# VI ESTIMACIÓN PUNTUAL DE PARÁMETROS POBLACIONALES

Desde el punto de vista de la Estadística Clásica, se considera a los parámetros de una población como cantidades fijas, las cuáles es posible *estimar* mediante estadísticos que son función de los datos obtenidos en una o más muestras y que no contienen cantidades desconocidas. A estos estadísticos se les conoce como estimadores del parámetro. Las técnicas mas utilizadas para encontrar estimadores para algún parámetro son el método de máxima verosimilitud y el método de los momentos. Frecuentemente se puede proponer más de un estimador para el mismo parámetro, así que es necesario determinar cual de ellos es el **mejor estimador** para ese parámetro; se han desarrollado criterios específicos para seleccionar al mejor estimador, si existe. En las siguientes secciones se estudian dichos criterios.

**Un Estimador Puntual** es cualquier estadístico que se utiliza para inferir un parámetro desconocido de la población y que proporciona un único valor como estimación de dicho parámetro. En general representaremos con $\theta$ a un parámetro cualquiera en el que estamos interesados, a su estimador con $\hat{\Theta}$ y al estimado o estimación con $\hat{\theta}$ (la notación puede variar en algunos casos).

Los criterios más usuales para evaluar un estimador son:

- Sesgo
- Varianza
- Eficiencia
- Consistencia
- Suficiencia

Error cuadrático medio

A continuación, se dan algunas definiciones sobre estos criterios, que permiten evaluar la bondad de un estimador. Conviene aquí recordar que un estimador es una variable aleatoria, cuya distribución de probabilidad se conoce como distribución de muestreo; y que para una variable aleatoria con distribución conocida es posible obtener su valor esperado y su varianza.

## ***Definiciones***

*Media de un estimador*: es el valor esperado del estimador.

$$\mu_{\hat{\Theta}} = E\{\hat{\Theta}\}$$

*Varianza de un estimador*: es el segundo momento con respecto a la media del estimador.

$$\sigma^2_{\hat{\Theta}} = Var\{\hat{\Theta}\}, \quad \text{es decir:} \qquad \sigma^2_{\hat{\Theta}} = E\{\hat{\Theta} - \mu_{\hat{\Theta}}\}^2$$

recordemos que a su vez, la varianza del estimador se puede expresar como:

$$\sigma^2_{\hat{\Theta}} = E\{\hat{\Theta}^2\} - [E\{\hat{\Theta}\}]^2$$

*Error estándar de un estimador*: El error estándar es una medida para indicar la precisión de un estimador y es simplemente la desviación estándar del estimador.

$$\mathrm{E}S(\hat{\Theta}) = \sqrt{Var(\hat{\Theta})}$$

Cuando el error estándar de un estimador contiene parámetros desconocidos, pero que es posible estimar, al sustituir esas estimaciones en la expresión para el error estándar, se obtiene un *error estándar estimado*.

*Sesgo: e*s la diferencia entre el valor esperado del estimador y el parámetro que pretende estimar.

$$Sesgo[\hat{\Theta}] = \theta - \mathrm{E}\{\hat{\Theta}\}$$

*Estimador insesgado*: Es aquel cuyo valor esperado es igual al parámetro que pretende estimar, y por lo tanto su sesgo es igual a cero.

Un estimador $\hat{\Theta}$ del parámetro $\theta$ es insesgado si $E\{\hat{\Theta}\} = \theta$, lo que también equivale a que su sesgo sea cero.

*Error Cuadrático medio*: Es una medida de la variabilidad del estimador, definida como:

$$\mathrm{ECM}\{\hat{\Theta}\} = \mathrm{E}\{(\hat{\Theta} - \theta)^2\}$$

Notar que aunque es parecido, difiere de $Var(\hat{\Theta})$

Si desarrollamos el cuadrado dentro del valor esperado tenemos:
$$\mathrm{ECM} = \mathrm{E}\{(\hat{\Theta}^2 - 2\theta\hat{\Theta} + \theta^2)\}$$

Distribuyendo el operador de valor esperado:
$$\mathrm{ECM} = \mathrm{E}\{\hat{\Theta}^2\} - 2\theta\mathrm{E}\{\hat{\Theta}\} + \theta^2$$
(ya que para efectos del valor esperado, $\theta^2$ es una constante).

