Fundamentos de Econometría

Ignacio Lobato

ITAM

Muestra Aleatoria

- Anteriormente se ha discutido toda la información contenida en la población a través de las distribuciones de probabilidad.
- Ahora reorientamos el análisis hacia las muestras aleatorias, definidas como un conjunto de variables aleatorias extraídas de una misma población que se distribuyen de forma idéntica e independiente.
- Iniciamos el análisis con el caso de una distribución de probabilidad univariada para una variable aleatoria X. En particular, sean $X_1, X_2, ..., X_n$ un conjunto de variables aleatorias extraídas de la distribución poblacional de X. Es decir, son variables aleatorias que resultan de la realización de un mismo experimento repetido n veces de manera independiente.
- Así, el vector $X = (X_1.X_2, ..., X_n)$ es llamado muestra aleatoria de tamaño n de la variable X, o de la población de X, o de la distribución de probabilidad de X. Los valores que X toma serán denotados como $x = (x_1, x_2, ..., x_n)$.

Muestra Aleatoria

- Si $X = (X_1, X_2, ..., X_n)$ es una muestra aleatoria de X, luego $X_i' s$ son **idéntica e independientemente distribuidos**. Así, si f(x) es la función de distribución de X, entonces la densidad conjunta para la muestra aleatoria de X es:

$$f_n(x) = f_n(x_1, x_2, x_3, ..., x_n) = f_1(x_1)...f_n(x_n) = f(x_1)...f(x_n) = \prod_i f(x_i)$$

- Por ejemplo, si X se distribuye como una Bernoulli con parámetro ρ, entonces la distribución de probabilidad de X es:

$$f(x) = \rho^x (1 - \rho)^{1-x}$$

Cuando x toma 0 o 1 y $f(x) = 0$ cuando x toma algún otro valor. Luego la distribución para una muestra aleatoria de tamaño n será:

$$f_n(x) = \prod_i \left[\rho^{x_i} (1 - \rho)^{1-x_i}\right] = \left[\prod_i \rho^{x_i}\right]\left[\prod_i (1 - \rho)^{1-x_i}\right]$$

$$= \rho^{\sum_i x_i} (1 - \rho)^{n - \sum_i x_i}$$

Estadísticos Muestrales

- Sea $T_n = h(X_1, .., X_n) = h(X)$ una función escalar de una muestra aleatoria. Luego T_n es llamado **estadístico muestral** o estadístico. Los valores que $T_n = h(X)$ toman serán denotados por $t_n = h(x)$. Algunos ejemplos de estadísticos muestrales son los siguientes:
- La media muestral:

$$\overline{X} = (X_1 + ... + X_n)/n = (1/n) \sum_i X_i$$

- La varianza muestral:

$$S_x^2 = (1/n) \sum_i (X_i - \overline{X})^2$$

- Los momentos muestrales centrados y no centrados (con r entero no negativo):

$$M_r = (1/n) \sum_i (X_i - \overline{X})^r \quad \text{y} \quad M_r' = (1/n) \sum_i X_i^r$$

Estadísticos Muestrales

- Note que todo estadístico muestral $T_n = h(X)$ es una variable aleatoria porque su valor es determinado por el resultado de un experimento. A la distribución probabilística de T_n se le conoce como **distribución muestral**, la cual es completamente determinada por $h(.)$, $f(x)$, y n.

- Por ejemplo, la distribución acumulada de $T_n = (X_1 + X_2)/2$ es:

$$G(t) = Pr(T_n \leq t) = Pr(X_1 + X_2 \leq 2t) = Pr(X_2 \leq 2t - X_1)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{2t-x_1} f(x_1) f(x_2) dx_2 dx_1 = \int_{-\infty}^{\infty} f(x_1) \left[\int_{-\infty}^{2t-x_1} f(x_2) dx_2 \right] dx_1$$

$$= \int_{-\infty}^{\infty} f(x_1) F(2t - x_1) dx_1 = \int_{-\infty}^{\infty} F(2t - x) f(x) dx$$

La Media Muestral

- La media muestral satisface las siguientes propiedades con las siguientes distribuciones específicas:

  (1). Si $X \sim \textit{Bernoulli}(\rho)$, entonces $Y = n\overline{X} \sim \textit{binomial}(n, \rho)$.
  (2). Si $X \sim N(\mu, \sigma^2)$, entonces $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$.
  (3). Si $X \sim \textit{exponencial}(\lambda)$, entonces $W \sim \chi^2(k)$, donde $k = 2n$ y $W = k\lambda\overline{X}$.

