

Muestreo Estadístico: π -Estimador^{*}

García Prado, Sergio
sergio@garciparedes.me

26 de septiembre de 2017

1. Demostración de la Inssegadez

Para una población población finita U formada por los individuos $\{1, \dots, i, \dots, N\}$ cuya variable de interés se denota por y , siendo y_k el valor que toma dicha variable en el individuo k . Se define el total poblacional como sigue:

$$t = \sum_{k \in U} y_k \quad (1)$$

Sea s una muestra extraída de la población U (por tanto $s \subset U$), se define la variable aleatoria I_k como se indica a continuación:

$$I_k = \begin{cases} 1 & \text{si } k \in s \\ 0 & \text{si } k \notin s \end{cases} \quad (2)$$

De esta manera, se puede utilizar el valor π_k para modelizar la probabilidad de que I_k tome el valor 1, es decir, de que el elemento k -ésimo esté en la muestra s . Por tanto, se puede afirmar que $I_k \sim B(\pi_k)$, una distribución de *Bernoulli* de parámetro π_k . A partir de dicha distribución se tiene que $E[I_k] = \pi_k$. Además, $\pi_k > 0 \forall k \in U$ para que se cumpla la restricción de muestreo probabilístico.

Se define el π -estimador como:

$$\hat{t}_\pi = \sum_{k \in s} \frac{y_k}{\pi_k} = \sum_{k \in U} I_k \frac{y_k}{\pi_k} \quad (3)$$

La demostración acerca de la inssegadez de este estimador deriva de la esperanza de la variable I_k , que toma el valor π_k , tal y como se indicaba anteriormente, lo cual produce que se anule dicho valor π_k con el π_k del divisor del estimador. Además, la restricción de muestreo probabilístico $\pi_k > 0 \forall k \in U$ hace que el valor \hat{t}_π exista siempre (no hay división entre 0). Por estas razones se puede decir que el estimador es inssegado. De manera matemática esto se describe a continuación:

$$E[t] = \sum_{k \in U} y_k \quad (4)$$

$$E[\hat{t}_\pi] = \sum_{k \in U} E[I_k] \frac{y_k}{\pi_k} = \sum_{k \in U} \pi_k \frac{y_k}{\pi_k} = \sum_{k \in U} y_k \quad (5)$$

$$sesgo(t, \hat{t}_\pi) = |E[t] - E[\hat{t}_\pi]| = \left| \sum_{k \in U} y_k - \sum_{k \in U} y_k \right| = 0 \quad (6)$$

^{*}URL: <https://github.com/garciparedes/statistical-sampling-pi-estimator>

Nótese por tanto, que el funcionamiento del π -estimador se basa en la “expansión” del conjunto de individuos seleccionados en la muestra s , de tal manera que se siga cumpliendo de la forma más precisa posible la distribución de probabilidades de la población U .

2. Demostración del Intervalo de Confianza

En este caso se pretende probar el intervalo de confianza de nivel $1 - \alpha$ para la distribución de probabilidades referida a la variable aleatoria t , es decir, a la suma del total poblacional para la variable Y . Es decir, se pretende probar la igualdad descrita en la ecuación (7). Dicha demostración se llevará a cabo siguiendo la notación de valores de probabilidad.

$$\left[\hat{t}_\pi \pm z_{1-\alpha/2} \sqrt{\widehat{Var}[\hat{t}_\pi]} \right] \equiv \Pr \left(t \in \left[\hat{t}_\pi \pm z_{1-\alpha/2} \sqrt{Var[\hat{t}_\pi]} \right] \right) = 1 - \alpha \quad (7)$$

Un paso previo para la demostración del intervalo de confianza es la demostración de la ecuación (8). Por el *Teorema Central del Límite* se puede asumir que la distribución t sigue una distribución normal con media \hat{t} y varianza $Var[\hat{t}_\pi]$. Por tanto, este puede ser tipificado a una distribución normal estándar. Esta propiedad se indica en la ecuación (8)

$$\frac{t - \hat{t}_\pi}{\sqrt{Var[\hat{t}_\pi]}} \sim N(0, 1) \quad (8)$$

Por tanto, para demostrar que el intervalo de confianza sigue un nivel $1 - \alpha$ basta con aplicar distintas propiedades algebraicas sobre este, hasta llegar a la distribución normal estándar tal y como se indicó en la ecuación (8), que debido a su simetría permite sustituir el intervalo multiplicando por 2 el valor de α . Por último, debido a las propiedades de la distribución normal estándar, queda demostrado que en hasta el percentil $z_{1-\alpha}$ se localiza la probabilidad $1 - \alpha$.

$$\Pr \left(t \in \left[\hat{t}_\pi \pm z_{1-\alpha/2} \sqrt{Var[\hat{t}_\pi]} \right] \right) = \quad (9)$$

$$= \Pr \left(t - \hat{t}_\pi \in \left[\pm z_{1-\alpha/2} \sqrt{Var[\hat{t}_\pi]} \right] \right) \quad (10)$$

$$= \Pr \left(\frac{t - \hat{t}_\pi}{\sqrt{Var[\hat{t}_\pi]}} \in [\pm z_{1-\alpha/2}] \right) \quad (11)$$

$$= \Pr (N(0, 1) \in [\pm z_{1-\alpha/2}]) \quad (12)$$

$$= \Pr (N(0, 1) \leq z_{1-\alpha}) \quad (13)$$

$$= 1 - \alpha \quad (14)$$

Referencias

- [SSW03] Carl-Erik Särndal, Bengt Swensson, and Jan Wretman. *Model assisted survey sampling*. Springer Science & Business Media, 2003.
- [TG18] Jesús Alberto Tapia García. Muestreo Estadístico 1, 2017/18.