Pero como
$Var(\hat{\Theta}) = \mathrm{E}\{\hat{\Theta}^2\} - [\mathrm{E}\{\hat{\Theta}\}]^2$ entonces $\mathrm{E}\{\hat{\Theta}^2\} = Var\{\hat{\Theta}\} + [\mathrm{E}\{\hat{\Theta}\}]^2$

y sustituyendo en la última expresión que tenemos que el ECM queda:
$$\mathrm{ECM} = Var\{\hat{\Theta}\} + [\mathrm{E}\{\hat{\Theta}\}]^2 - 2\theta\mathrm{E}\{\hat{\Theta}\} + \theta^2$$

donde si agrupamos los últimos tres términos, notamos que corresponden al desarrollo de un binomio, es decir:
$\left[E\{\hat{\Theta}\}\right]^2 - 2\theta E\{\hat{\Theta}\} + \theta^2 = \left(\theta - E\{\hat{\Theta}\}\right)^2$, que es justamente el cuadrado del sesgo de $\hat{\Theta}$; así que el error cuadrático medio se puede expresar como:

$$ECM = Var\{\hat{\Theta}\} + \left(\theta - E\{\hat{\Theta}\}\right)^2 \quad \text{ó} \quad ECM = Var\{\hat{\Theta}\} + \left(Sesgo\{\hat{\Theta}\}\right)^2$$

Por lo tanto el error cuadrático medio es igual a la varianza del estimador más el sesgo al cuadrado.

Se deduce que si un estimador es insesgado, entonces su error cuadrático medio es igual a su varianza.

Ahora bien, tratándose de comparar dos estimadores $\hat{\Theta}_1$ y $\hat{\Theta}_2$ para el mismo parámetro $\theta$, se preferirá aquel con menor error cuadrático medio, ya que es una medida que combina tanto el sesgo como la varianza de ambos estimadores. Si definimos la ***eficiencia relativa*** de $\hat{\Theta}_1$ con respecto a y $\hat{\Theta}_2$ como el cociente de los errores cuadráticos medios respectivos:

$$\eta_{\Theta_2/\Theta_1} = \frac{ECM\{\hat{\Theta}_2\}}{ECM\{\hat{\Theta}_1\}}$$

entonces si $\eta_{\Theta_2/\Theta_1}$ es menor a uno implica que $ECM\{\hat{\Theta}_2\} < ECM\{\hat{\Theta}_1\}$ y se dice que $\hat{\Theta}_2$ es más eficiente que $\hat{\Theta}_1$; si es mayor que uno, entonces $\hat{\Theta}_1$ es más eficiente; y si es igual a uno, entonces ambos estimadores son igualmente eficientes. Naturalmente es preferible elegir al estimador más eficiente, ya que es el que presenta menor variabilidad (menor ECM).

*Estimador óptimo*: es aquel que tiene el menor error cuadrático medio de entre todos los estimadores que existan para estimar al parámetro $\theta$, para cualquier valor de $\theta$.

*Estimador insesgado de varianza mínima uniforme*: Si existe un estimador insesgado $\hat{\Theta}$ para el parámetro $\theta$, tal que la varianza de $\hat{\Theta}$ sea la menor de entre todos los estimadores insesgados de $\theta$, para cualquier valor de $\theta$, entonces $\hat{\Theta}$ se denomina el estimador de varianza mínima uniforme.

Una forma de determinar la varianza del estimador insesgado de varianza mínima la proporciona un resultado que asegura que la varianza de cualquier estimador insesgado siempre será mayor o a lo sumo igual que cierto límite conocido como *cota inferior de Cramér-Rao*, Es decir, si $\hat{\Theta}$ es un estimador insesgado del parámetro $\theta$, y además se cumple que:

i) $\frac{\partial}{\partial\theta} \ln f(x;\theta)$ existe para toda $x$ y para toda $\theta$.

ii) $0 < E\left\{\left[\frac{\partial}{\partial\theta}\ln f(x;\theta)\right]^2\right\} < \infty$ para toda θ.

donde *f(x; θ)* es la distribución de la población, que es función del parámetro θ, entonces:

$$Var(\hat{\Theta}) \geq \frac{1}{nE\left[\left(\frac{\partial}{\partial\theta}\ln f(x;\theta)\right)^2\right]}$$

El lado derecho de la desigualdad es la cota inferior de Cramér-Rao, y puede asegurarse que, bajo los supuestos iniciales i) e ii), no existe ningún estimador insesgado para θ con varianza menor a ese límite.

