QÜESTIIÓ, Vol 17, 2 pp. 203-219, 1993

NUEVOS ESTIMADORES DE LA VARIANZA EN POBLACIONES FINITAS

M. RUIZ ESPEJO*

Universidad Complutense de Madrid

Obtenemos una expresión de la varianza de una población finita en función de los tamaños relativos, varianzas y medias de los estratos o conglomerados en que puede ser dividida la población. Como consecuencia de esta nueva expresión, podemos desarrollar varios estimadores consistentes y no negativos de la varianza poblacional en muestreo estratificado y muestreo por conglomerados con o sin submuestreo. En cada caso, los estimadores de la varianza poblacional son insesgados o conservativos (en el sentido de Wolter, 1985). También se derivan dos nuevos controles de la estimación de la media poblacional en la línea de Ruiz (1987). Finalmente, comparamos los estimadores de la varianza de las estrategias intermedias propuestas por Ruiz y Santos (1989) con respecto al clásico estimador de grupos aleatorios (Wolter, 1985). El primero resulta asintóticamente más preciso si el tamaño n de cada muestra parcial independiente crece suficientemente, cuando el tamaño poblacional N es muy grande.

New estimators of the variance in finite populations.

Key words: Control de la estimación, estrategias intermedias, insesgación, muestreo de poblaciones finitas, propiedades de los estimadores, varianza.

AMS Classification: 62 D 05.

---

*M. Ruiz Espejo. Departamento de Estadística e Investigación Operativa. Facultad de Ciencias Económicas y Empresariales. Universidad Complutense (Campus de Somosaguas). 28223 Madrid.

-Article rebut el setembre de 1992.
-Acceptat el juny de 1993.

INTRODUCCIÓN

Usualmente las técnicas de muestreo se enfocan a la estimación de la media poblacional. Aunque se han dedicado menos espacios a la estimación de la varianza poblacional o de la varianza de los estimadores de la media poblacional, los avances recientes en este campo prometen un mayor aprovechamiento de la información proporcionada por muestreo.

En este artículo presentamos avances metodológicos en la estimación de la varianza poblacional en el muestreo estratificado y por conglomerados de poblaciones finitas (una referencia bibliográfica del tema es el trabajo de Ruiz y Ruiz, 1992), en la sección 1.

También en la sección 2 justificamos la utilidad del estimador de la varianza para estrategias intermedias sugerido por Ruiz y Santos (1989) que asintóticamente será mejor en precisión que el estimador clásico por grupos aleatorios (una referencia útil de estimadores de la varianza es la de Wolter, 1985).

1. ESTIMACIÓN DE LA VARIANZA POBLACIONAL

1.1. Planteamiento

A lo largo de la historia matemática y estadística se han ido ofreciendo varias expresiones de la varianza de una población finita. Una de ellas, debida a Ruiz (1987), es concretamente

$$\sigma^2 = \sum_{h=1}^{L} P_h \sigma_h^2 + \sum_{h=1}^{L-1} P_h \left(\mu_h - \frac{\sum_{h=1}^{L-1} P_h \mu_h}{1 - P_L} \right)^2 + \frac{1}{P_L (1 - P_L)} \left[\sum_{h=1}^{L-1} P_h (\mu_h - \mu) \right]^2 , \tag{1.1}$$

donde L es el número de estratos o conglomerados, $P_h = N_h/N$ es el peso relativo del estrato h, N_h es el tamaño del estrato h y N el tamaño de la población finita. También μ_h y μ representan la media del estrato h y de la población global respectivamente. Finalmente σ_h^2 es la varianza del estrato h.

Esta nueva relación, (1.1), permitió diseñar un control o comprobación de la estimación de la media poblacional en el muestreo estratificado estándar.

Además, ahora podemos desarrollar una nueva fórmula de la varianza poblacional inspirándonos en una relación clásica,

$$N^2\sigma^2 = \sum_{i<j}^{N}\sum (X_i - X_j)^2 \tag{1.2}$$

que puede verse por ejemplo en Murthy (1963) o Chaudhuri (1978), donde N es el tamaño de la población finita, y σ^2 es la varianza de la población finita.

La anterior relación clásica (1.2) permitió a Murthy (1963) proponer un estimador de la varianza poblacional, insesgado y no negativo, razonando de un modo similar a como se hace con el estimador Horvitz-Thompson (1952),

$$\hat{\sigma}^2 = \frac{1}{N^2}\sum_{i<j\in s}\sum \frac{(X_i - X_j)^2}{\pi_{ij}}$$

siendo s la muestra seleccionada de acuerdo con un diseño p no ordenado (Cassel *et al.*, 1977) y π_{ij} es la probabilidad de inclusión de las unidades i y j en la muestra; la admisibilidad de este estimador fue justificada por Sankaranarayanan (1980). Posteriormente Liu y Thompson (1983) trataron de nuevo este problema.