- **Teorema de la Media Muestral:** En una muestra aleatoria de tamaño n y de cualquier población con $E(X) = \mu$ y $V(X) = \sigma^2$, la media muestral $\overline{X}$ tiene como esperanza $E(\overline{X}) = \mu$ y $V(\overline{X}) = \sigma^2/n$.

Momentos Muestrales

- El teorema de la media muestral se cumple también para otros estadísticos muestrales.
- De forma análoga al caso poblacional, se define el momento muestral no centrado como:

$$M_r^{'} = (1/n)\sum_i X_i^r$$

- Así, sea $Y = X^r$, al que $Y_i = X_i^r$. Luego $M_r^{'} = (1/n)\sum_i Y_i = \overline{Y}$ es una media muestral y $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$ es una muestra aleatoria de la variable aleatoria Y. Luego el teorema de la media muestral debe aplicarse a $M_r^{'}$. Así:

$$E(Y) = E(X^r) = \mu_r^{'},$$
$$V(Y) = E(Y^2) - E^2(Y) = E(X^{2r}) - E^2(X^r) = \mu_{2r}^{'} - (\mu_r^{'})^2.$$

Entonces

$$E(M_r^{'}) = \mu_r^{'} \quad \text{y} \quad V(M_r^{'}) = [\mu_{2r}^{'} - (\mu_r^{'})^2]/n$$

Momentos Muestrales

- Recuérdese también la definición de los momentos poblacionales alrededor de la media: $E(X-\mu)^r = \mu_r$. De esta forma, se construye la siguiente medida de interés:

$$M_r^* = (1/n)\sum_i (X_i - \mu)^r$$

- Sea $Y = (X-\mu)^r$, tal que $Y_i = (X_i - \mu)^r$. Entonces $M_r^* = (1/n)\sum_i Y_i = \overline{Y}$ es una media muestral en la muestra aleatoria de la variable Y. Así, aplicando el teorema de la media muestral:

$$E(M_r^*) = \mu_r \quad \text{y} \quad V(M_r^*) = (\mu_{2r} - \mu_r^2)/n$$

- En particular, si r=2:

$$E(M_2^*) = \mu_2 \quad \text{y} \quad V(M_2^*) = (\mu_4 - \mu_2^2)/n$$

Nos referiremos a M_2^* como la varianza muestral ideal dado que en la práctica esta no puede ser computada ya que μ no suele ser conocida.

Momentos Muestrales

- Ahora consideremos los momentos muestrales centrados alrededor de la media muestral:

$$M_r = (1/n)\sum_i (X_i - \overline{X})^r$$

- Sea $Y = (X - \overline{X})^r$, tal que $Y_i = (X_i - \overline{X})^r$ y $M_r = \overline{Y}$. Sin embargo, note que las $Y_i's$ no son independientes entre sí pues dados $U_1 = X_1 - \overline{X}$ y $U_2 = X_2 - \overline{X}$ tenemos que:

$$C(U_1, U_2) = C(X_1, X_2) + C(\overline{X}, \overline{X}) - C(X_1, \overline{X}) - C(X_2, \overline{X}) = -V(X)/n$$

- De este modo, el teorema de la media muestral no puede aplicarse a momentos centrados alrededor de la media muestral. No obstante, la varianza de dichos momentos puede obtenerse algebraicamente.

Varianza Muestral

- Considere el segundo momento muestral alrededor de la media muestral para X, el cual se denominará la varianza muestral y se denotará por S_x^2. De este modo:

$$S_x^2 = M_2 = (1/n)\sum_i (X_i - \overline{X})^2$$

Además

$$\sum_i (X_i - \overline{X})^2 = \sum_i [(X_i - \mu) - (\overline{X} - \mu)]^2 =$$

$$\sum_i (X_i - \mu)^2 + n(\overline{X} - \mu)^2 - 2(\overline{X} - \mu)\sum_i (X_i - \mu) = \sum_i (X_i - \mu)^2 - n(\overline{X} - \mu)^2$$

Entonces $M_2 = M_2^* - (\overline{X} - \mu)^2$, así nos queda que:

$$E(M_2) = E(M_2^*) - E(\overline{X} - \mu)^2 = \mu_2 - V(\overline{X}) = \mu_2 - \mu_2/n = \sigma^2(1 - 1/n)$$