Si existe el estimador $\hat{\Theta}'$ tal que su varianza sea exactamente igual a la cota inferior de Cramér-Rao, entonces ese estimador se denomina el ***estimador eficiente*** de θ; es decir:

Si existe $\hat{\Theta}'$ insesgado tal que $Var(\hat{\Theta}') = \frac{1}{nE\left[\left(\frac{\partial}{\partial\theta}\ln f(x;\theta)\right)^2\right]}$;

entonces $\hat{\Theta}$ es el estimador eficiente de θ.

La utilidad inmediata de la cota inferior de Cramér-Rao consiste en que al calcularla, puede conocerse al menos la varianza del estimador insesgado de varianza mínima, y así estar en posibilidades de encontrar ese estimador; en lugar de buscar cualquier estimador, se buscará únicamente aquel cuya varianza satisface la cota citada, con la seguridad de que no habrá otro estimador insesgado mejor. Por otro lado, si ya se cuenta con un estimador insesgado, al comparar su varianza con la cota inferior de Cramér-Rao, puede saberse si ése es el mejor estimador insesgado o es posible que exista otro mejor.

*Estimador consistente*: Intuitivamente, un estimador $\hat{\Theta}$ del parámetro será consistente para estimar al parámetro θ si a medida que aumenta el tamaño de la muestra, la diferencia entre el valor proporcionado por el estimador y el verdadero valor del parámetro tiende a cero. Mas formalmente un estimador $\hat{\Theta}$ es consistente para estimar al parámetro θ si:

$$\lim_{n\to\infty} P\left(\left|\hat{\Theta} - \theta\right| < \varepsilon\right) = 1$$

donde ε es cualquier número real mayor que cero.

Una alternativa para determinar si un estimador es consistente es utilizar el hecho de que aquellos estimadores cuyo error cuadrático medio tiende a cero cuando el tamaño de la muestra tiende a infinito, son consistentes; es decir, si $\lim_{n\to\infty} ECM(\hat{\Theta}) = 0$, entonces el estimador $\hat{\Theta}$ es consistente para θ.

<u>*Estadísticas suficientes*</u>: Se dice que una estadística es suficiente para estimar a algún parámetro si utiliza toda la información que proporciona la muestra y la resume en un solo valor. Una definición de estadística suficiente es:

> Sea $\underline{X} = \{X_1, X_2, ...X_n\}$ es una muestra aleatoria simple extraída de una población con distribución $f(x;\theta)$, y sea $T=t(X_1, X_2, ...X_n)$ una estadística , T es *suficiente* para estimar θ si la distribución condicional de $X_1, X_2, ...X_n$ dado T no depende de θ.

Un teorema útil para determinar si una estadística $T=t(X_1, X_2, ...X_n)$ es suficiente para estimar θ es el siguiente:

> <u>*Teorema de factorización de Neyman-Pearson*</u>
>
> Si $\underline{X} = \{X_1, X_2, ...X_n\}$ es una muestra aleatoria simple extraída de una población con distribución $f(x;\theta)$ y la función de verosimilitud de la muestra es $L(\underline{X};\ \theta)$[1]; la estadística $T=t(X_1, X_2, ...X_n)$ es suficiente para $\theta$ si y solo si la función de verosimilitud se puede factorizar en dos funciones tales que, una de ellas sea función de la muestra y la otra no contenga al parámetro en cuestión.
>
> $$L(\underline{X},\ \theta)=h(t;\theta)g(\underline{X})$$
>
> donde $h(t;\theta)$ es una función no negativa que depende únicamente de t y del parámetro $\theta$; y $g(\underline{X})$ es una función no negativa y no contiene al parámetro $\theta$.

Un resultado importante es que si T es una estadística suficiente para θ y definimos otra estadística W(t) que es función de la estadística suficiente T, entonces W(T) también es suficiente para estimar θ.

Ejemplo: demostrar que si X es una variable aleatoria con función de densidad de probabilidad f(x) normal, con media μ, entonces $T = \sum_{i=1}^{n} X_i$ es suficiente para μ, y por lo tanto $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} Xi$, también es suficiente para μ ya que como se ve es función de la estadística suficiente T.

---

[1] Recordar que la función de verosimilitud de la muestra $L(\underline{X};\theta)$ se define como la función de probabilidad conjunta de la observaciones de la muestra $X_1, X_2, ...,X_n$,. Además si las $X_i$ son independientes e idénticamente distribuidas con función de probabilidad $f(x;\theta)$, como en el caso del muestreo aleatorio simple con reemplazo, entonces $L(\underline{X};\theta)=f(x_1;\theta)f(x_2;\theta)..f(x_n;\theta)$.