1.2. Nueva expresión de la varianza poblacional

Como es bien conocido, la varianza de una población finita admite la descomposición usual de la varianza total en variación dentro de estratos y variación entre estratos siguiente

$$\sigma^2 = \sum_{h=1}^{L} P_h\sigma_h^2 + \sum_{h=1}^{L} P_h(\mu_h - \mu)^2. \tag{1.3}$$

De (1.2) tenemos

$$\frac{1}{N}\sum_{h=1}^{L} N_h(\mu_h - \mu)^2 = \frac{1}{N^2}\sum_{h<g}^{L}\sum N_h N_g(\mu_h - \mu_g)^2,$$

y ahora de (1.3) podemos sustituir la última relación para concluir que

$$\sigma^2 = \sum_{h=1}^{L} P_h\sigma_h^2 + \sum_{h<g}^{L}\sum P_h P_g(\mu_h - \mu_g)^2. \tag{1.4}$$

Como consecuencia de esta nueva descomposición del análisis de la varianza, podemos construir varios estimadores consistentes, no negativos y conservativos de la varianza poblacional, así como insesgados; en la sección siguiente 1.3 damos dos controles de la estimación de la media poblacional en muestreo estratificado usual.

1.3. Dos nuevas expresiones de la media poblacional

Igualando los segundos miembros de (1.1) y de (1.4), tenemos

$$\sum_{h<g}^{L}\sum P_h P_g(\mu_h - \mu_g)^2 =$$

$$= \sum_{h=1}^{L-1} P_h \left(\mu_h - \frac{\sum_{h=1}^{L-1} P_h \mu_h}{1 - P_L} \right)^2 + \frac{1}{P_L(1 - P_L)} \left[\sum_{h=1}^{L-1} P_h \mu_h - (1 - P_L)\mu \right]^2 .$$

De aquí, tenemos despejando

$$\mu = \frac{1}{1 - P_L} \left\{ \sum_{h=1}^{L-1} p_h \mu_h \mp \right.$$

$$\left. \mp \sqrt{P_L(1 - P_L) \left[\sum_{h<g}^{L}\sum P_h P_g(\mu_h - \mu_g)^2 - \sum_{h=1}^{L-1} P_h \left(\mu_h - \frac{\sum_{h=1}^{L-1} P_h \mu_h}{1 - P_L} \right)^2 \right]} \right\}$$

(1.5)

y, como en Ruiz (1987), si el orden de las medias de los estratos es creciente, o más simplemente si

$$\mu_L > \frac{1}{1 - P_L} \sum_{h=1}^{L-1} P_h \mu_h ,$$

omitiremos el signo menos previo a la raíz cuadrada de (1.5). Otra expresión válida para la media poblacional es

$$\mu = \pm \sqrt{\sum_{h=1}^{L} P_h \mu_h^2 - \sum_{h<g}^{L}\sum P_h P_g(\mu_h - \mu_g)^2}, \tag{1.6}$$

fórmula más sencilla que (1.5), porque (1.6) no requiere los cálculos tan complejos como (1.5). Obviamente la relación más simple es la clásica

$$\mu = \sum_{h=1}^{L} P_h \mu_h. \tag{1.7}$$

No obstante, las fórmulas (1.5) y (1.6) pueden ser consideradas como nuevos controles del cálculo de la media poblacional μ, y por tanto de su estimación al sustituir μ_h por $\hat{\mu}_h$ en (1.5), (1.6) ó (1.7) en el sentido propuesto por Ruiz (1987).