Varianza Muestral

- De la misma forma y bajo la misma mecánica, se pueda calcular la varianza de la varianza muestral como sigue:

$$V(M_2) = (n-1)^2\{\mu_4 - [(n-3)/(n-1)]\mu_2^2]\}/n^3$$

- Note que $S_x^2 = M_2 = M_2^* - (\overline{X} - \mu)^2$, por lo tanto $S_x^2 \leq M_2^*$ en cualquier muestra.
- Asimismo, si n es grande, se cumple que:

$$E(M_2) \approx \mu_2 \quad \text{y} \quad V(M_2) \approx (\mu_4 - \mu_2^2)/n = V(M_2^*)$$

- Esto sugiere que cuando el temaño de la muestra es grande, las distribuciones de S^2 y M_2^* deberían ser similares. Estos resultados se formalizarán cuando se vea teoría asintótica.

Distribución Chi-cuadrado

- Si $Z_1, Z_2, ..., Z_k$ son variables aleatorias normales estándar independientes, y $W = \sum_{i=1}^{k} Z_i^2$ entonces la función de distribución de W es:

$$g_k(w) = (1/2)(w/2)^{(k/2)-1} exp(-w/2)/\Gamma(k/2) \quad \text{para } w > 0$$

  con $g_k(w) = 0$ en cualquier otro punto. $\Gamma(n)$ denota la función gamma.
- De este modo, decimos que $W \sim \chi^2(k)$. Asimismo, W puede reescribirse como la suma de los cuadrados de k variables independientes que se distribuyen N(0,1).
- Por otro lado, se tiene que la esperanza y varianza de W son:

$$E(W) = \sum_{i=1}^{k} E(Z_i^2) = k \quad \text{y} \quad V(W) = \sum_{i=1}^{k} V(Z_1^2) = 2k$$

- k usualmente es llamado "los grados de libertad", que en este caso es la esperanza de W.

Distribución t-student

- Para el caso de una variable aleatoria $W \sim \chi^2(k)$, se tiene que:

$$E(1/W) = 1/(k-2) \quad \text{para } k > 2 \quad ,$$

$$E(1/W^2) = 1/[(k-2)(k-4)] \quad \text{para } k > 4 \quad ,$$

- Si $Z \sim N(0,1)$ y $W \sim \chi^2(k)$, con Z y W independientes, y $U = Z/\sqrt{(W/k)}$, entonces decimos que $U \sim t(k)$ y su función de densidad es:

$$g_k(u) = \frac{\Gamma[(k+1)/2]}{\sqrt{k}\Gamma(k/2)\Gamma(1/2)}(1+u^2/k)^{-[(k+1)/2]}$$

- A esta distribución se le denomina t-student y es de un solo parámetro. En particular, k denota los grados de libertad de dicha distribución.
- Dos puntos a resaltarse de la distribución t-student son: 1) está centrada en cero y 2) es simétrica. Ambas propiedades las tiene también la distribución normal estándar.

Distribución t-student

- Dada una variable aleatoria $U \sim t(k)$, tenemos que su esperanza y su varianza son calculadas utilizando los momentos de la distribución ξ^2 y de la normal estándar. Así, tenemos que:

  $$E(U) = \sqrt{k}E(Z)E(1/\sqrt{W}) = \sqrt{k}0E(1/\sqrt{W}) = 0$$

  y

  $$V(U) = E(U^2) = E(kZ^2/W) = kE(Z^2)E(1/W) = k/(k-2)$$

  para $k > 2$
- Note que $E(U) = 0$ y que cuando $k \to \infty$, $V(U) \approx 1$. De este modo, la distribución t-student 'converge' a la distribución normal estándar. La formalización del concepto de 'convergencia' se verá a detalle en el capítulo de teoría asintótica.

Muestreo desde una Distribución Normal Estándar

- Dada una variable $X \sim N(0,1)$, supongamos que tenemos una muestra aleatoria de tamaño n dada por $X_1, X_2, ..., X_n$. Entonces tenemos que dada la media muestral $\overline{X} = \sum_i X_i/n$ y la varianza muestral $S_x^2 = (\sum_i (X_i - \overline{X})^2)/n$:

  F1*: $\sqrt{n}\overline{X} \sim_E N(0,1)$.