**Ejercicio 1**: Sea una muestra aleatoria $\underline{X} = \{x_1, x_2, ...x_n\}$ extraída con reemplazo de una población con distribución *f(x)*, con media $\mu$ y varianza $\sigma^2$. Demostrar que:

a) la media muestral $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ es un estimador insesgado de la media poblacional $\mu$

b) $S_n{}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{\mathrm{X}})^2$ es un estimador sesgado de la varianza poblacional $\sigma^2$ y encontrar el sesgo.

c) La varianza muestral definida como $S_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{\mathrm{X}})^2$ es un estimador insesgado de la varianza poblacional $\sigma^2$.

***Resolución:*** lo primero que se debe considerar, es que como la muestra es aleatoria con reemplazo, entonces la distribución de probabilidad de cada observación $x_i$ es idéntica a la distribución de la población, $f(x_i)=f(x)$, y por lo tanto, el valor esperado y la varianza de las $x_i$ son también iguales respectivamente a la media y la varianza de la población:

$$E[x_i] = \mu \qquad \text{y} \qquad Var[x_i] = \sigma^2$$

a) para saber si $\overline{X}$ es insesgado para $\mu$, se debe determinar su valor esperado:

$E[\overline{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right]$ pero como las variables aleatorias $x_i$ son independientes, se puede distribuir el valor esperado en la sumatoria: $E[\overline{X}] = \frac{1}{n}\sum_{i=1}^{n} E[x_i]$, donde $E[x_i] = \mu$ se tiene: $E[\overline{X}] = \frac{1}{n}\sum_{i=1}^{n}\mu$ , con lo que $E[\overline{X}] = \frac{1}{n} n\mu$ y por último tenemos que $E[\overline{X}] = \mu$.

Como el valor esperado de $\overline{X}$ es igual al parámetro $\mu$, entonces la media muestral $\overline{X}$ es un estimador insesgado de la media poblacional $\mu$.

b) Para determinar si $S_n{}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{\mathrm{X}})^2$ es un estimador insesgado para la varianza poblacional $\sigma^2$ también debemos encontrar su valor esperado.

$$\mathrm{E}\{S_n^2\} = \mathrm{E}\left\{\frac{1}{n}\sum_{I=1}^{n}(x_i - \overline{X})^2\right\}$$

El factor 1/n puede salir como constante del valor esperado:

$$E\left\{S_n^2\right\}=\frac{1}{n}\mathrm{E}\left\{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2\right\}$$

Dentro del binomio cuadrado, podemos sumar y restar $\mu$ y agrupar algunos términos:

$$E\left\{S_n^2\right\}=\frac{1}{n}E\left\{\sum_{i=1}^{n}\left[\left(x_i-\mu\right)-\left(\overline{X}-\mu\right)\right]^2\right\}$$

En seguida desarrollamos el binomio cuadrado:

$$E\left\{S_n^2\right\}=\frac{1}{n}\mathrm{E}\left\{\sum_{i=1}^{n}\left[\left(x_i-\mu\right)^2-2\left(x_i-\mu\right)\left(\overline{X}-\mu\right)+\left(\overline{X}-\mu\right)^2\right]\right\}$$

Ahora distribuimos la sumatoria a cada término, reconociendo que para efectos de la sumatoria los términos $(\overline{X}-\mu)$ y $(\overline{X}-\mu)^2$ son constantes:

$$E\left\{S_n^2\right\}=\frac{1}{n}E\left[\sum_{i=1}^{n}\left(x_i-\mu\right)^2-2\left(\overline{X}-\mu\right)\sum_{i=1}^{n}\left(x_i-\mu\right)+n\left(\overline{X}-\mu\right)^2\right]$$