1.4. Estimación de la varianza en muestreo por conglomerados

1.4.1. Un estimador insesgado en muestreo por conglomerados sin submuestreo

Si tenemos a la población finita de tamaño N clasificada en L conglomerados de tamaños $N_h (h = 1, 2, \ldots, L)$ con

$$N = \sum_{h=1}^{L} N_h,$$

un estimador insesgado y no negativo de la varianza poblacional σ^2 en muestreo por conglomerados sin submuestreo es

$$\hat{\sigma}^2_{\text{cl}} = \sum_{h \in s_1} P_h \frac{\sigma_h^2}{\pi_h} + \sum \sum_{h<g \in s_1} P_h P_g \frac{(\mu_h - \mu_g)^2}{\pi_{hg}}, \tag{1.8}$$

donde s_1 es la muestra de unidades primarias, y π_h y $\pi_{hg} (> 0)$ son las probabilidades de inclusión de las unidades primarias h y h, g respectivamente en la muestra. El estimador (1.8) es exactamente insesgado, pues si introducimos la variable aleatoria auxiliar

$$e_h = \begin{cases} 1 & \text{si } h \in s_1 \\ 0 & \text{si } h \notin s_1, \end{cases}$$

entonces $\mathrm{E}(e_h) = \pi_h$, y $\mathrm{E}(e_h\, e_g) = \pi_{hg}$, y consecuentemente

$$\begin{aligned} \mathrm{E}(\hat{\sigma}^2_{\text{cl}}) &= \sum_{h=1}^{L} P_h \frac{\sigma_h^2}{\pi_h} \mathrm{E}(e_h) + \sum \sum_{h<g}^{L} P_h P_g \frac{(\mu_h - \mu_g)^2}{\pi_{hg}} \mathrm{E}(e_h\, e_g) = \\ &= \sum_{h=1}^{L} P_h \sigma_h^2 + \sum \sum_{h<g}^{L} P_h P_g (\mu_h - \mu_g)^2 = \sigma^2, \end{aligned}$$

debido a (1.4).

1.4.2. Un estimador conservativo, consistente y no negativo en muestreo por conglomerados con submuestreo

Puede verse directamente que el estimador no negativo de σ^2 (si $\hat{\sigma}_h^2$ son no negativos),

$$\hat{\sigma}^2_{\text{cls}} = \sum_{h\in s_1} P_h \frac{\hat{\sigma}_h^2}{\pi_h} + \sum\sum_{h<g\in s_1} P_h P_g \frac{(\hat{\mu}_h - \hat{\mu}_g)^2}{\pi_{hg}}, \tag{1.9}$$

es consistente si $\hat{\sigma}_h^2$ y $\hat{\mu}_h$ son consistentes respectivamente para σ_h^2 y μ_h. Hemos denotado s_1 a la muestra de conglomerados o unidades primarias, y π_h y π_{hg} son las probabilidades de inclusión de unidades primarias en la muestra. Seguidamente, $\hat{\sigma}_h^2$ y $\hat{\mu}_h$ se calculan a partir de las unidades secundarias obtenidas por muestreo, asumiendo que estos estimadores son insesgados respectivamente para σ_h^2 y μ_h. Entonces, $\hat{\sigma}^2_{\text{cls}}$ sobreestima a σ^2,

$$\begin{aligned}
\mathrm{E}(\hat{\sigma}^2_{\text{cls}}) &= \mathrm{E}_1\mathrm{E}_2\left[\sum_{h=1}^{L} P_h \frac{\hat{\sigma}_h^2}{\pi_h} e_h + \sum\sum_{h<g}^{L} P_h P_g \frac{(\hat{\mu}_h - \hat{\mu}_g)^2}{\pi_{hg}} e_h e_g\right] = \\
&= \mathrm{E}_1\left[\sum_{h=1}^{L} P_h \frac{\mathrm{E}_2(\hat{\sigma}_h^2)}{\pi_h} e_h + \sum\sum_{h<g}^{L} P_h P_g \frac{\mathrm{E}_2(\hat{\mu}_h - \hat{\mu}_g)^2}{\pi_{hg}} e_h e_g\right] \geq \\
&\geq \mathrm{E}_1\left[\sum_{h=1}^{L} P_h \frac{\sigma_h^2}{\pi_h} e_h + \sum\sum_{h<g}^{L} P_h P_g \frac{(\mu_h - \mu_g)^2}{\pi_{hg}} e_h e_g\right] = \sigma^2,
\end{aligned}$$

porque $\mathrm{E}_1(e_h) = \pi_h$ y $\mathrm{E}_1(e_h\, e_g) = \pi_{hg}$ en la última igualdad, aplicando previamente la desigualdad de Jensen. Además, $\hat{\sigma}^2_{\text{cls}}$ generaliza a $\hat{\sigma}^2_{\text{cl}}$ dado en (1.8).

