

ANÁLISIS DE PRECISIÓN DE ESTIMADORES EN TÉCNICAS DE MUESTREO

Lic. Jaime Tito Pinto Ajhuacho

✉ titojaime_pinto@yahoo.com

RESUMEN

La información permite adquirir el conocimiento necesario para la toma de decisiones en diversas áreas donde se lo requiera, existen diversas técnicas para la captura de datos, como las encuestas por muestreo, que se apoyan diseños muestrales con diversidad de técnicas, en ellos es bueno analizar la precisión del estimador, porque interesa conocer o aproximarse al parámetro poblacional.

Analizar y comentar las técnicas sobre la precisión es importante, sobre todo comparar entre ellas y ver cual se aproxima más al parámetro poblacional.

Para este propósito apoyado en una información muestral, se analizó y comparo técnicas como el Muestreo Aleatorio Simple, Muestreo Estratificado y el Método de Postestratificación.

PALABRAS CLAVE

Comparación de precisión de estimadores

ABSTRACT

The information allows to acquire the necessary knowledge for decision-making in various areas where it is required, there are various techniques for data capture, such as sample surveys, which are supported by sample designs with a variety of techniques, in which it is good to analyze the precision of the estimator, because it is of interest to know or approximate the population parameter.

Analyzing and commenting on the precision techniques is important, especially comparing between them and seeing which one is closest to the population parameter.

For this purpose, supported by sample information, techniques such as Simple Random Sampling, Stratified Sampling and the Post-Stratification Method were analyzed and compared.

KEYWORDS

Estimator precision comparison

La información que es un grupo organizado de datos procesados que integran un mensaje sobre un determinado ente o fenómeno; permiten adquirir el conocimiento necesario para la toma de decisiones en diversas áreas donde se lo requiera.

El Dato, que es una expresión que explica las características de algo que se esté analizando

o simplemente se quiere conocer, es viabilizado acudiendo a diversas técnicas para su captura de datos, como las encuestas por muestreo, que se apoyan diseños muestrales con diversidad de técnicas, en ellos es bueno analizar la precisión del estimador, porque interesa conocer o aproximarse al parámetro poblacional.

Las encuestas por muestreo, al igual que cualquier investigación profunda, se ven afectadas por el error ajeno al muestreo y el error que influye en los resultados de las investigaciones por muestreo, lo constituye los errores muestrales, los que están estrechamente relacionados con el diseño estadístico utilizado para la selección de la muestra; y que mediante un buen esquema de muestreo y proceso de estimación, es posible reducirlos considerablemente.

Por lo que el soporte básico de los estudios por muestreo, es el de proporcionar a partir de una muestra, resultados o estimaciones.

Analizar y comentar las técnicas sobre la precisión es importante, sobre todo comparar entre ellas cual se aproxima más al parámetro poblacional.

Para este propósito apoyado en una información muestral, se analizó y comparo técnicas como la del Muestreo Aleatorio Simple, Muestreo Estratificado y el Método de Postestratificación.

El Muestreo Estratificado brinda una buena precisión con sus estimadores, pero a veces se da el caso que no se dispone de la información para la realización de los estratos y realizamos la aplicación de otra técnica como por ejemplo el Muestreo Aleatorio Simple, y posteriormente podemos con la información muestral establecer los estratos, porque el muestreo estratificado ofrece una variación relativa menor.

Recordemos que el muestreo estratificado realiza primero una partición de la población en subpoblaciones que se denominan estratos, y dentro de cada estrato se realiza el muestreo de forma independiente.

Las utilidades del muestreo estratificado son:

- Sirve cuando se quiere obtener una precisión distinta para cada subpoblación. De esta forma se puede controlar qué muestra pertenece a cada estrato, y así controlar su precisión.
- Se utiliza también cuando es necesario plantear distintas tácticas de muestreo según las subpoblaciones.
- Si los estratos que se utilizan son más homogéneos que la población, la utilización del muestreo estratificado permite ganar precisión frente al aleatorio simple.

Planteando metodologías que minimicen la varianza del estimador, se puede a posteriori asignar las unidades de una muestra a los estratos, que trabaja el Muestreo Estratificado y de este modo obtener mejores estimaciones; se puede advertir que los tamaños muestrales en cada estrato varían de tamaño.

