Fundamentos de Econometría

Ignacio Lobato

ITAM

Estimación de Parámetros

- En este capítulo nos enfocaremos en inferir hechos de la población a partir de una muestra aleatoria.
- En particular, tendremos una muestra aleatoria $(y_1, y_2, y_3, ..., y_n)'$, que proviene de una población desconocida $Y \sim f(y)$.
- Estamos interesados en alguna característica de la población (un parámetro θ). Nuestro objetivo es estimar dicho parámetro.

- La estimación será una función de la muestra. La pregunta en esta sección será cuál función de la muestra debemos elegir para estimar.
- $y = (y_1, y_2, ..., y_n)'$ será una observación del vector aleatorio $Y = (Y_1, Y_2, ..., Y_n)'$.
- Para una función h(Y), la estimación que calculamos será $t = h(y_1, y_2, ..., y_n)$, la cual será una observación de la variable aleatoria T=h(Y). La variable aleatoria T será referida como el estimador y t será la estimación.

Principio del Análogo Muestral

- Quizás la forma más natural de seleccionar un estimador es a través del principio de analogía. Un parámetro es una característica de la población. Para estimar esto, usamos la característica análoga en la muestra.

  Algunas aplicaciones de este principio son los siguientes:

  1. Estimar algún momento poblacional usando el respectivo momento muestral. Para $\theta = \mu$, usar $T = \overline{Y}$. Para $\theta = \sigma^2$, usar $T = S_y^2$.

  2. Estimar una función de momentos poblacionales usando la respectiva función de momentos muestrales. Para $\beta = \sigma_{xy}/\sigma_x^2$, usar $B = S_{xy}/S_x^2$.

  3. Para estimar el BLP $E^*(Y/X) = \alpha + \beta X$, usar la línea que minimiza la media de las desviaciones al cuadrado en la muestra: $\hat{Y}$=A+BX.

Principio del Análogo Muestral

- Adoptar el principio del análogo muestral como punto de partida para encontrar estimadores nos lleva a formularnos algunas preguntas.
- ¿Cuán confiables son los estimadores obtenidos por el principio del análogo muestral? ¿Qué debemos hacer si dichos estimadores no son factibes de obtener o no son confiables? ¿Qué hacer cuando existen varios estimadores para un mismo parámetro?
- Para responder estas preguntas, primero definimos criterios para evaluar a un estimador.

Criterios para un Estimador-ECM

- Sea $T = h(Y_1, Y_2, ..., Y_n)$ un estadístico muestral con una pdf g(t) y momentos E(T), V(T), etc.
- Una medida natural de distancia entre la variable aleatoria T y el parámetro θ es el **Error Cuadrático Medio (ECM)**, $E(T-\theta)^2$.
- El punto de referencia de distancia en este caso es un punto fijo (el parámetro) y ya no una variable aleatoria.
- Nuevamente, lo deseable es que el ECM sea pequeño. Recordemos que el ECM puede escribirse de la siguiente forma:

$$E(T-\theta)^2 = V(T) + [E(T)-\theta]^2$$

- Defínase el **Sesgo del Estimador** T de θ como:

$$E(T) - \theta = E(T-\theta)$$

Por lo tanto, el ECM del estimador T es igual a su varianza más el cuadrado de su sesgo.

Criterios para un Estimador-Insesgadez

- Note que en general el sesgo y la varianza de un estimador depende del parámetro desconocido θ y por lo tanto no es factible encontrar una T que minimice el ECM para todo θ. No obstante, es deseable tener un estimador T con un ECM pequeño.
- También parece deseable tener un estimador para el cual la desviación esperada respecto a ese parámetro es cero. Esto nos lleva a la siguiente definición:

  Insesgadez: T es un estimador insesgado de θ si y solo si $E(T-\theta)=0$, para todo θ.
- Note que para un estimador insesgado T, ECM=V(T), lo cual conduce a un criterio asociado a la menor varianza de estimadores insesgados que se detalla a continuación:

Criterios para un Estimador-MVUE

- T es un **Estimador Insesgado de Mínima Varianza (MVUE)** de θ si y solo si:

  (i) $E(T-\theta)=0$ para todo θ, y

  (ii) $V(T) \leq V(T^*)$ para todo T^*, tal que $E(T^*-\theta)=0$

- Se puede tener un estimador MVUE incluso cuando no se está minimizando el ECM. Usualmente se habla de los estimadores insesgados de varianza mínima restringiendo el conjunto posible de estimadores (por ejemplo los lineales).

