_i) \in \mathcal{X} \times \mathcal{Y}$.
Nota: esto no implica independencia entre vectores.

Observación 2.3

Decimos que T es un estimador *insesgado* si y solo si $\mathbb{E}[T - \theta] = 0$ para todo θ.

Definición 2.5 (Estimador insesgado de varianza mínima). *Decimos que T es un **estimador insesgado de varianza mínima** si y solo si:*

(I) $\mathbb{E}[T - \theta] = 0$ *para todo* θ.
(II) $\mathrm{Var}(T) \leq \mathrm{Var}(T^*)$, *para todo* T^*, *tal que* $\mathbb{E}[T^* - \theta] = 0$.

Teorema 2.4

En una muestra aleatoria de tamaño n, de cualquier población, la media muestral es el estimador lineal insesgado de menor varianza respecto a la media poblacional.

3. Teoría asintótica

Sea T_n una sucesión de variables aleatorias con distribuciones acumuladas $G_n(t) = \mathbb{P}(T_n \le t)$ con esperanzas $\mathbb{E}[T_n]$ y varianzas $\mathrm{Var}(T_n)$:

Definición 3.1 (Convergencia en probabilidad). *Decimos que T_n **converge en probabilidad**† a una constante c si y solo si:*

$$\lim_{n\to\infty} G_n(t) = 0, \quad \forall t < c,$$
$$\lim_{n\to\infty} G_n(t) = 1, \quad \forall t \ge c.$$

O bien,

$$\lim_{n\to\infty} \mathbb{P}(|T_n - c| \ge \varepsilon) = 0, \quad \forall \varepsilon > 0.$$

† $T_n \xrightarrow{p} c.$

Definición 3.2 (Convergencia en media r). *Decimos que T_n **converge en media r**† a una constante c si y solo si:*

$$\lim_{n\to\infty} \mathbb{E}\left[(T_n - c)^r\right] = 0.$$

† $T_n \xrightarrow{L^r} c.$

Definición 3.3 (Convergencia en distribución). *Decimos que T_n **converge en distribución**† a $G(\cdot)$ si y sólo si:*

$$\lim_{n\to\infty} G_n(t) = G(t), \quad \forall t.$$

† $T_n \xrightarrow{d} G(\cdot).$

Corolario 3.1

Sea T_n una sucesión de variables aleatorias tal que $\lim_{n\to\infty} \mathbb{E}[T_n] = c$ y $\lim_{n\to\infty} \mathrm{Var}(T_n) = 0$, entonces, T_n converge en *media cuadrática*† a c.

† $T_n \xrightarrow{L^2} c.$

Corolario 3.2

Si $T_n \xrightarrow{L^r} c$, entonces, $T_n \xrightarrow{p} c$.

Teorema 3.1: Ley de los grandes números

En una muestra aleatoria de cualquier población con esperanza μ y varianza σ^2, la media muestral converge en probabilidad a la media poblacional.

$$\text{i.e.} \quad \bar{X} \xrightarrow{p} \mu.$$

Caso multivariado:

$$(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_n) \xrightarrow{p} (\mu_{\mathrm{x}_1}, \mu_{\mathrm{x}_2}, \dots, \mu_{\mathrm{x}_n}).$$

Teorema 3.2: Teorema central del límite

En una muestra aleatoria de cualquier población con esperanza μ y varianza σ^2, la media muestral estandarizada tiende en distribución a una normal estándar.

$$\text{i.e.} \quad \frac{\sqrt{n}\,(\bar{X} - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0,1).$$

Equivalentemente,

$$\sqrt{n}\,(\bar{X} - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

Caso multivariado:

$$\sqrt{n}\,(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}).$$

Proposición 3.1

Decimos que la *distribución asintótica* de la media muestral es $\mathcal{N}(\mu, \sigma^2/n)$.

$$\text{i.e.} \quad \bar{X} \overset{A}{\sim} \mathcal{N}(\mu, \sigma^2/n).$$

Proposición 3.2

El análogo muestral al **mejor predictor lineal** es:

$$\mathbb{E}^*\left[Y \mid X\right] = A + BX,$$

donde

$$A = \bar{Y} - B\bar{X} \qquad B = \frac{S_{\mathrm{xy}}}{S_{\mathrm{x}^2}}.$$

Definición 3.4 (Consistencia). *Decimos que T_n es un estimador consistente de θ si y solo si:*

$$T_n \xrightarrow{p} \theta.$$

Teorema 3.3: de Slutsky

Sean T_n, V_n y W_n sucesiones variables aleatorias, $h(\cdot)$ una función y c una constante — tal que estas últimas dos no dependen de n; lo siguiente se cumple:

(i) Si $T_n \xrightarrow{p} c$ y $h(T_n)$ es continua en c, entonces, $h(T_n) \xrightarrow{p} h(c)$.

(ii) Si $V_n \xrightarrow{p} c_1$, $W_n \xrightarrow{p} c_2$ y $h(V_n, W_n)$ es continua en (c_1, c_2), entonces, $h(V_n, W_n) \xrightarrow{p} h(c_1, c_2)$.