  F2*: $nS_x^2 \sim_E \chi^2(n-1)$.

  F3*: $\overline{X}$ y S_x^2 son independientes.

  F4*: $\sqrt{n-1} * \overline{X}/\sqrt{S_x^2} \sim_E t(n-1)$

- Los resultados se prueban utilizando que:

$$\sqrt{n} * \overline{X} = Z_1 \quad \text{y} \quad nS_x^2 = Z_2^2 + ... + Z_n^2$$

  donde $Z_1, Z_2, ..., Z_n$ son variables aleatorias independientes $N(0,1)$.

Muestras aleatorias desde una Población Normal general

- Ahora supongamos que $X \sim N(\mu, \sigma^2)$ y $\overline{X}$ y S_x^2 se definen de la misma manera que en el caso anterior. Luego tenemos que:

  F1*: $\overline{X} \sim_E N(\mu, \sigma^2/n)$.

  F2*: $W = nS^2/\sigma^2 \sim_E \chi^2(n-1)$.

  F3*: $\overline{X}$ y S_x^2 son independientes.

  F4*: $U = \sqrt{n-1} * (\overline{X} - \mu)/S \sim_E t(n-1)$

- Los resultados se prueban escribiendo X_i como función de una distribución normal estándar. Específicamente, $X = \mu + \sigma Z$ con $Z \sim N(0,1)$.

Distribuciones Muestrales: El caso bivariado

- Ahora extendemos el análisis de los momentos muestrales de una población univariada a una población bivariada. Consideremos una población bivariada en la cual la función de distribución conjunta de dos variables aleatorias (X,Y) es f(x,y). Luego los primeros y segundos momentos incluyen:

$$E(X) = \mu_x \qquad E(Y) = \mu_y$$

$$V(X) = \sigma_x^2 \quad V(Y) = \sigma_y^2 \quad C(X,Y) = \sigma_{xy}$$

- Una muestra aleatoria de esta población consiste en la colección de vectores aleatorios provenientes de dicha distribución. Así, para una muestra aleatoria de tamaño n, (X_i, Y_i) para i=1,...,n son vectores idéntica e independientemente distribuidos.
- Note que la independencia es entre vectores y no necesariamente entre X e Y. Es decir, la independencia se da a lo largo de las observaciones y no a lo largo de cada vector.

Momentos Muestrales

- Los estadísticos muestrales son funciones de la muestra aleatoria. En el caso bivariado se resaltarán los momentos muestrales que se obtuvieron en el caso univariado; es decir, $\overline{X}, \overline{Y}, S_x^2, S_y^2$ y también los estadísticos conjuntos como la covarianza muestral que involucra a ambos componentes de los vectores aleatorios.
- La covarianza muestral se define como:

$$S_{xy} = (1/n) \sum_i (X_i - \overline{X})(Y_i - \overline{Y})$$

- Note que la covarianza entre las dos medias muestrales son:

$$C(\overline{X}, \overline{Y}) = (1/n^2) \sum_i \sum_h C(X_i, Y_h) = (1/n^2) \sum_i C(X_i, Y_i)$$

$$= (1/n^2) n\sigma_{xy} = \sigma_{xy}/n$$

Covarianza Muestral

- Note que S_{xy} puede ser escrito de la siguiente manera:

$$S_{xy} = (1/n)\sum_i (X_i - \mu_x)(Y_i - \mu_y) - (\overline{X} - \mu_x)(\overline{Y} - \mu_y)$$

  Entonces

$$E(S_{xy}) = \sigma_{xy} - C(\overline{X}, \overline{Y}) = \sigma_{xy} - C(X, Y)/n = (1 - 1/n)\sigma_{xy}$$

  Asimismo, un cálculo directo nos lleva a lo siguiente:

$$V(S_{xy}) = (n-1)^2(\mu_{22} - \mu_{11}^2)/n^3 + 2(n-1)(\mu_{20}\mu_{02})/n^3$$

- Se pueden definir otros estadísticos como el ratio de medias muestrales dado por $T = \overline{X}/\overline{Y}$.
- Un estadístico que será de interés es el análogo muestral de la pendiente del mejor predictor lineal de Y sobre X dado por $\beta = \sigma_{xy}/\sigma_x^2$. Así, el análogo muestral de β es:

$$B = S_{xy}/S_x^2$$