Estimación de la Media Poblacional

- Supongamos que tenemos una muestra aleatoria sobre la variable Y, donde $E(Y) = \mu$ y $V(Y) = \sigma^2$, los cuales son desconocidos.
- Para estimar μ, el principio de analogía sugiere que usemos la media muestral $\overline{Y}$ y note que:

$$E(\overline{Y}) = E(Y) = \mu$$

Por lo tanto, $\overline{Y}$ es un estimador insesgado de μ. Sin embargo, hay muchos estimadores lineales insesgados de μ. Así, sea $T = \sum_i c_i Y_i$, donde las c's son constantes. Entonces:

$$E(T) = E(\sum_i c_i Y_i) = \sum_i E(c_i Y_i) = \mu \sum_i c_i,$$

$$V(T) = \sum_i c_i^2 V(Y_i) = \sigma^2 \sum_i c_i^2$$

Estimación de la Media Poblacional

- Por lo tanto, cualquier función lineal de las Y_i's con intercepto cero y con una suma de pendientes igual a 1 será insesgada. Para escoger uno de ellos, escogemos las c_i's que minimizan $\sum_i c_i^2$ sujeto a que $\sum_i c_i = 1$. Sea:

$$Q = \sum_i c_i^2 - \lambda(\sum_i c_i - 1),$$

Donde λ es el multiplicador de Lagrange. Luego:

$$\partial Q/\partial c_i = 2c_i - \lambda \quad \text{para i=1,..,n}$$

$$\partial Q/\partial \lambda = -(\sum_i c_i - 1).$$

Igualando las derivadas a cero, obtenemos que $\sum_i c_i = 1$ y que $c_i = \lambda/2$. Por lo tanto, $\lambda = 2/n$ y $c_i = 1/n$ para todo i=1,...,n. Por lo tanto la elección óptima a este problema de minimización es:

$$T = \sum_i (1/n) Y_i = \overline{Y}$$

Estimación de la Media Poblacional

- Lo anterior nos lleva al siguiente teorema:

  Teorema: En una muestra aleatoria de tamaño n, de cualquier población, la media muestral es el estimador lineal insesgado de menor varianza (MVLUE) de la media poblacional.

- Recuerde que este resultado solo toma en cuenta a estimadores lineales.

- Note que este teorema puede extenderse para cualquier momento poblacional no centrado.

Estimación de la Varianza Poblacional

- Ahora supongamos que $\theta = E(Y - \mu_y)^2 = \sigma^2$ con μ_y desconocido. El principio de analogía sugiere $M_2 = S^2$.
- No obstante, anteriormente vimos que la varianza muestral no puede ser interpretada como una media muestral. Y en efecto, este estimador no es insesgado pues:

$$E(S^2) = \sigma^2(1 - 1/n)$$

- De este modo, podemos definir la varianza muestral ajustada, la cual sí será insesgada:

$$S^{*2} = \sum_i (Y_i - \overline{Y})^2/(n-1) = nS^2/(n-1)$$

Note que $E(S^{*2}) = \sigma^2$ pero no se ha dicho nada respecto a si es el MVUE.

Estimación de la Covarianza y otros momentos

- Para estimar la covarianza poblacional, podemos emplear como estimador insesgado a la covarianza muestral $S_{xy}^* = nS_{xy}/(n-1)$ (no necesariamente es el MVUE).

- En el caso del máximo poblacional $\theta = max(Y)$, el principio de analogía sugiere $T = max(Y_1, Y_2, ..., Y_n)$ como estimador. Pero $T \leq \theta$ y $Pr(T = \theta) < 1$, por lo que $E(T) < \theta$. Por lo tanto, el estimador es sesgado y no hay una manera obvia de remover el sesgo.

