APUNTES SOBRE PRECISIONES DE ESTIMADORES INDIRECTOS DE LA VARIANZA EN MUESTREO

Lic. Pinto Ajhuacho, Jaime Tito

✉ *titojaime_pinto@yahoo.com*

RESUMEN

Los diseños muéstrales plantean procedimientos para conocer el universo de estudio, tamaño y selección de la muestra, y estimación de parámetros poblacionales, así como la precisión de los estimadores, buscando siempre aproximarse al parámetro poblacional.

El minimizar la varianza de los estimadores, sugiere la búsqueda de cálculos de medición de precisión de los estimadores, donde las técnicas de muestreo cada vez van mejorando, tal es el caso de los estimadores indirectos, cuyo aporte ayuda en las encuestas por muestreo.

PALABRAS CLAVE

Estimadores indirectos, varianza de muestreo, regresión lineal, presición.

ABSTRACT

The sample designs propose procedures to know the universe of study, size and selection of the sample, and estimation of population parameters, as well as the accuracy of the estimators, always seeking to approximate the population parameter.

The minimization of the variance of the estimators suggests the search for calculations of precision measurement of the estimators, where sampling techniques are increasingly improving, such is the case of indirect estimators, whose contribution helps in sample surveys.

KEYWORDS

Indirect estimators, sampling variance, linear regression, precision.

1. INTRODUCCIÓN

Un diseño muestral es el bosquejo del conocimiento del universo o población de estudio, es la determinación de una muestra representativa, selección de las unidades de la muestra, estimación de parámetros poblacionales y el cálculo de precisión.

El tamaño de la muestra y precisión de las estimaciones guardan estrecha relación entre sí y cualquier modificación de una de ellas se manifiesta en la otra. La precisión hace referencia a la concentración de los valores estimados en torno al valor que se trata de estimar, de tal forma que la distancia entre el valor a estimar y el valor estimado sea pequeña.

2. ESTIMADORES

Recordando los conceptos de consistencia y estimadores insesgados asintóticos de la teoría general de la inferencia estadística, se puede decir que por lo general un estimador puntual no es idéntico al parámetro que se estima, esto es debido a la presencia del error de muestreo que es dado por $e = \hat{\theta} - \theta$, donde $\hat{\theta}$ es el dato estimado y θ, el parámetro poblacional.

Se espera que un buen estimador tenga su valor muy cercano al valor verdadero del parámetro o por lo menos tenga una alta probabilidad de acercarse. En otras palabras, un buen estimador debe tener la propiedad de ***consistencia***, que dice que una sucesión $\hat{\theta}_n = t(x_1,....,x_n)$ de estimadores de un parámetro

θ es consistente, si dados ε > 0 y δ > 0 tan pequeños como se quiera, existe un número real tan grande N > 0 , tal que $\forall\, n > N$,

se tiene: $P\left(\left| \hat{\theta}_n - \theta \right| > \varepsilon \right) \leq \delta$ o equivalente

$$\lim_{n \to \infty} P\left(\left| \hat{\theta} - \theta \right| > \varepsilon \right) = 0$$

Se observa que no es muy simple de aplicar esta definición para verificar la consistencia de un estimador, en la práctica se dice que $\hat{\theta}_n = t(x_1, \ldots, x_n)$ es una sucesión consistente de estimadores de θ, si:

i).- $\lim_{n \to \infty} E\left[\hat{\theta}_n\right] = \theta$ ii).- $\lim_{n \to \infty} Var\left[\hat{\theta}_n\right] = 0$

Si $\hat{\theta}$ es un estimador insesgado, obviamente la primera condición estará satisfecha.

Si el estimador es **conocido** y es **asintóticamente insesgado** entonces puede ser considerado como **insesgado** cuando ***n*** es bastante grande, luego la distribución de muestreo del estimador puede considerarse fuertemente concentrada en torno a θ cuando ***n*** es bastante grande.

Un tratamiento completo sobre la precisión no se presenta; pero se muestra una idea del marco conceptual para el razonamiento práctico en un diseño muestral. En el presente análisis, se recuerda los estimadores de la media y el total en los métodos indirectos en comparación con el Muestreo Aleatorio Simple[1].