Estratificando a “a posteriori” se puede estimar la media poblacional $\bar{y}$ mediante una media pos estratificada $\bar{y}_{post}$. Así, el tamaño muestra en el estrato h , n_h es aleatorio antes de seleccionar la muestra y fijo una vez seleccionada; el estimador postestratificado será similar al estratificado.

Este método de postestratificación, ofrece una buena precisión frente a otras técnicas de muestreo, se dan algunas condiciones para garantizar su efectividad:

- a).- La muestra debe ser suficientemente grande. Si los estratos varían mucho en tamaño, la muestra debe ser más grande para garantizar que caen suficientes observaciones en cada estrato.
- b).- Las ponderaciones W_h no están exentas

de errores, pues N_h suele ser una estimación, pero se supone que el nivel de error cometido en esta estimación es despreciable.

El estimador de la media y su varianza son:

1).- $\bar{y}_{post}$ es un estimador insesgado de $\bar{y}$

$$2).- V(\bar{y}_{post}) \approx \frac{1-f}{n} \sum_{h=1}^L W_h S_h^2 + \frac{1}{n^2} \sum_{h=1}^L (1-W_h) S_h^2$$

El estimador de la varianza del estimador de la media es:

$$\hat{V}(\bar{y}_{post}) = \frac{1-f}{n} \sum_{h=1}^m W_h \cdot s_h^2 + \frac{1}{n^2} \cdot \sum_{h=1}^m (1-W_h) \cdot s_h^2$$

Donde es s_h^2 información muestral.

Para mostrar las aplicaciones de las técnicas mencionadas, acudimos a una información muestral de una encuesta de Uso de la tierra en el Departamento de Chuquisaca, año 2008, cuya muestra estaba conformada por 834 registros.

La variable que se analizó fue “Superficie cultivada” (Has.) y se procedió a la estimación de la media y el Total, utilizando un nivel de confianza del 95%.

PRECISIÓN POR EL MUESTREO ALEATORIO SIMPLE.-

Estimación de la superficie media cultivada.

N=73.388 Unidades de Producción

n = 84 Unidades de Producción.

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = 2,0814 \text{ Has.}$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = 7,2944$$

$$\hat{V}(\bar{y}) = \frac{s^2}{n} \cdot \frac{(N-n)}{N} = 0,008646$$

$$\sqrt{\hat{V}(\bar{y})} = \sqrt{\frac{s^2}{n} \cdot \frac{(N-n)}{N}} = 0,09299 \text{ Has.}$$

$$\hat{Y}_{Inferior} = \bar{y} - z \sqrt{\hat{V}(\bar{y})} = 1,8991$$

$$\hat{Y}_{Superior} = \bar{y} + z \sqrt{\hat{V}(\bar{y})} = 2,2637$$

$$CV(\bar{y}) = \frac{\sqrt{\hat{V}(\bar{y})}}{\bar{y}} * 100 = 4,4677$$

Estimación del total de Superficie Cultivada.

$$\hat{Y} = N \bar{y} = 73388 (2,0814) = 152.749,8 \text{ Has.}$$

$$\hat{Y}_{Inferior} = N \left(\bar{y} - z \sqrt{\hat{V}(\bar{y})} \right) = 139.371$$

$$\hat{Y}_{Superior} = N \left(\bar{y} + z \sqrt{\hat{V}(\bar{y})} \right) = 166.128$$

$$\hat{V}(\hat{Y}) = \hat{V}(N\bar{y}) = N^2 \hat{V}(\bar{y}) = 46.565.614,21$$

$$\sqrt{\hat{V}(N\bar{y})} = \sqrt{46565614,21} = 6.823,90$$

$$CV(N\bar{y}) = \frac{\sqrt{\hat{V}(N\bar{y})}}{\hat{Y}} * 100 = 4,4677 \%$$

PRECISIÓN POR EL ESTRATIFICADO.

En el *muestreo estratificado*, para el procedimiento de estimación, se consultó sobre una estratificación en el área de estudio, dando el siguiente corte de productores, de 0 a 1 Has.(Pequeños), de 1 a 4 Has.(Medianos) y Mayor a 4 Has.(Grandes), bajo esta referencia se dividió en tres estratos.