1.5. Estimación de la varianza en muestreo estratificado

1.5.1. Estimadores conservativos y consistentes en muestreo estratificado

El estimador no negativo (cuando son $\hat{\sigma}_h^2$ no negativos),

$$\hat{\sigma}^2_{\text{st}} = \sum_{h=1}^{L} P_h \hat{\sigma}_h^2 + \sum\sum_{h<g}^{L} P_h P_g (\hat{\mu}_h - \hat{\mu}_g)^2 \tag{1.10}$$

es consistente para estimar la varianza poblacional σ^2, si los estimadores $\hat{\mu}_h$ y $\hat{\sigma}_h^2$ son consistentes. Por ejemplo, siendo

$$\hat{\mu}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} X_{hi}, \quad \text{y}$$

$$\hat{\sigma}_h^2 = \frac{N_h - 1}{N_h} \cdot \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(X_{hi} - \hat{\mu}_h)^2,$$

donde n_h es el tamaño muestral en el estrato h, y donde las observaciones obtenidas en este estrato por muestreo aleatorio simple sin reemplazamiento (mas) son $X_{h1}, X_{h2}, \ldots, X_{hn_h}$. Si además, como en el ejemplo, $\hat{\mu}_h$ y $\hat{\sigma}_h^2$ son insesgados, el estimador $\hat{\sigma}_{st}^2$ es conservativo en el sentido de que sobreestima σ^2.

En efecto,

$$\begin{aligned} \mathrm{E}(\hat{\sigma}_{st}^2) &= \mathrm{E}\left[\sum_{h=1}^{L} P_h\hat{\sigma}_h^2 + \sum_{h<g}\sum^{L} P_h P_g(\hat{\mu}_h - \hat{\mu}_g)^2\right] = \\ &= \sum_{h=1}^{L} P_h\ \mathrm{E}(\hat{\sigma}_h^2) + \sum_{h<g}\sum^{L} P_h P_g\ \mathrm{E}(\hat{\mu}_h - \hat{\mu}_g)^2 \geq \\ &\geq \sum_{h=1}^{L} P_h\sigma_h^2 + \sum_{h<g}\sum^{L} P_h P_g(\mu_h - \mu_g)^2 = \sigma^2, \end{aligned}$$

porque $\mathrm{E}(\hat{\sigma}_h^2) = \sigma_h^2$ y $\mathrm{E}(\hat{\mu}_h - \hat{\mu}_g)^2 \geq [\mathrm{E}(\hat{\mu}_h) - \mathrm{E}(\hat{\mu}_g)]^2 = (\mu_h - \mu_g)^2$, por la desigualdad de Jensen.

Otro estimador diferente, consistente y no negativo (cuando $\hat{\sigma}_h^2$ son no negativos) es

$$\hat{\sigma}_{st}^{2\prime} = \sum_{h=1}^{L} P_h\hat{\sigma}_h^2 + \sum_{h=1}^{L} P_h(\hat{\mu}_h - \hat{\mu}_{st})^2 \tag{1.11}$$

el cual es conservativo en las mismas hipótesis que el anterior dado en (1.10).

Los estimadores (1.10) y (1.11) son mucho más simples en la práctica que los dados por Mirás (1985) y Hedayat y Sinha (1991), si bien estos son insesgados y todos ellos aplicables en muestreo estratificado.

1.5.2. Estimación insesgada en muestreo estratificado

Una visión integradora de estimadores insesgados de la varianza poblacional en muestreo estratificado puede derivarse de la relación (1.4). En efecto, si $\overline{x}_h$

es la media muestral en el estrato h, tenemos que si $h \neq g$, con diseño de muestreo aleatorio simple con reemplazamiento (masr) o sin reemplazamiento (mas) dentro de los estratos,

$$\mathrm{E}(\overline{x}_h - \overline{x}_g)^2 = \mathrm{E}(\overline{x}_h^2 - 2\overline{x}_h\overline{x}_g + \overline{x}_g^2) = (\mu_h - \mu_g)^2 + \mathcal{V}(\overline{x}_h) + \mathcal{V}(\overline{x}_g),$$

por lo que un estimador insesgado de la varianza poblacional es

$$(1.12)\quad \hat{\sigma}^{2''}_{st} = \sum_{h=1}^{L} P_h \hat{\sigma}_h^2 + \sum_{h<g}^{L}\sum P_h P_g \left\{(\overline{x}_h - \overline{x}_g)^2 - \left[\hat{\mathcal{V}}(\overline{x}_h) + \hat{\mathcal{V}}(\overline{x}_g)\right]\right\},$$

siendo

$$\hat{\sigma}_h^2 = \begin{cases} s_h^2 & \text{con diseño masr} \\ \dfrac{N_h - 1}{N_h} s_h^2 & \text{con diseño mas} \end{cases}$$

y

$$\hat{\mathcal{V}}(\overline{x}_h) = \begin{cases} \dfrac{s_h^2}{n_h} & \text{con diseño masr} \\ \dfrac{N_h - n_h}{N_h} \dfrac{s_h^2}{n_h} & \text{con diseño mas} \end{cases}$$

siendo s_h^2 la cuasivarianza muestral en el estrato h.