Estimadores del muestreo aleatorio simple

La media muestral $\bar{y}$ es un estimador insesgado de $\bar{Y}$, también $\hat{Y} = N\bar{y}$ es un estimador insesgado del total de la población Y.

La varianza de la media $\bar{y}$ de una muestra aleatoria simple es:

$$V(\bar{y}) = \frac{s^2}{n}\frac{(N-n)}{N} = \frac{s^2}{n}(1-f)$$

El error estándar de $\bar{y}$ es

$$\hat{\sigma}_{\bar{y}} = \frac{s}{\sqrt{n}}\sqrt{(N-n)/N} = \frac{s}{\sqrt{n}}\sqrt{1-f}$$

La varianza de $\hat{Y} = N\bar{y}$, como un estimador del total de la población Y, es

$$\hat{V}(\hat{Y}) = \frac{N^2 s^2}{n}\frac{(N-n)}{N} = \frac{N^2 s^2}{n}(1-f)$$

El error estándar de $\hat{Y}$ es

$$\hat{\sigma}_{\hat{Y}} = \frac{Ns}{\sqrt{n}}\sqrt{(N-n)/N} = \frac{Ns}{\sqrt{n}}\sqrt{1-f}$$

En una muestra aleatoria simple

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$

es una estimación insesgada de S^2.

Límites de confianza

Generalmente se presupone que las estimaciones $\bar{y}$ y $\bar{Y}$ se distribuyen en forma normal alrededor del valor correspondiente de la población. Si la suposición es verdadera, los límites de confianza superior e inferior para la media y total de la población son como sigue:

$$\text{Media:}\quad \begin{aligned} Limite\ Inferior &= \bar{y} - \frac{zs}{\sqrt{n}}\sqrt{1-f} \\ Limite\ Superior &= \bar{y} + \frac{zs}{\sqrt{n}}\sqrt{1-f} \end{aligned}$$

$$\text{Total:}\quad \begin{aligned} Limite\ Inferior &= N\bar{y} - \frac{zNs}{\sqrt{n}}\sqrt{1-f} \\ Limite\ Superior &= N\bar{y} + \frac{zNs}{\sqrt{n}}\sqrt{1-f} \end{aligned}$$

[1] Estractado del libro de "Técnicas de muestreo", William G. Cochran

Probabilidad de confianza (%)

Escenarios	E_1	E_2	E_3	E_4	E_5
*	50	80	90	95	99
**	0,67	1,28	1,64	1,96	2,58

* Nivel de confianza %
** Valor de z asignado.

Estimadores de razón

En el método de razón se obtiene una variable auxiliar x_i, correlacionada con y_i para cada unidad de la muestra. El total de la población X de las x_i debe conocerse.

La estimación de razón de Y, el total de la población para los y_i es:

$$\hat{Y}_R = \frac{\bar{y}}{\bar{x}}.\mu.N$$

Donde y y x son los totales de muestra de y_i y x_i respectivamente.

La estimación de la Razón es $\hat{R} = \frac{\bar{y}}{\bar{x}}$

La varianza estimada del total es:

$$\hat{V}(\hat{Y}_R) = \frac{N^2(1-f)}{n(n-1)}\left(\sum_{i=1}^{n} y_i^2 + \hat{R}^2\sum_{i=1}^{n} x_i^2 - 2\hat{R}\sum_{i=1}^{n} y_i x_i\right)$$

Si la cantidad que se ha de estimar es $\bar{Y}$, el valor de la media de la población de y_i, la estimación de razón es $\hat{\bar{Y}}_R = \frac{\bar{y}}{\bar{x}}.\mu$

Varianza estimada de la media es:

$$\hat{V}(\hat{\bar{Y}}_R) = \frac{(1-f)}{n(n-1)}\left(\sum_{i=1}^{n} y_i^2 + \hat{R}^2\sum_{i=1}^{n} x_i^2 - 2\hat{R}\sum_{i=1}^{n} y_i x_i\right)$$

Límites de confianza

Si la muestra es lo suficientemente grande, de modo que se aplique la aproximación normal, se pueden obtener límites de confianza para el total y la media:

$$\text{Total: } Y: \hat{Y}_R \pm z\sqrt{\hat{V}(\hat{Y}_R)}$$

$$\text{Media: } \bar{Y}: \hat{\bar{Y}}_R \pm z\sqrt{\hat{V}(\hat{\bar{Y}}_R)}$$

Estimadores de regresión lineal

Al igual que la estimación del método de razón, la de regresión lineal se ha diseñado para aumentar la precisión con el uso de una variable auxiliar x_i correlacionada con y_i.