- En el caso de la pendiente de la proyección lineal $\beta = \sigma_{xy}/\sigma_x^2$ (con medias poblacionales desconocidas), el principio de analogía sugiere $B = S_{xy}/S_x^2$ como estimador. En general este también será un estimador sesgado y no hay una manera obvia de remover dicho sesgo. Un caso especial en el cual este estimador es insesgado es cuando la esperanza condicional de Y respecto a X es lineal.

Criterios Asintóticos - Consistencia

- Como hemos visto, evaluar estimadores obtenidos por el principio de analogía en base a momentos exactos puede ser no conveniente, pues la distribución de estos estimadores puede depender de algunas características específicas de la población
- De este modo, tomar criterios aproximados como los asintóticos pueden resultar convenientes. Así, un primer criterio de evaluación de estimadores bajo este enfoque es el siguiente:

Consistencia: T_n es un estimador consistente de θ si y solo si:

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

Este concepto nos dice que a medida que el tamaño de la muestra se incrementa indefinidamente, la distribución del estimador se concentra completamente en el valor del parámetro.

Criterios Asintóticos - Consistencia

- Note que por la Ley de Grandes Números, la media muestral $\overline{Y}$ es un estimador consistente de μ_y, y que cualquier momento muestral no centrado es un estimador consistente de su respectivo momento poblacional no centrado.

- Asimismo, por los teoremas de Slutsky, S^2 será un estimador consistente de σ^2 y $B = S_{xy}/S_x^2$ será un estimador consistente de β.

- Naturalmente puede haber más de un estimador consistente para un mismo parámetro. Para discriminar entre estos, debemos analizar sus distribuciones asintóticas.

- Típicamente, los estimadores consistentes tendrán una distribución asintótica normal centrada en el parámetro que se quiere estimar.

- Para comparar entre distintos estimadores consistentes con distribución asintótica normal, utilizamos el siguiente criterio:

  Definición: T_n es el Mejor Estimador Asintóticamente Normal (BAN) de θ si y solo si:

  (i) $T_n \overset{a}{\sim} N(\theta, \phi^2/n)$, y

  (ii) $\phi^2 \leq \phi^{*2}$ para todo T_n^* tal que $T_n^* \xrightarrow{d} N(\theta, \phi^{*2}/n)$

- Note que el criterio BAN es la versión asintótica del MVUE.

- Este criterio también es referido como Eficiencia Asintótica.

Intervalos de Confianza

- Al momento de reportar la estimación de un parámetro, es buena idea brindar información respecto a la confiabilidad de dicho estimador.
- Una forma de hacer esto es indicar el rango bajo el cual varía de muestra a muestra y la medida usual de ello es la desviación o error estándar del estimador.
- Usualmente esta información es presentada en forma de intervalo de confianza.
- Supongamos que:

$$Y \sim N(\mu, \sigma^2/n) \quad \text{y} \quad Z = \sqrt{n}(\overline{Y} - \mu)/\sigma \sim N(0, 1)$$

Sea $A = \{| z | \leq 1{,}96\}$, Luego Pr(A)=0.95. Además note que :

$$A = \{| \overline{Y} - \mu | \leq 1{,}96\sigma/\sqrt{n}\} = \{\overline{Y} - 1{,}96\sigma/\sqrt{n} \leq \mu \leq \overline{Y} + 1{,}96\sigma/\sqrt{n}\}$$

Por lo tanto, μ se encuentra en el intervalo $\overline{Y} \pm 1{,}96\hat{\sigma}/\sqrt{n}$ con un 95 % de probabilidad. Dicho intervalo será un intervalo de confianza al 95 % para μ.

Intervalos de Confianza

- Para obtener un intervalo de confianza a un distinto nivel de confianza, solo se debe usar un nuevo punto de cut-off. Por ejemplo, al 10 % de confianza el cut-off será 1.645.
- Asimismo, note que estos intervalos no son computables directamente puesto que σ^2 es desconocido. Por ello, se computa un intervalo de confianza **aproximado**, en el cual se computa dicho intervalo reemplazando σ^2 por S^2, pues este último es consistente para σ^2.