El estimador insesgado (1.12) es el mismo propuesto por Mirás (1985), y reformulado por Hedayat y Sinha (1991), aunque más simple que ambos desde un punto de vista operacional.

2. ESTIMACIÓN DE LA VARIANZA PARA ESTRATEGIAS INTERMEDIAS

2.1. Planteamiento

En un reciente trabajo, Ruiz y Santos (1989) han introducido las llamadas "estrategias intermedias de muestreo" que permiten estimar la media de una población finita con una precisión y un coste esperado intermedio a las estrategias

clásicas masr y mas del mismo tamaño muestral. La técnica propuesta consiste en seleccionar $m(m \geq 2)$ muestras aleatorias simples sin reemplazamiento (independientes) de tamaño $n(n \geq 2)$ cada una, de manera que $mn \leq N$. Si llamamos $\overline{X}_h$ a la media muestral de la $h-$ésima muestra aleatoria simple sin reemplazamiento, $h = 1, 2, \ldots, m$, entonces el estimador insesgado propuesto para la media poblacional (de la población finita de tamaño N), μ, será

$$\hat{\mu} = \frac{1}{m}\sum_{h=1}^{m}\overline{X}_h.$$

Llamando s_h^2 a la cuasivarianza muestral en el diseño $h-$ésimo, $h = 1, 2, \ldots, m$, entonces se propuso como estimador insesgado de la varianza poblacional, σ^2, a

$$\hat{\sigma}_i^2 = \frac{1}{m}\sum_{h=1}^{m}\frac{N-1}{N}s_h^2. \tag{2.1}$$

Con este diseño intermedio la varianza $\mathcal{V}(\hat{\mu})$ se sitúa entre las varianzas clásicas $\mathcal{V}(\overline{X}_{\underline{s}})$ y $\mathcal{V}(\overline{X}_s)$, siendo $\underline{s}$ la muestra ordenada de tamaño nm obtenida por diseño masr, y s es la muestra de tamaño efectivo nm seleccionada por diseño mas. Las expresiones $\overline{X}_{\underline{s}}$ y $\overline{X}_s$ corresponden a las medias muestrales. Además, como se justifica en Ruiz y Santos (1989), el coste esperado de la estrategia intermedia propuesta se sitúa entre los costes esperados de las estrategias clásicas anteriores.

Una ventaja del "diseño intermedio" es que permite estimar la varianza poblacional σ^2, sin modificaciones en cuanto a la recogida de información dada por el diseño propuesto. Una sugerencia hecha en el mismo trabajo es que llamando p al diseño intermedio, su varianza es

$$\mathcal{V}(p, \hat{\mu}) = \frac{N-n}{m(N-1)n}\sigma^2,$$

por lo que al poder estimar σ^2 por $\hat{\sigma}_i^2$ dado en (2.1), es obvio que un estimador insesgado de $\mathcal{V}(p, \hat{\mu})$ será

$$\hat{\mathcal{V}}_1(p, \hat{\mu}) = \frac{N-n}{m(N-1)n}\cdot\frac{1}{m}\sum_{h=1}^{m}\frac{N-1}{N}s_h^2 = \frac{N-n}{m^2Nn}\sum_{h=1}^{m}s_h^2. \tag{2.2}$$

Por otro lado, en las condiciones de selección muestral propuesta vemos que ésta se ajusta a las hipótesis clásicas del método de estimación de la varianza por "grupos aleatorios" en la simbología americana o de "muestreo interpenetrante" en la notación tradicional india. Desde esta perspectiva un estimador insesgado,

diferente del propuesto en (2.2) para $\mathcal{V}(p,\hat{\mu})$ e históricamente anterior, sería (véase Wolter, 1985)

$$\hat{\mathcal{V}}(p,\hat{\mu}) = \frac{1}{m(m-1)}\sum_{h=1}^{m}(\overline{X}_h - \hat{\mu})^2. \tag{2.3}$$

La pregunta que surge de modo natural es qué estimador insesgado de la varianza $\mathcal{V}(p,\hat{\mu})$ es más preciso o deseable. El propósito de las siguientes secciones será dar una respuesta satisfactoria a esta pregunta.

