ESTUDIO DE VARIOS ESTIMADORES DEL TIPO DE RAZON BAJO EL ENFOQUE DE LA INFERENCIA BASADA EN MODELOS Y CONSIDERANDO DOS DISEÑOS MUESTRALES, EL ALEATORIO SIMPLE Y EL ALEATORIO DOBLE

Larisa Zamora Matamoros*, Dpto. Matemática, Universidad de Oriente

ABSTRACT
In the present work a comparative study between four ratio type estimators is presented. The estimators were obtained under two sampling design: the simple random sampling design and the double sampling design, using the superpopulation approach. The conditions under which one estimator is better than another, from the view point of the mean square error, were established.

Key words: ratio type estimator, double sampling, superpopulation approach.

RESUMEN
En el presente trabajo se realiza un estudio comparativo entre varios estimadores del tipo de razón, obtenidos a través de los diseños muestrales aleatorio simple y doble, empleando el enfoque de la inferencia basada en modelos, estableciéndose condiciones según las cuales algunos de ellos producen, desde el punto de vista del error cuadrático medio, estimaciones más precisas que los otros.

MSC: 62D05

1. INTRODUCCION

El grado de énfasis dedicado al uso de la información auxiliar para mejorar la precisión de las estimaciones es algo que distingue a la teoría del muestreo de la teoría estadística general. Por tal motivo se han desarrollado no solamente métodos de selección de muestras, los cuales incorporan la información auxiliar a través del diseño muestral, sino también métodos de estimación en los que la información auxiliar es incorporada a través de la forma funcional del estimador, como es por ejemplo el δ-estimador, definido por Villagómez en 1992.

En Villagómez se estudia el comportamiento del estimador clásico de razón cuando no se dispone del conocimiento exacto de la información auxiliar, para lo cual el autor se propuso analizar las consecuencias asociadas al utilizar en la estimación $\overline{X}\delta$ en lugar de $\overline{X}$, donde δ es un parámetro desconocido, positivo y distinto de la unidad, definiendo lo que él denominó el δ-estimador.

De la teoría clásica del muestreo se conoce que cuando no se dispone del conocimiento de la media poblacional de la variable auxiliar $(\overline{X})$, resulta necesario reemplazar ésta por un estimador conveniente. Usualmente una muestra grande inicial de tamaño n1 es tomada y la media poblacional $\overline{X}$ es estimada usando la media muestral $\overline{x}_1$, luego una submuestra de tamaño n es tomada a partir de la muestra inicial usando el diseño muestral aleatorio simple. Este procedimiento recibe el nombre de diseño muestral aleatorio doble.

Por lo antes expuesto, nos propusimos investigar, bajo el enfoque de la ξ-inferencia o inferencia basada en modelos, las condiciones para las cuales el δ-estimador produce una estimación más precisa, desde el punto de vista del error cuadrático medio, que la obtenida al utilizar el estimador clásico de razón bajo el muestreo aleatorio doble. Además de presentar la comparación entre estos estimadores, se realizaron también comparaciones entre los siguientes estimadores: $\overline{y}$ (media por unidad), $\overline{y}_R$ (estimador clásico de razón) y $\overline{y}_\delta$ (δ-estimador).

*Email: larisa@csd.uo.edu.cu

2. DESARROLLO

Bajo el diseño muestral aleatorio doble, el estimador clásico de razón adopta la siguiente expresión:

$$\bar{y}_R' = \frac{\bar{y}}{\bar{x}}\bar{x}_1 \qquad (1)$$

donde: $\bar{y}$ y $\bar{x}$ representan las medias muestrales de la característica de interés (Y) y de la variable auxiliar (X) en la submuestra, respectivamente, y

$\bar{x}_1$ es la media de la variable auxiliar en la muestra inicial.

El δ-estimador se define como

$$\bar{y}_\delta = \frac{\bar{y}}{\bar{x}}\bar{X}\delta \qquad (2)$$

donde: $\bar{y}$ y $\bar{x}$ representan las medias muestrales de la característica de interés (Y) y de la variable auxiliar (X), respectivamente,

$\bar{X}$ representa la media poblacional de la variable auxiliar, y

δ es un parámetro desconocido que toma valores en el intervalo (0,1),

Bajo el muestreo aleatorio doble el δ-estimador adopta la siguiente expresión:

$$\bar{y}_\delta' = \frac{\bar{y}}{\bar{x}}\bar{x}_1\delta \qquad (3)$$

donde $\bar{y}$, $\bar{x}$ y $\bar{x}_1$ están definidos como en (1) y δ como en (2).