Por otro lado, observemos que:

$$\sum_{i=1}^{n}\left(x_i-\mu\right)=\sum_{i=1}^{n}\left(x_i\right)-n\mu=n\frac{\sum_{i=1}^{n}\left(x_i\right)}{n}-n\mu=n\overline{X}-n\mu=n\left(\overline{X}-\mu\right)$$

por lo que el valor esperado de $S_n^2$ se puede expresar como:

$$E\left\{S_n^2\right\}=\frac{1}{n}E\left[\sum_{i=1}^{n}\left(x_i-\mu\right)^2-2\left(\overline{X}-\mu\right)n\left(\overline{X}-\mu\right)+n\left(\overline{X}-\mu\right)^2\right]$$

$$E\left\{S_n^2\right\}=\frac{1}{n}E\left[\sum_{i=1}^{n}\left(x_i-\mu\right)^2-2n\left(\overline{X}-\mu\right)^2+n\left(\overline{X}-\mu\right)^2\right]$$

$$E\left\{S_n^2\right\}=\frac{1}{n}E\left[\sum_{i=1}^{n}\left(x_i-\mu\right)^2-n\left(\overline{X}-\mu\right)^2\right]$$

$$E\left\{S_n^2\right\}=\frac{1}{n}\left\{\sum_{i=1}^{n}E\left[\left(x_i-\mu\right)^2\right]-nE\left[\left(\overline{X}-\mu\right)^2\right]\right\}$$

Aquí, debemos notar que $E\left[\left(x_i-\mu\right)^2\right]$ es la varianza de cada observación $x_i$, que es igual a $\sigma^2$; y también $E\left[\left(\overline{X}-\mu\right)^2\right]$ es la varianza de la variable aleatoria $\overline{X}$, que es igual a $\frac{\sigma^2}{n}$, con lo que el valor esperado de $E\left\{S_n^2\right\}$ queda:

$$E\left\{S_n^2\right\} = \frac{1}{n}\left\{\sum_{i=1}^{n}\sigma^2 - n\frac{\sigma^2}{n}\right\} , \qquad E\left\{S_n^2\right\} = \frac{1}{n}\left\{n\sigma^2 - \sigma^2\right\}$$

Y finalmente:

$$E\left\{S_n^2\right\} = \frac{(n-1)\sigma^2}{n}$$

Es claro que $E\left\{S_n^2\right\} \neq \sigma^2$, con lo que se concluye que $S_n{}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{\mathrm{X}}\right)^2$ no es un estimador insesgado de $\sigma^2$.

Para encontrar el sesgo basta con encontrar la diferencia entre el parámetro y el estimador propuesto:

$$Sego\left\{S_n^2\right\} = \sigma^2 - E\left\{S_n^2\right\} , \qquad Sego\left\{S_n^2\right\} = \sigma^2 - \frac{(n-1)\sigma^2}{n} , \qquad Sego\left\{S_n^2\right\} = \frac{\sigma^2}{n}$$

c) Para demostrar que la varianza muestral definida como $S_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{\mathrm{X}}\right)^2$ es un estimador insesgado, utilizamos el resultado anterior $E\left\{S_n^2\right\} = \frac{(n-1)\sigma^2}{n}$;

notando que $S_{n-1}^2 = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2$, es decir que $S_{n-1}^2 = \frac{n}{n-1}S_n^2$

Entonces $E\left[S_{n-1}^2\right] = E\left[\frac{n}{n-1}S_n^2\right]$; $\quad E\left[S_{n-1}^2\right] = \frac{n}{n-1}E\left[S_n^2\right]$; $\quad E\left[S_{n-1}^2\right] = \frac{n}{(n-1)}\frac{(n-1)\sigma^2}{n}$

Finalmente: $E\left[S_{n-1}^2\right] = \sigma^2$ con lo que queda demostrado que $S_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{\mathrm{X}}\right)^2$ es un estimador insesgado de la varianza poblacional.

Como conclusión puede verse que **no es adecuado** usar $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2$ como estimador de la varianza poblacional, ya que resulta ser un estimador sesgado; y que el estimador insesgado es $S_{n-1}^2 = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2$ que es el que debe usarse para estimar la varianza de la población cuando se disponga de una muestra aleatoria simple de tamaño n. En adelante para hacer referencia al estimador insesgado se utilizará únicamente la notación $S^2$ conocido como varianza muestral, y se entenderá por: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2$.

Bibliografía:

1. Canavos G., Probabilidad y Estadística Aplicaciones y Métodos,
McGraw Hill, México 1988

2. Hines W., Montgomety D., G., Probabilidad y Estadística para Ingeniería y Administración,
3ra. Ed., CECSA, México 1988

2. Hines W., Montgomety D., G., Probabilidad y Estadística para Ingeniería y Administración,
3ra. Ed., CECSA, México 1988
Mood A., Graybill F, Boes D., Introduction to the Theory of Statistics
3rd. Ed., Mc Graw Hill.