2.2. Varianza de $\hat{\mathcal{V}}_1$

Directamente, de (2.2), debido a la independencia de las m muestras parciales,

$$\mathcal{V}(\hat{\mathcal{V}}_1) = \frac{(N-n)^2}{m^4N^2n^2}\sum_{h=1}^{m}\mathcal{V}(s_h^2), \tag{2.4}$$

y como $\mathcal{V}(s_h^2)$ es constante independientemente del valor que tome $h(h = 1, 2, \ldots, m)$ y su valor viene recogido por Hansen *et al.* (1953), concluímos que asintóticamente (para N muy grande)

$$\mathcal{V}(\hat{\mathcal{V}}_1) \doteq \frac{1}{m^3n^2}\sigma^4\mathcal{V}_{s^2}^2,$$

siendo

$$\mathcal{V}_{s^2}^2 \doteq \frac{1}{n}\left(\frac{\mu_4}{\sigma^4} - \frac{n-3}{n-1}\right) \doteq \frac{1}{n}\frac{\mu_4 - \sigma^4}{\sigma^4}$$

aproximación útil bajo los diseños masr y mas (Hansen *et al.*, 1953), si N es muy grande y $n(< N)$ suficientemente grande. Así resulta que asintóticamente

$$\mathcal{V}(\hat{\mathcal{V}}_1) \doteq \frac{1}{m^3n^3}(\mu_4 - \sigma^4) \tag{2.5}$$

donde hemos despreciado los infinitésimos de órdenes superiores a $m^{-3}n^{-3}$, y siendo μ_4 el momento central de orden 4 para la variable de interés en la población finita,

$$\mu_4 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^4.$$

2.3. Varianza de $\hat{\mathcal{V}}_2$

De la fórmula (2.3) deducimos que

$$(2.6)\quad \mathcal{V}(\hat{\mathcal{V}}_2) = \frac{1}{m^2}\mathcal{V}\left[\frac{1}{m-1}\sum_{h=1}^{m}(\overline{X}_h - \hat{\mu})^2\right] = \frac{1}{m^2}\frac{\mathcal{V}(\overline{X}_h)}{m} = \frac{1}{m^3}\frac{N-n}{N-1}\frac{\sigma^2}{n}$$

que asintóticamente se comporta (cuando N es muy grande),

$$(2.7)\qquad \mathcal{V}(\hat{\mathcal{V}}_2) \doteq \frac{1}{m^3 n}\sigma^2.$$

2.4. Comparaciones

Para comparar $\mathcal{V}(\hat{\mathcal{V}}_1)$ y $\mathcal{V}(\hat{\mathcal{V}})$ tendríamos que hacerlo con sus expresiones exactas (2.4) y (2.6), lo cual resulta ciertamente complicado a la vista de sus desarrollos. No obstante podemos predecir sus comportamientos usando sus expresiones asintóticas (2.5) y (2.7). Efectivamente, para valores de N muy grandes, $\mathcal{V}(\hat{\mathcal{V}}_1)$ es un infinitésimo de orden $m^{-3}n^{-3}$ mientras que $\mathcal{V}(\hat{\mathcal{V}}_2)$ lo es de $m^{-3}n^{-1}$. Esto quiere decir que aumentando el tamaño muestral, n, de cada muestra parcial $h = (1, 2, \ldots, m)$, el estimador $\hat{\mathcal{V}}_1$ tiene una varianza que tiende a cero de modo mucho más rápido que $\mathcal{V}(\hat{\mathcal{V}}_2)$, concretamente converge con un orden de n^{-2} más rápido, lo cual hace de la varianza o dispersión de $\hat{\mathcal{V}}_1$ que sea más pequeña, por lo que el estimador $\hat{\mathcal{V}}_1$ es más deseable y preciso que el clásico $\hat{\mathcal{V}}_2$.

Sin embargo conviene destacar un hecho; para valores de n pequeños, $\mathcal{V}(\hat{\mathcal{V}}_2)$ puede resultar más pequeña que $\mathcal{V}(\hat{\mathcal{V}}_1)$ pues $\mu_4 - \sigma^4$ puede ser superior a σ^2. Por tanto, el tamaño muestral parcial n debe ser suficientemente grande para que el infinitésimo n^{-2} haga de su factor $(\mu_4 - \sigma^4)/\sigma^2$ un valor inferior a 1, y consecuentemente $\mathcal{V}(\hat{\mathcal{V}}_1) < \mathcal{V}(\hat{\mathcal{V}}_2)$ y así el estimador sugerido por Ruiz y Santos (1989) será superior en precisión al clásico estimador de grupos aleatorios o muestreo interpenetrante, para estrategias intermedias.