En el presente trabajo se parte de considerar una población de tamaño N, la cual ha sido extraída de una superpoblación con estructura estocástica dada a través del siguiente modelo:

$$\text{Modelo } \xi\text{:}\ Y_i = \beta X_i + e_i,\quad \xi[e_i] = 0;\quad \xi[e_i e_j] = 0;\quad \xi[e_i^2] = \sigma^2 x_i, \qquad (4)$$

donde $\xi[\cdot]$ denota valor esperado según el modelo ξ.

Para este modelo se obtuvieron las expresiones de la (5) a la (9) para el error cuadrático medio de los estimadores $\bar{y}$, $\bar{y}_R$, $\bar{y}_R'$, $\bar{y}_\delta$ y el $\bar{y}_\delta'$, respectivamente, en las cuales:

$$K = \frac{\sigma^2}{N} + \beta^2\bar{X} \quad \text{y} \quad K_1 = \frac{\sigma^2}{n} + \beta^2\bar{x},$$

$$\xi\left[\bar{y} - \bar{Y}\right]^2 = \bar{x}K_1 - 2\bar{x}K + \bar{X}K \qquad (5)$$

$$\xi\left[\bar{y}_R - \bar{Y}\right]^2 = \frac{\bar{X}^2}{\bar{x}}K_1 - \bar{X}K \qquad (6)$$

$$\xi\left[\bar{y}_R' - \bar{Y}\right]^2 = \frac{\bar{x}_1^2}{\bar{x}}K_1 - 2\bar{x}_1K + \bar{X}K \qquad (7)$$

$$\xi\left[\bar{y}_\delta - \bar{Y}\right]^2 = A\delta^2 + B\delta + C \qquad (8)$$

donde:

$$A = \frac{\bar{X}^2}{\bar{x}} K_1;\ B = -2\bar{X}K \text{ y } C = \bar{X}K\ ,$$

$$\xi\left[\bar{y}'_{\delta} - \bar{Y}\right]^2 = A'\delta^2 + B'\delta + C' \tag{9}$$

donde:

$$A' = \frac{\bar{x}_1^2}{\bar{x}} K_1;\quad B' = -2\bar{x}_1 K \text{ y } C' = \bar{X}K\ ,$$

A continuación se presentan los resultados obtenidos al comparar entre sí estos estimadores.

Comparación entre la media por unidad y el estimador clásico de razón

De las expresiones (5) y (6) obtenemos que:

$$\xi\left[\bar{y}_R - \bar{Y}\right]^2 - \xi\left[\bar{y} - \bar{Y}\right]^2 = (\bar{X} - \bar{x})\left\{\frac{K_1}{\bar{x}}\left(\bar{X} + \bar{x}\right) - 2K\right\} \tag{10}$$

de donde se puede establecer que:

- *si* $\bar{X} > \bar{x}$ *y* $K_1(\bar{x} + \bar{X}) < 2K\bar{x}$ *o si* $\bar{X} < \bar{x}$ *y* $K_1(\bar{x} + \bar{X}) > 2K\bar{x}$, *la estimación obtenida por el estimador clásico de razón es más precisa que la obtenida a través de la media por unidad.*

Comparación entre el estimador clásico de razón bajo el muestreo aleatorio doble, y el estimador clásico de razón

De las expresiones (5) y (6) obtenemos que:

$$\xi\left[\bar{y}'_R - \bar{Y}\right]^2 - \xi\left[\bar{y}_R - \bar{Y}\right]^2 = \left(\bar{x}_1 - \bar{X}\right)\left\{\frac{K_1}{\bar{x}}\left(\bar{x}_1 + \bar{X}\right) - 2K\right\} \tag{11}$$

de donde se puede establecer que:

- *si* $\bar{X} > \bar{x}_1$ *y* $K_1(\bar{X} + \bar{x}_1) > 2K\bar{x}$ *o si* $\bar{X} < \bar{x}_1$ *y* $K_1(\bar{X} + \bar{x}_1) < 2K\bar{x}$, *la estimación obtenida por el estimador clásico de razón, bajo el muestreo aleatorio doble, es más precisa que la obtenida a través del estimador clásico de razón.*

Comparación entre el estimador clásico de razón bajo el muestreo aleatorio doble y el estimador media por unidad