(iii) Si $V_n \xrightarrow{p} c$ y W_n tiene una distribución límite, entonces, la distribución límite de $V_n + W_n$ es la misma que la correspondiente a $c + W_n$.

(iv) Si $V_n \xrightarrow{p} c$ y W_n tiene una distribución límite, entonces, la distribución límite de $V_n W_n$ es la misma que la correspondiente a cW_n.

(v) Si $\sqrt{n}\,[T_n - \theta] \xrightarrow{d} \mathcal{N}(0, \phi^2)$ y $h(T_n)$ es continuamente diferenciable en θ, entonces:

$$\sqrt{n}\,[h(T_n) - h(\theta)] \xrightarrow{d} \mathcal{N}\left(0, \phi^2 \cdot \left[h'(\theta)\right]^2\right),$$

tal que

$$T_n \overset{A}{\sim} \mathcal{N}\left(h(\theta), \left[h'(\theta)\right]^2 \phi^2/n\right).$$

Caso multivariado: si $\sqrt{n}\,[\boldsymbol{T}_n - \boldsymbol{\theta}] \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma})$ y $h(\boldsymbol{T}_n)$ es continuamente diferenciable en $\boldsymbol{\theta}$, entonces:

$$\sqrt{n}\,[h(\boldsymbol{T}_n) - h(\boldsymbol{\theta})] \xrightarrow{d} \mathcal{N}\left(0, \nabla h(\boldsymbol{\theta})^T \cdot \boldsymbol{\Sigma} \cdot \nabla h(\boldsymbol{\theta})\right),$$

tal que

$$\boldsymbol{T}_n \overset{A}{\sim} \mathcal{N}\left(h(\boldsymbol{\theta},), \nabla h(\boldsymbol{\theta})^T \cdot \frac{1}{n}\boldsymbol{\Sigma} \cdot \nabla h(\boldsymbol{\theta})\right).$$

Definición 3.5 (Mejor estimador asintóticamente normal). *Decimos que T_n es el **mejor estimador asintóticamente normal** si y solo si:*

(i) $T_n \overset{A}{\sim} \mathcal{N}(\theta, \sigma^2/n)$.

(ii) $\sigma^2 \le \sigma^{*2}$, *para todo T_n^*, tal que $T_n^* \xrightarrow{d} \mathcal{N}(\theta, \sigma^{*2}/n)$.*

*A este criterio también se le conoce como **eficiencia asintótica**.*

Definición 3.6 (Intervalo de confianza). *Dado un conjunto de observaciones $x_1, \dots, x_n$ de las variables aleatorias $X_1, \dots, X_n$, y sean θ nuestro parámetro de interés y $\gamma \in (0,1)$. Si existen estadísticos muestrales $L_n = g(X_1, \dots, X_n)$ y $U_n = h(X_1, \dots, X_n)$ tales que:*

$$\mathbb{P}(L_n < \theta < U_n) = \gamma, \quad \forall \theta;$$

*entonces, decimos que (l_n, u_n) es un **intervalo de confianza** $\gamma \times 100\,\%$ del parámetro θ, donde $l_n = g(x_1, \dots, x_n)$ y $u_n = h(x_1, \dots, x_n)$.*

Algunas distribuciones asintóticas

(I) $S^2 \overset{A}{\sim} \mathcal{N}\left(\sigma^2, \dfrac{\mu_4 - \mu_2^2}{n}\right)$.

(II) $S_{\mathrm{xy}} \overset{A}{\sim} \mathcal{N}\left(\sigma_{\mathrm{xy}}, \dfrac{\mu_{22} - \mu_{11}^2}{n}\right)$.

(III) $(\bar{X}, \bar{Y}) \overset{A}{\sim} \mathcal{N}\left(\mu_{\mathrm{x}}, \mu_{\mathrm{y}}; \sigma_{\mathrm{x}}^2/n, \sigma_{\mathrm{y}}^2/n, \sigma_{\mathrm{xy}}/n\right)$.

(IV) $\dfrac{\bar{X}}{\bar{Y}} \overset{A}{\sim} \mathcal{N}\left(\theta, \dfrac{\phi^2}{n}\right)$ tal que $\theta = \dfrac{\mu_{\mathrm{x}}}{\mu_{\mathrm{y}}}$ y

$\phi^2 = \left(\dfrac{1}{\mu_{\mathrm{y}}}\right)^2 \left(\sigma_{\mathrm{x}}^2 + \theta^2\sigma_{\mathrm{y}}^2 - 2\theta\sigma_{\mathrm{xy}}\right)$.

(V) $\dfrac{S_{\mathrm{xy}}}{S_{\mathrm{x}}^2} \overset{A}{\sim} \mathcal{N}\left(\beta, \dfrac{\phi^2}{n}\right)$ donde $\beta = \dfrac{\sigma_{\mathrm{xy}}}{\sigma_{\mathrm{x}}^2}$ y

$\phi^2 = \left(\dfrac{1}{\mu_{20}^2}\right)\left(\mu_{22} + \beta^2\mu_{40} - 2\beta\mu_{31}\right)$.

4. Inferencia en el modelo lineal