Finalmente, el incremento de m (número de muestras parciales), para n fijo, no altera asintóticamente las preferencias por uno u otro estimador de la varianza.

3. CONCLUSIONES

Hemos propuesto dos nuevos controles (en el sentido de Ruiz, 1987) para el estimador usual de la media poblacional en muestreo estratificado. Además (1.8), (1.9), (1.10), (1.11) y (1.12) son nuevos estimadores de la varianza poblacional que en determinadas condiciones son todos no negativos, consistentes y conservativos en el sentido de Wolber (1985); además los estimadores recogidos en (1.8) y (1.12) son exactamente insesgados para estimar la varianza poblacional σ^2. Su interés en la práctica es amplio pues son controles o estimadores aplicables en muestreo estratificado convencional o muestreo por conglomerados, a los que sin modificar el diseño muestral es posible añadir estimaciones complementarias aprovechando la información muestral ya obtenida y no siendo necesario otros estudios independientes para su estimación, con sus consecuentes nuevos presupuestos que del modo explicado se economizan.

En la sección segunda, se presenta la superioridad asintótica cuando N es muy grande y $n(< N)$ crece suficientemente, frente al método de grupos aleatorios, del estimador de la varianza para estrategias intermedias propuestas por Ruiz y Santos (1989).

AGRADECIMIENTOS

Expreso mi agradecimiento a los evaluadores por sus constructivos y valiosos consejos que han permitido mejorar la calidad del artículo.

REFERENCIAS

[1] **Cassel, C.M., Särndal, C.E. y Wretman, J.H.** (1977). *Foundations of Inference in Survey Sampling.* Nueva York: Wiley.

[2] **Chaudhuri, A.** (1978). "On estimating the variance of a finite population". *Metrika*, **25**, 65–76.

[3] **Hansen, M.H., Hurwitz, W.N. y Madow, W.G.** (1953). *Sample Survey Methods and Theory.* (Volumen II). Nueva York: Wiley.

[4] **Hedayat, A.S. y Sinha, B.K.** (1991). *Design and Inference in Finite Population Sampling.* Nueva York: Wiley.

[5] **Horvitz, D.G. y Thompson, D.J.** (1952). "A generalisation of sampling without replacement from a finite universe". *J. Amer. Statist. Assoc.*, **47**, 663–685.

[6] **Liu, T.P. y Thompson, M.E.** (1983). "Properties of estimators of quadratic finite populations functions: the batch approach". *Ann. Statist.*, **11**, 275–285.

[7] **Mirás, J.** (1985). *Elementos de Muestreo para Poblaciones Finitas.* Madrid: I.N.E.

[8] **Murthy, M.N.** (1963). "Generalised unbiased estimation in sampling from finite populations". *Sankhyā Ser. B*, **25**, 245–262.

[9] **Ruiz, M.** (1987). "A control in stratified sampling". *Statistics*, **18**, 287–291.

[10] **Ruiz, M. y Ruiz, M.M.** (1992). "Equilibrated strategy for population variance estimation". *Test*, **1**, 79–91.

[11] **Ruiz, M. y Santos, J.** (1989). "Estrategias intermedias de muestreo". *Estadíst. Española*, **31**, nº 121, 227–235.

[12] **Sankaranarayanan, K.** (1980). "A note on the admissibility of some non-negative quadratic estimators". *J. Roy. Statist. Soc. B*, **42**, 387–389.

[13] **Wolter, K.M.** (1985). *Introduction to Variance Estimation.* Nueva York: Springer-Verlag.

ENGLISH SUMMARY:

NEW ESTIMATORS OF THE VARIANCE IN FINITE POPULATIONS

M. Ruiz Espejo

INTRODUCTION

Although the sampling techniques for finite populations are usually applied to estimate the population mean, attention has recently focused on estimating variances, obtaining greater benefit from the information supplied by the sampling.

1. ESTIMATION OF POPULATION VARIANCE

1.1. Planning

Formula (1.1) expresses the finite population variance already given by Ruiz (1987), and in (1.2) a known relation is also given.

1.2. New expression of population variance

From the usual decomposition of the total variability within and among strata given in (1.3), we can deduce the following expression (1.4)

$$\sigma^2 = \sum_{h=1}^{L} P_h \sigma_h^2 + \sum_{h<g}^{L} \sum P_h P_g (\mu_h - \mu_g)^2,$$

where σ^2 is the population variance, σ_h^2 the variance of stratum h, μ_h the mean of stratum h, P_h the relative size of stratum h and L the number of strata.