De las expresiones (5) y (7) obtenemos que:

$$\xi\left[\bar{y}'_R - \bar{Y}\right]^2 - \xi\left[\bar{y} - \bar{Y}\right]^2 = \left(\bar{x}_1 - \bar{x}\right)\left\{\frac{K_1}{\bar{x}}\left(\bar{x}_1 + \bar{x}\right) - 2K\right\} \tag{12}$$

por lo tanto,

- *si* $\bar{x}_1 > \bar{x}$ *y* $K_1(\bar{x}_1 + \bar{x}) < 2K\bar{x}$ *o si* $\bar{x}_1 < \bar{x}$ *y* $K_1(\bar{x}_1 + \bar{x}) > 2K\bar{x}$, *la estimación obtenida por estimador clásico de razón, bajo el muestreo aleatorio doble, es más precisa que la obtenida a partir de la media por unidad.*

De las comparaciones anteriores podemos concluir que:

- *Si* $\bar{x} < \bar{X} < \bar{x}_1$ *y* $K_1\bar{x}_1 < K\bar{x}$, *o si* $\bar{x}_1 < \bar{X} < \bar{x}$ *y* $K\bar{x} < K_1\bar{x}_1$, *entonces* $\bar{y}'_R << \bar{y}_R << \bar{y}$, donde "<<" lo usaremos para significar menor error cuadrático medio según modelo, es decir, bajo estas condiciones *el estimador clásico de razón, bajo el muestreo aleatorio doble, es más preciso que el clásico de razón, y éste más que la media por unidad.*

Comparación entre el δ-estimador y el clásico de razón bajo el muestreo aleatorio doble

De (7) y (8) obtenemos que:

$$\xi\left[\bar{y}_\delta-\bar{Y}\right]^2-\xi\left[\bar{y}'_R-\bar{Y}\right]^2=A\delta^2+B\delta+\tilde{C} \tag{13}$$

donde A y B coinciden con las expresiones dadas en (8) y $\tilde{C}=2\bar{x}_1K-\dfrac{\bar{x}_1^2}{\bar{x}}K_1$.

La expresión (13) representa una forma cuadrática, la cual posee dos raíces reales dadas por:

$$\delta_1^a=\frac{2\bar{x}K-\bar{x}_1K_1}{\bar{X}K_1}\quad \text{y}\quad \delta_2^a=\frac{\bar{x}_1}{\bar{X}}$$

y como el coeficiente del término cuadrático (A) es una magnitud positiva, se obtiene que:

$$\xi\left[\bar{y}_\delta-\bar{Y}\right]^2<\xi\left[\bar{y}'_R-\bar{Y}\right]^2\quad \text{para}\quad \delta\in\left(\delta_1^a,\delta_2^a\right),$$

es decir, *para valores del parámetro δ pertenecientes al intervalo antes definido; se obtendrá a partir del δ-estimador, una estimación más precisa que la obtenida con el clásico de razón bajo el muestreo aleatorio doble.*

Comparación entre el δ-estimador bajo el muestreo aleatorio doble y el clásico de razón

De (6) y (9) obtenemos que:

$$\xi\left[\bar{y}'_\delta-\bar{Y}\right]^2-\xi\left[\bar{y}_R-\bar{Y}\right]^2=A\delta^2+B\delta+\approx{C} \tag{14}$$

donde A y B están definidos como en (9) y $\approx{C}=2\bar{X}K-\dfrac{\bar{X}^2}{\bar{x}}K_1$.

Esta ecuación cuadrática posee dos raíces reales dadas por:

$$\delta_1^b=\frac{2\bar{x}K-\bar{X}K_1}{\bar{x}_1K_1}\quad \text{y}\quad \delta_2^b=\frac{\bar{X}}{\bar{x}_1},$$

y como el coeficiente del término cuadrático es no negativo, obtenemos que:

$$\xi\left[\bar{y}'_\delta-\bar{Y}\right]^2<\xi\left[\bar{y}_R-\bar{Y}\right]^2\quad \text{para}\quad \delta\in\left(\delta_1^b,\delta_2^b\right),$$

es decir, *para valores del parámetro δ pertenecientes al intervalo antes especificado, el δ-estimador bajo el muestreo aleatorio doble produce una estimación de la media poblacional más precisa que la obtenida a través del estimador clásico de razón.*

Comparación entre el δ-estimador bajo el muestreo aleatorio doble y el δ estimador bajo el muestreo aleatorio simple

De (8) y (9) obtenemos que:

$$\xi\left[\bar{y}'_\delta-\bar{Y}\right]^2-\xi\left[\bar{y}_\delta-\bar{Y}\right]^2=\left(\bar{x}_1-\bar{X}\right)\delta\left\{\frac{K_1}{\bar{x}}\left(\bar{x}_1+\bar{X}\right)\delta-2K\right\} \tag{15}$$

por lo tanto:

- *si* $\overline{x}_1 > \overline{X}$ *y* $\delta < \delta^c$ *o si* $\overline{x}_1 < \overline{X}$ *y* $\delta > \delta^c$, *la estimación obtenida por el δ-estimador bajo el muestreo aleatorio doble, es más precisa que la obtenida bajo el muestreo aleatorio simple,*

$$\delta^c = \frac{2K\overline{x}}{K_1(\overline{x}_1 + \overline{X})}.$$

Comparación entre el δ-estimador y el estimador clásico de razón, bajo el muestreo aleatorio doble

De (7) y (9) obtenemos que:

$$\xi\left[\overline{y}_\delta^{'} - \overline{Y}\right]^2 - \xi\left[\overline{y}_R^{'} - \overline{Y}\right]^2 = A\delta^2 + B\delta + \overset{\equiv}{C} \tag{16}$$

donde A y B están definidos como en (9) y $\overset{\equiv}{C} = 2\overline{x}_1 K - \frac{\overline{x}_1^2}{\overline{x}} K_1$.

Esta ecuación cuadrática posee dos raíces reales dadas por:

$$\delta_1^d = \frac{2\overline{x}K - \overline{x}_1 K_1}{\overline{x}_1 K_1} \text{ y } \delta_2^d = 1,$$

y como el coeficiente del término cuadrático es no negativo, obtenemos que:

$$\xi\left[\overline{y}_\delta^{'} - \overline{Y}\right]^2 < \xi\left[\overline{y}_R^{'} - \overline{Y}\right]^2 \quad \text{si} \quad \delta \in \left(\delta_1^d, 1\right).$$

es decir, para valores de δ pertenecientes al intervalo $(\delta_1^d, 1)$, *el δ-estimador bajo el muestreo aleatorio doble produce una estimación, de la media poblacional más precisa que la obtenida a través del estimador clásico de razón utilizando el mismo diseño.*

De las comparaciones anteriores podemos resumir que:

1. Si se satisfacen las desigualdades siguientes:

D1: $K\overline{x} < K_1\overline{x}_1$,

D2: $\overline{x}_1 < \overline{X}$,

D3: $\overline{X} < \overline{x}$,

D4: $K_1\overline{x} < 2K\overline{x}$,

se cumplirá que:

- $\delta_1^a > 0$ y $\delta_1^d > 0$ por D2, D3 y D4; $\delta_1^b > 0$ por D3 y D4, y $\delta^c > 0$ independientemente de que se satisfagan o no las condiciones D1-D4.
- $\delta_1^a < \delta_2^a < 1$, por lo tanto, para $\delta \in (\delta_1^a, \delta_2^a)$, $\overline{y}_\delta << \overline{y}_R^{'} << \overline{y}_R << \overline{y}$,
- $\delta_1^b < 1$ y $\delta_2^b > 1$, entonces para $\delta \in (\delta_1^b, 1)$, $\overline{y}_\delta^{'} << \overline{y}_R << \overline{y}$,
- $\delta^c < 1$, por lo que para $\delta > \delta^c$, $\overline{y}_\delta^{'} << \overline{y}_\delta$,
- $\delta_1^d < 1$, por lo que para $\delta \in (\delta_1^d, 1)$, $\overline{y}_\delta^{'} << \overline{y}_R^{'}$,

- $\delta_1^a < \delta^c$, pero $\delta^c < \delta_2^a$ si se cumple que $2K\bar{x}\bar{X} < K_1\bar{x}_1(\bar{x}_1 + \bar{X})$, desigualdad que denotaremos por D5, e implica, junto con D2, la desigualdad D1; por lo que:

  - si se cumplen las desigualdades D2-D4, $\bar{y}'_\delta \ll \bar{y}_\delta \ll \bar{y}'_R \ll \bar{y}_R \ll \bar{y}$ para $\delta \in (\delta^c, \delta_2^a)$,
  - si se satisfacen las desigualdades D1-D3, pero no la D4, entonces como $\delta_1^b < \delta^c$ y $\delta_1^d > \delta^c$, se cumplirá que para $\delta \in (\delta_1^d, 1)$, $\bar{y}'_\delta \ll \bar{y}'_R \ll \bar{y}_R \ll \bar{y}$, $\bar{y}'_\delta \ll \bar{y}_\delta$; $\bar{y}'_R \ll \bar{y}_\delta$.