Estimación de la media de población:

$$\bar{y}_{st} = \frac{\sum_{h=1}^L N_h \bar{y}_h}{N} = \sum_{h=1}^L W_h \bar{y}_h$$

La varianza de la estimación $\bar{y}_{st}$:

$$V(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h=1}^L N_h(N_h - n_h) \frac{S_h^2}{n_h}$$

$\hat{Y}_{st} = N \cdot \bar{y}_{st}$ es la estimación del Total de la población Y , su varianza es:

$$V(\hat{Y}_{st}) = \sum_{h=1}^L N_h(N_h - n_h) \frac{S_h^2}{n_h}$$

La estimación insesgada de la varianza de $\bar{y}_{st}$ es:

$$\hat{V}(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h=1}^L N_h(N_h - n_h) \frac{\hat{S}_h^2}{n_h}$$

Media de la población: $\bar{y}_{st} \pm z \sqrt{\hat{V}(\bar{y}_{st})}$

Total de la población:

$$N \cdot \bar{y}_{st} \pm N z \sqrt{\hat{V}(\bar{y}_{st})} = N \left(\bar{y}_{st} \pm z \sqrt{\hat{V}(\bar{y}_{st})} \right)$$

Con la información muestral proporcionada, se armó el siguiente Cuadro:

Cuadro 1.

ESTRATIFICACIÓN DE LA SUPERFICIE CULTIVADA (Has.)

	ESTRATO 1	ESTRATO 2	ESTRATO 3	TOTAL
nh	339	410	85	834
$\bar{y}_h$	0,64011	2,1374	7,56011	2,0814
s_h^2	0,09981	0,4951	30,7518	
S_h	0,3159	0,7036	5,5454	
CV	49,28	32,87	72,91	
Nh	29.831	40.038	3.519	73.388
W_h	0,40648	0,54557	0,04795	1
$W_h \bar{y}_h$	0,260196	1,16612	0,362512	1,78883
$\hat{V}(\bar{y}_{st})$				0,0012155

Fuente: Elaboración Propia

El estimador de la media estratificada:

$$\bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h = 1,78883$$

El error de estimación:

$$\hat{V}(\bar{y}_{st}) = \sum_{h=1}^L N_h(N_h - n_h) \frac{\hat{S}_h^2}{n_h} = 0,001216$$

$$\text{Lim Inf}(\bar{y}) = \bar{y}_{st} - z \sqrt{\hat{V}(\bar{y}_{st})} = 1,7205$$

$$\text{Lim Sup}(\bar{y}) = \bar{y}_{st} + z \sqrt{\hat{V}(\bar{y}_{st})} = 1,8572$$

$$\hat{CV}(\bar{y}_{st}) = \frac{\sqrt{\hat{V}(\bar{y}_{st})}}{\bar{y}_{st}} 100 = \frac{\sqrt{0,001216}}{1,78883} 100 = 1,9494\%$$

El Total estimado de Superficie cultivada:

$$\hat{Y}_{st} = N \bar{y}_{st} = 73.388 (1,78883) = 131.278,6 \text{ Has.}$$

El error de estimación:

$$\hat{V}(\hat{Y}_{st}) = N^2 \hat{V}(\bar{y}_{st}) = 73.388^2 (0,001216) = 6.549.13$$

$$\text{Lim Inf}(\hat{Y}) = \hat{Y}_{st} - z \sqrt{\hat{V}(\hat{Y}_{st})} = 126.262,72$$

$$\text{Lim Sup}(\hat{Y}) = \hat{Y}_{st} + z \sqrt{\hat{V}(\hat{Y}_{st})} = 136.294,47$$

$$\hat{CV}(\hat{Y}_{st}) = \frac{\sqrt{\hat{V}(\hat{Y}_{st})}}{\hat{Y}_{st}} 100 = \frac{\sqrt{6.549.131,03}}{131.278,6} 100 = 1,94$$

PRECISIÓN POR POSTESTRATIFICACIÓN

Utilizamos el estimador de la media estratificada.

$$\bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h = 1,78883$$

El error de esta estimación:

$$\hat{V}(\bar{y}_{post}) \approx \frac{1-f}{n} \sum_{h=1}^L W_h s_h^2 + \frac{1}{n^2} \sum_{h=1}^L (1-W_h) s_h^2$$

$$\hat{V}(\bar{y}_{post}) = 0,0008971$$

$$\sqrt{\hat{V}(\bar{y}_{post})} = \sqrt{0,0008971} = 0,0290517$$

$$\text{Lim Inf}(\bar{y}) = \bar{y}_{st} - z \sqrt{\hat{V}(\bar{y}_{post})} = 1,731888$$

$$\text{Lim Sup}(\bar{y}) = \bar{y}_{st} + z \sqrt{\hat{V}(\bar{y}_{post})} = 1,845771$$

$$\hat{CV}(\bar{y}_{post}) = \frac{\sqrt{\hat{V}(\bar{y}_{post})}}{\bar{y}_{post}} 100 = 1,6743 \%$$