1.3. Two new expressions of population mean

Formulas (1.5) and (1.6) can be considered as new controls for obtaining the population mean μ, together with the classical formula (1.7).

1.4. Estimation of variance in cluster sampling

1.4.1. An unbiased estimator in cluster sampling without subsampling

From (1.8),

$$\hat{\sigma}^2_{\text{cl}} = \sum_{h \in s_1} P_h \frac{\sigma_h^2}{\pi_h} + \sum \sum_{h<g \in s_1} P_h P_g \frac{(\mu_h - \mu_g)^2}{\pi_{hg}},$$

where s_1 is the sample of primary units, and π_h and $\pi_{hg} (> 0)$ are the inclusion probabilities of the primary units h and h, g respectively in the sample.

1.4.2. A conservative, consistent and non-negative estimator in cluster sampling with subsampling

From (1.9),

$$\hat{\sigma}^2_{\text{cls}} = \sum_{h \in s_1} P_h \frac{\hat{\sigma}_h^2}{\pi_h} + \sum \sum_{h<g \in s_1} P_h P_g \frac{(\hat{\mu}_h - \hat{\mu}_g)^2}{\pi_{hg}},$$

where $\hat{\sigma}_h^2$ and $\hat{\mu}_h$ are consistent estimators of σ_h^2 and μ_h respectively.

1.5. Estimation of variance in stratified sampling

1.5.1. Conservative and consistent estimators in stratified sampling

When $\hat{\sigma}_h^2$ are non-negative ($h = 1, 2, \ldots, L$), and $\hat{\mu}_h$ and $\hat{\sigma}_h^2$ are consistent, the estimator (1.10)

$$\hat{\sigma}^2_{\text{st}} = \sum_{h=1}^{L} P_h \hat{\sigma}_h^2 + \sum \sum_{h<g}^{L} P_h P_g (\hat{\mu}_h - \hat{\mu}_g)^2$$

is consistent for σ^2. Moreover, if $\hat{\mu}_h$ and $\hat{\sigma}_h^2$ are unbiased, then $\hat{\sigma}_{st}^2$ is conservative for σ^2. Another estimator with similar characteristics is (1.11),

$$\hat{\sigma}_{st}^{2'} = \sum_{h=1}^{L} P_h \hat{\sigma}_h^2 + \sum_{h=1}^{L} P_h (\hat{\mu}_h - \hat{\mu}_{st})^2,$$

where $\hat{\mu}_{st}$ is the usual estimator of μ in stratified sampling.

1.5.2. Unbiased estimation in stratified sampling

This estimator is (1.12)

$$\hat{\sigma}_{st}^{2''} = \sum_{h=1}^{L} P_h \hat{\sigma}_h^2 + \sum_{h<g}^{L} \sum P_h P_g \left\{ (\overline{x}_h - \overline{x}_g)^2 - \left[\hat{\mathcal{V}}(\overline{x}_h) + \hat{\mathcal{V}}(\overline{x}_g) \right] \right\},$$

where $\hat{\sigma}_h^2$ and $\hat{\mathcal{V}}(\overline{x}_h)$ are unbiased for σ_h^2 and $\mathcal{V}(\overline{x}_h)$ respectively.

2. VARIANCE ESTIMATION FOR INTERMEDIATE STRATEGIES

2.1. Planning

The theory of intermediate sampling strategies due to Ruiz and Santos (1989) is revised so as to estimate the population mean. We also give two unbiased variance estimators, $\hat{\mathcal{V}}_1$ (2.2) and $\hat{\mathcal{V}}_2$ (2.3).

2.2. Variance of $\hat{\mathcal{V}}_1$

This can be obtained asymptotically (if $1 \ll n \ll N$),

$$\mathcal{V}(\hat{\mathcal{V}}_1) \doteq \frac{1}{m^3 n^3} (\mu_4 - \sigma^4).$$

2.3. Variance of $\hat{\mathcal{V}}_2$

Similarly

$$\mathcal{V}(\hat{\mathcal{V}}_2) \doteq \frac{1}{m^3 n}\sigma^2.$$

2.4. Comparisons

Asymptotically, if $1 \ll n \ll N$, then

$$\mathcal{V}(\hat{\mathcal{V}}_1) < \mathcal{V}(\hat{\mathcal{V}}_2),$$

although the other inequality can appear for small values of n.