2. Si se satisfacen las desigualdades siguientes:

D6: $K\bar{x} < K_1\bar{X}$,

D7: $\bar{x}_1 > \bar{X}$,

D8: $\bar{x}_1 < \bar{x}$,

D9: $K_1\bar{x}_1 < 2K\bar{x}$,

se cumplirá que:

- $\delta_1^a > 0$ y $\delta_1^d > 0$ por D9; $\delta_1^b > 0$ por D7 y D9, y $\delta^c > 0$ independientemente de que se satisfagan o no las desigualdades D6-D9.
- $\bar{y}_R \ll \bar{y}'_R$, $\bar{y}_R \ll \bar{y}$, $\bar{y}'_R \ll \bar{y}$,
- $\delta_1^a < 1$ y $\delta_2^a > 1$, por lo tanto, para $\delta \in (\delta_1^a, 1)$ se cumplirá que $\bar{y}_\delta \ll \bar{y}'_R$,
- $\delta_1^b < \delta_2^b < 1$, entonces para $\delta \in (\delta_1^b, \delta_2^b)$ se cumplirá que $\bar{y}'_\delta \ll \bar{y}_R$,
- $\delta^c < 1$, por lo que para $\delta < \delta^c$, $\bar{y}'_\delta \ll \bar{y}_\delta$,
- $\delta_1^d < 1$, por lo que para $\delta \in (\delta_1^d, 1)$ se cumplirá que $\bar{y}'_\delta \ll \bar{y}'_R$,
- $\delta_1^a < \delta^c$, por lo que para $\delta \in (\delta_1^a, \delta^c)$, $\bar{y}'_\delta \ll \bar{y}_\delta \ll \bar{y}'_R$; $\bar{y}_R \ll \bar{y}'_R$,
- $\delta_1^a < \delta_1^b$, $\delta_1^b < \delta_2^b$, $\delta_1^b < \delta^c$ y $\delta^c < \delta_2^b$ si se satisface la desigualdad $2K\bar{x}\bar{x}_1 < K_1\bar{X}(\bar{x}_1 + \bar{X})$, la cual denotaremos por D10. Por lo que:

  - si se satisfacen las desigualdades D6-D10, $\bar{y}'_\delta \ll \bar{y}_\delta \ll \bar{y}'_R$; $\bar{y}'_\delta \ll \bar{y}_R \ll \bar{y}$; $\bar{y}_R \ll \bar{y}'_R$ para $\delta \in (\delta_1^b, \delta^c)$.
  - si se satisfacen las desigualdades D6-D8 y D10, pero no se satisface D9, $\bar{y}'_\delta \ll \bar{y}_\delta \ll \bar{y}'_R$; $\bar{y}'_\delta \ll \bar{y}_R \ll \bar{y}$; $\bar{y}_R \ll \bar{y}'_R$ para $\delta \in \left(\delta_1^b, \delta_2^b\right)$.

Debemos señalar que en el caso 1, si la desigualdad D4 no se satisface, en los intervalos confidenciales obtenidos para el parámetro δ, que posean su extremo inferior distinto de δ^c, éste deberá ser reemplazado por cero, ya que en estos casos dichos extremos son magnitudes negativas y δ es una magnitud positiva. Lo mismo ocurrirá si en el caso 2 no se verifica la desigualdad D9.

3. CONCLUSIONES

En este trabajo se ha realizado un estudio comparativo entre los estimadores: media por unidad, clásico de razón, clásico de razón bajo el muestreo aleatorio doble, δ estimador y δ estimador bajo el muestreo aleatorio doble, considerando el enfoque de la inferencia basada en modelos. Dicho estudio nos permitió establecer condiciones para las cuales unos estimadores son más precisos que los restantes, desde el punto de vista del error cuadrático medio.

REFERENCIAS

1. COCHRAN, W.G. (1977): **Sampling techniques**, 3ra ed. John Wiley. New York.

2. HEDAYAT, A.S. and B.K. SINHA (1991): **Design and inference in finite population sampling**, John Wiley, New York.

3. VILLAGOMEZ, J. (1992): "Una contribución en el uso de la información auxiliar en la estimación del tipo de razón", Tesis de Doctorado, Universidad de La Habana, Cuba.

4. ZAMORA, L.M. (1997): "Estudio de dos estimadores del tipo de razón bajo los enfoques de la $p\xi$- y la $\xi-$inferencia", Tesis de Doctorado, Universidad de Oriente, Cuba.