$$\bar{y}_{rl} = \bar{y} + b(\mu - \bar{x})$$

Cuando se examine la relación entre y_i y x_i podrá encontrarse que, aunque se trata de una relación aproximadamente lineal, la línea no pasa por el origen. Esto sugiere una estimación basada sobre la regresión lineal de y_i respecto a x_i.

La estimación de regresión lineal de $\bar{Y}$, media de la población de los y_i es:

$$\hat{Y}_{rl} = N(\bar{y} + b(\mu - \bar{x}))$$

El valor de ***b*** debe calcularse a partir de la muestra, esto es:

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

La varianza estimada de la media es:

$$\hat{V}(\bar{y}_{lr}) = \frac{(1-f)}{n(n-1)}\left(\sum_{i=1}^{n}(y_i - \bar{y})^2 + b^2\sum_{i=1}^{n}(x_i - \bar{x})^2 - 2b\left(\sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i\right)\right)$$

La estimación del total de población Y, es:

$$Y = N(\mu + (\bar{y} - \bar{x}))$$

La varianza estimada del total es:

$$\hat{V}(N\bar{y}_{lr}) = \frac{N^2(1-f)}{n(n-1)}\left(\sum_{i=1}^{n}(y_i - \bar{y})^2 + b^2\sum_{i=1}^{n}(x_i - \bar{x})^2 - 2b\left(\sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i\right)\right)$$

Limites de confianza

Los límites de confianza para el total y la media son:

$$\text{Total: } Y: \hat{Y}_{lr} \pm z\sqrt{\hat{V}(\hat{Y}_{lr})}$$

$$\text{Media: } \bar{Y}: \hat{\bar{Y}}_{lr} \pm z\sqrt{\hat{V}(\hat{\bar{Y}}_{lr})}$$

Estimadores de regresión por diferencia

Estimador de la media es $\bar{y}_{rl} = \bar{y} + (\mu - \bar{x}) = \mu + (\bar{y} - \bar{x})$

La varianza estimada

$$\hat{V}(\bar{y}_{lr}) = \frac{(1-f)}{n(n-2)}\left(\sum_{i=1}^{n}(y_i - \bar{y})^2 - \frac{\left[\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right]^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$

El estimador del total es $\hat{Y}_{rl} = N(\bar{y} + b(\mu - \bar{x}))$

La estimación de la varianza del total es:

$$\hat{V}(N\bar{y}_{lr}) = \frac{N^2(1-f)}{n(n-2)}\left(\sum_{i=1}^{n}(y_i - \bar{y})^2 - \frac{\left[\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right]^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$

Mediante un ejercicio didáctico para los estimadores de la media y el total, se muestra y compara el logro de la ganancia de precisión, gracias a estos métodos de estimación mencionados.

Ejercicio.- Se tiene las Unidades de Producción Agropecuarias (UPAs) de una región del país, para las cuales se plantea el análisis de la estimación de la media y del total de la uperficie cultivada de papa en Has.

Para verificar los Parámetros Poblacionales mediante el muestreo, se recabó la siguiente información:

Información Poblacional.

- N=239 Productores.
- x = Superficie cultivada de papa 2008.
- X = 0,3960 Has. Media superficie poblacional (Año 2008).
- y = Superficie cultivada de papa 2012.
- Y= 96,9410 Has. Total superficie poblacional (Año 2012)
- Y= 0,406 Has. Media superficie poblacional (Año 2012)

Información muestral:

- n=130 Productores

Tabla 1
Cálculo de estimador del total

Descripción	MAS	Razón	Regresión por diferencia	Regresión
R (estimado)		1,0292		
B				0,8375
1-f				0,4561
(1-f)/n				0,0035
Estimador del total	**96,4089**	**97,4118**	**97,3833**	**97,2250**
Estimador de Varianza del total	10,2301	4,5999	4,3220	4,2885
Error de muestreo (Desviación estándar)	3,1985	2,1447	2,0789	2,0709
Límite inferior del estimador total	90,1399	93,2081	93,3086	93,1661
Límite superior del estimador total	102,6779	101,6155	101,4580	101,2839
Rango del intervalo de confianza	12,5379	8,4073	8,1494	8,1178
Diferencia entre valor real y estimado	0,5321	0,4708	0,4423	0,2840
Error relativo (real-estimado)	0,5489%	0,4857%	0,4563%	0,2929%

