MÉTODO ESTADÍSTICO PARA ESTIMAR PARÁMETROS

Lic. César Valcárcel Rodríguez

El presente documento fue expuesto con ocasión del Seminario de Investigación realizado en la UNIFÉ, en el mes de marzo del 2001.
Se trata de un método práctico, fácil de entender y de aplicación en las diferentes áreas de la investigación estadística.

Un aspecto del trabajo estadístico, es la inferencia estadística, que es el proceso de hacer uso de los resultados de una investigación en base a muestras aleatorias, con la finalidad de obtener conclusiones sobre los hechos o fenómenos que ocurren en una población. La inferencia supone la estimación y prueba de hipótesis que son procedimientos estadísticos para la toma de decisiones.

En este documento nos referiremos específicamente a la estimación de parámetros poblacionales, para lo cual presentamos algunas definiciones inherentes al tema, a fin de comprender mejor este procedimiento estadístico.

La estimación es la aproximación de los valores de los parámetros, que se efectúa con la información captada en un conjunto de observaciones muestrales y de acuerdo a ciertos procedimientos establecidos por indicadores que son llamados estimadores.

PARÁMETRO.

Es una medida de las características de la población.
Ejemplo: la media poblacional (); la proporción poblacional (P); la varianza poblacional (), entre otros.

ESTIMADOR.

Es una función de las observaciones muestrales que no dependen de parámetro alguno.

Un estimador define un procedimiento para resumir la información captada en una muestra con la finalidad de obtener una estimación o aproximación del valor de un parámetro.

Los estimadores son variables aleatorias cuyos valores dependen de las observaciones captadas en una muestra aleatoria.
Las estimaciones son valores particulares de las variables aleatorias llamadas estimadores.

CONSIDERACIONES DE LOS ESTIMADORES

1) ESTIMADOR INSESGADO

Un estimador $\hat{\theta}$ se dice que es insesgado del parámetro θ, si $E[\hat{\theta}] = \theta$
Ejemplo:

a) Si $\overline{X} = \frac{\sum_{1}^{n} X_i}{n}$, entonces $E[\overline{X}] = \mu$.

b) Si $S^2 = \frac{\sum_{1}^{n}(X_i - \overline{X})^2}{n-1}$, entonces $E[S^2] = \sigma^2$.

Por consiguiente, X y S^2 son estimadores insesgados de los parámetros μ y σ^2

ESTIMADOR CONSISTENTE

Un estimador $\hat{\theta}$ del parámetro se dice que es consistente, si al tomar un tamaño de muestra «n» cada vez más grande, el valor $\hat{\theta}$ se aproxima cada vez más al valor del parámetro ; es decir,

si $\lim_{n \to \infty} \hat{\theta} = \theta$

Ejemplo:

Si $S_1^2 = \frac{\sum_1^n (X_i - \bar{X})^2}{n}$, entonces $E[S_1^2] = \frac{n-1}{n}\sigma^2$,

es decir S_1^2 es un estimador sesgado de σ^2, pero el sesgo $\frac{n-1}{n}$ decrece a medida que "n" es grande.

2) ESTIMADOR EFICIENTE

Se dice que es un estimador eficiente del parámetro , si es el de menor varianza entre todos aquellos estimadores insesgados que se pueden establecer.

Lo que quiere decir, si $Var(\hat{\theta}) = Var(\hat{\theta}_1)$, donde $\hat{\theta}_1$ es otro estimador insesgado de θ.
Por consiguiente el estimador $\hat{\theta}$ es más eficiente que el estimador $\hat{\theta}_1$, de esta forma $\hat{\theta}$ es un estimador eficiente.

La eficiencia de un estimador se evalúa por su varianza.

Ejemplo.

Dada una muestra $X_1; X_2; \ldots ; X_n$, se define

$\hat{\theta}_1 = (X_1 + X_2 + X_3 + X_4 + X_5)/5$ y $\hat{\theta}_2 = \bar{X}$

$Var(\hat{\theta}_1) = \frac{\sigma^2}{5}$ y $Var(\hat{\theta}_2) = \frac{\sigma^2}{n}$; de donde se desprende que $\hat{\theta}_2$ es más eficiente que $\hat{\theta}_1$ ya que la $Var(\hat{\theta}_2) \leq Var(\hat{\theta}_1)$

3) ESTIMADOR SUFICIENTE

Se dice que $\hat{\theta}$ es un estimador suficiente, si es una función de todas las observaciones de una muestra, la cual se captó para estimar el parámetro $\boldsymbol{\theta}$.

Una definición más rigurosa indica que si $\hat{\theta}$ y $\hat{\theta}_1$ son dos estimadores del parámetro $\boldsymbol{\theta}$, donde $\hat{\theta}_1$ no es función de $\hat{\theta}$ y $P[\hat{\theta}_1/\hat{\theta}]$ no depende del parámetro $\boldsymbol{\theta}$, entonces se dice que $\hat{\theta}$ es un estimador suficiente para $\boldsymbol{\theta}$.

Ejemplo.

Dada una muestra $X_1, X_2, \ldots, X_n$, se define $\hat{\theta}_1 = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$ y $\hat{\theta}_2 = \frac{X_1 + X_2 + \ldots + X_n}{n}$, como $\hat{\theta}_1$ solo contiene a $X_1 + X_2 + X_3 + X_4 + X_5$; luego no es suficiente; pero $\hat{\theta}_2$ es una función de todas las observaciones muestrales y por tanto es un estimador suficiente.

Nota: Todo estimador suficiente e insesgado, es consistente.