Análisis de precisión de estimadores en técnicas de muestreo

La superficie cultivada total estimada:

$$\hat{Y}_{post} = N \bar{y}_{post} = 73.388 (1,78883) = 131.278,6$$

El error de estimación:

$$\hat{V}(\hat{Y}_{post}) = N^2 \hat{V}(\bar{y}_{post}) = 4.831.599,9$$

$$\sqrt{\hat{V}(\hat{Y}_{post})} = \sqrt{4.831.599,9} = 2.198,09$$

$$\text{Lim Inf}(\hat{Y}) = \hat{Y}_{post} - z \sqrt{\hat{V}(\hat{Y}_{post})} = 126.970,34$$

$$\text{Lim Sup}(\hat{Y}) = \hat{Y}_{post} + z \sqrt{\hat{V}(\hat{Y}_{post})} = 135.586,86$$

$$\hat{CV}(\hat{Y}_{post}) = \frac{\sqrt{\hat{V}(\hat{Y}_{post})}}{\hat{Y}_{post}} 100 = 1,6743 \%$$

Se trata de un error muy aceptable. Los resultados de las tres técnicas podemos sintetizarlo en el siguiente cuadro.

Cuadro 2

ANÁLISIS DE LAS ESTIMACIONES

ESTIMACIÓN DE SUPERFICIE CULTIVADA (Has.), DESVIACIÓN ESTANDAR, COEFICIENTE DE VARIACIÓN E INTERVALO DE CONFIANZA EN EL DEPARTAMENTO DE CHUQUISACA, SEGUN VARIABLE DE ESTUDIO, ENCUESTA DE USO DE LA TIERRA AÑO 2008.

TÉCNICA	ESTIMADOR	ESTIMACIÓN DE SUPERFICIE CULTIVADA	DESVIACIÓN ESTÁNDAR	COEFICIENTE DE VARIACIÓN (%)	INTERVALO DE CONFIANZA 95%	
					LÍMITE INFERIOR	LÍMITE SUPERIOR
M.A.S	Media	2,0814	0,0922	4,4677	1,8991	2,2637
Estratificado	Media	1,78883	0,03486	1,9494	1,7205	1,8572
Postestratificado	Media	1,78883	0,0290517	1,6743	1,731888	1,845771
M.A.S	Total	152.749	6.823,90	4,4677	139.371,15	166.128,4
Estratificado	Total	131.278,6	2.559,12	1,9494	126.262,72	136.294,47
Postestratificado	Total	131.278,6	2.198,09	1,6743	126.970,34	135.586,86

Fuente: Elaboración Propia

En el cuadro 2 se puede observar que la superficie cultivada, utilizando el muestreo aleatorio es 152.749 Has. conteniendo un error estándar de 6.823 Has., siendo su coeficiente de variación de 4,46 por ciento.

Muestreo aleatorio simple y la variabilidad es más menor en el Post estratificado, mostrándonos que hay mayor precisión, es decir su estimación está más cerca del parámetro poblacional.

El Estratificado, da un valor estimado de 131.278 Has., con un error estándar de 2.559 Has. y su coeficiente de variación es 1,94 por ciento, y se ve que el Método de Post estratificación, estima la superficie cultivada en 131.278 Has, con un error estándar de 2.198 Has y un coeficiente de variación de 1,67 por ciento.

Corresponde determinar una estimación con cierto nivel de error de muestreo que sea útil para la toma de decisiones, de acuerdo con el grado de fiabilidad que precisa, y se ve que analizando el Método de Postestratificación, ofrece una mayor precisión.

Se puede apreciar $\hat{V}_{POST} < \hat{V}_{ESTRA} < \hat{V}_{MAS}$ el estratificado en variabilidad es menor que el

BIBLIOGRAFÍA

1. Cochran, Willian G. (1996). Técnicas de muestreo. Ed. Continental, 10ed, México, 513p.
2. Pérez López, Cesar. (2000). Técnicas de muestreo estadístico. Alfaomega, México, 603p.
3. Kish, Leslie. (1979). Muestreo de encuestas. Trillas, México, 739p.
4. Azorín Poch, Francisco. (1972). Curso de muestreo y aplicaciones. Aguilar, Madrid, 375p.