Fuente: Elaboración propia con datos del ejercicio propuesto.

$$\sum_{i=1}^{n} x = 50{,}9500 \qquad \sum_{i=1}^{n} y = 52{,}4400$$

$$\sum_{i=1}^{n} x^2 = 25{,}4215 \qquad \sum_{i=1}^{n} y^2 = 27{,}7390$$

$$\sum_{i=1}^{n} xy = 25{,}1194 \qquad \sum_{i=1}^{n} (x_i - \bar{x})^2 = 5{,}4530$$

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = 6{,}5855 \qquad \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = 4{,}5670$$

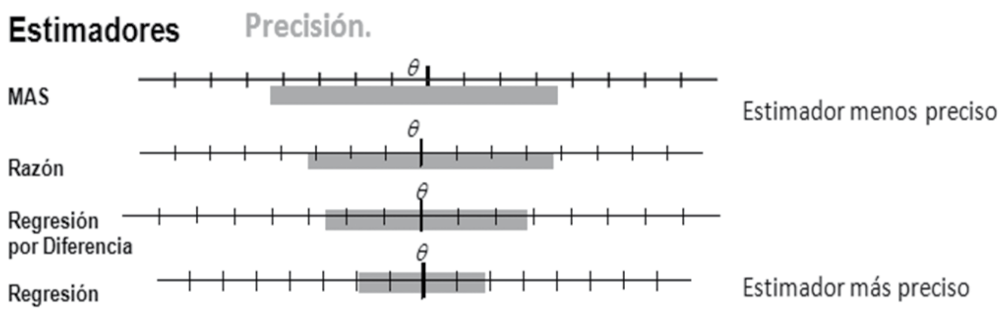
- Covarianza $(x_i y_i) = 0{,}0351$
- Coeficiente de Correlación = 0,7621
- Nivel de confianza z (95%) =1,96

Tabla 2
Cálculo de estimador de la media

Descripción	MAS	Razón	Regresión por diferencia	Regresión
Estimador de la media	**0,4034**	**0,4076**	**0,4075**	**0,4068**
Estimador de Varianza de la media	0,0002	0,0001	0,0001	0,0001
Error de muestreo (Desviación estándar)	0,0134	0,0090	0,0087	0,0087
Límite inferior del estimador de la media	0,3772	0,3900	0,3904	0,3898
Límite superior del estimador de la media	0,4296	0,4252	0,4245	0,4238
Rango del intervalo	0,0525	0,0352	0,0341	0,0340
Diferencia entre valor real y estimado	0,0026	0,0016	0,0015	0,0008
Error relativo (real-estimado)	0,6442%	0,3894%	0,3600%	0,1968%

Fuente: Elaboración propia con datos del ejercicio propuesto.

Gráfico 1
Diagrama de la presición por tipo de estimador



Fuente: Elaboración propia

3. CONCLUSIONES

Sobre la base de la información muestral, y la utilización de los métodos indirectos, el cálculo de los estimadores proporciona, mayor precisión.

El intervalo de confianza, es cada caso es de un rango menor y siempre contiene al parámetro poblacional.

Las aplicaciones de los métodos de estimación indirectos aportan para encontrar valores más próximos posibles al parámetro poblacional.

Con estos métodos, los diseños muestrales se fortalecen, y la práctica de ellas es un aporte estadístico en el manejo de la información.

BIBLIOGRAFÍA

- Abad de Servin Adela, Servin Andrade Luis A. (1978), "*Introducción al muestreo*", Editorial Limusa, México.
- Cochran William G.(1981), *"Técnicas de Muestreo"*, Compañía Editorial Continental, S.A. México, 2da. Impresión.
- Francisco Azorín y José Luis Sánchez Crespo (1986), *"Métodos y aplicaciones del muestreo"*, Alianza Editorial, S.A. Madrid.