Ejemplo:

La $\bar{X}$ y la S^2 son estimadores suficientes, también son consistentes.

CLASES DE ESTIMACIÓN

Para realizar las estimaciones se usan dos tipos de estimación: la estimación puntual y la estimación por intervalos, que a continuación se desarrollan:

A) ESTIMACIÓN PUNTUAL

Es la estimación del valor de un parámetro por medio de un único valor, el cual es obtenido mediante el cálculo o evaluación de un estimador para una muestra específica.

Ejemplo.-

Si para una muestra se obtiene $\bar{X}=26$ y $S^2=25$, entonces se puede decir que la $\bar{X}$ es una estimación puntual del parámetro $\boldsymbol{\mu}$;

S^2 es una estimación puntual del parámetro σ^2.

B) ESTIMACIÓN POR INTERVALOS

Es la estimación del valor de un parámetro mediante un conjunto de valores contenidos en un intervalo, el cual se obtiene a partir de una información muestral de modo tal, que brinde una confianza establecida de contener al valor del parámetro.

NOTA.

a) La estimación puntual tiene el inconveniente de no poder precisar el error de estimación del parámetro.
b) La estimación por intervalos tiene la posibilidad de poder estimar el error de estimación del parámetro.

A continuación se presenta los procedimientos para la estimación de la media y la proporción poblacional, en otra ocasión se hará con otras stimaciones.

PROCEDIMIENTO PARA LA ESTIMACIÓN POR INTERVALOS DE LA MEDIA POBLACIONAL (M)

A) Cuando la población es infinita (se conoce σ^2)

$$\bar{x} - Z\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le +Z\left(\frac{\sigma}{\sqrt{n}}\right)$$

B) Cuando la población es finita (se conoce σ^2)

$$\bar{x} - Z\left(\frac{\sigma}{\sqrt{n}}\right)\left(\frac{N-n}{N-1}\right) \le \mu \le \bar{x} + Z\left(\frac{\sigma}{\sqrt{n}}\right)\left(\frac{N-n}{N-1}\right)$$

PROCEDIMIENTO PARA LA ESTIMACIÓN POR INTERVALOS DE LA PROPORCIÓN POBLACIONAL (P)

A) Cuando la población es infinita (n > 30)

$$p - Z\sqrt{\frac{pq}{n}} \le P \le p + Z\sqrt{\frac{pq}{n}}$$

B) Cuando la población es finita (n > 30)

$$p - Z\sqrt{\frac{N-n}{N-1}}\sqrt{\frac{pq}{n}} \le P \le p + Z\sqrt{\frac{N-n}{N-1}}\sqrt{\frac{pq}{n}}$$

EJEMPLOS DE ESTIMACIÓN DE LA MEDIA Y LA PROPORCIÓN POBLACIONAL

A) ESTIMACIÓN DE LA MEDIA POBLACIONAL

Un analista de investigación de mercado, desea estimar el promedio del ingreso familiar mensual de una determinada población con una confianza del 95%, así como determinar el error de la estimación.

Se seleccionó para ello una muestra aleatoria de tamaño igual a 80 familias de esta población, el mismo que arrojó un promedio de ingreso familiar de 500 soles.

Se supone que el ingreso familiar mensual se distribuye normalmente con una desviación estándar de 100 soles.

DESARROLLO

Para una confianza del 95 % el valor de la variable aleatoria estandarizada Z=1.96

El intervalo de estimación tiene la siguiente estructura:

$$\bar{x} - Z\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{x} + Z\left(\frac{\sigma}{\sqrt{n}}\right)$$

Reemplazando valores se tiene lo siguiente:

$$500 - 1.96\left(\frac{100}{\sqrt{81}}\right) \leq \mu \leq 500 + 1.96\left(\frac{100}{\sqrt{81}}\right)$$

$$500 - 21.78 \leq \mu \leq 500 + 21.78$$

$$478.22 \leq \mu \leq 521.78$$

Por consiguiente, el ingreso mensual de las familias va de 478.22 a 521.78 soles.

El error de estimación es de 4.36 %

B) ESTIMACIÓN DE LA PROPORCIÓN POBLACIONAL

Se ha realizado una investigación en Lima para conocer la opinión de los Ingenieros de Sistemas sobre un nuevo producto informático que está promocionando la IBM del Perú. En una muestra aleatoria de 400 Ingenieros, 220 declararon que dicho producto es muy bueno, se pide:

a) Hallar el intervalo de confianza del 95 % para la proporción P a favor de dicho producto, en toda la población de Ingenieros de Sistemas de Lima.

b) Determinar el error de estimación de la proporción.

DESARROLLO

a) La estructura del intervalo de estimación de la proporción es:

$$p - Z\left(\sqrt{\frac{p(1-p)}{n}}\right) \leq P \leq p + Z\left(\sqrt{\frac{p(1-p)}{n}}\right)$$

$p = \frac{220}{400} = 0.55;$ q= 1 - 0.55 = 0.45

Reemplazando valores:

$$0.55 - 1.96\left(\sqrt{\frac{(0.55)(0.45)}{400}}\right) \leq P \leq 0.55 + 1.96\left(\sqrt{\frac{(0.55)(0.45)}{400}}\right)$$

$$0.55 - 0.0488 \leq P \leq 0.55 + 0.0488$$

$$0.5012 \leq P \leq 0.5988$$

Por consiguiente la proporción de Ingenieros de Sistemas que opinan a favor de dicho producto va de 0.5012 a 0.5988, o sea entre el 50.12 y 59.88 por ciento aproximadamente.

b) El error de estimación de la proporción es:

E = 0.0488 o sea el 4.88 %