ESTIMACION DE LA MEDIA POBLACIONAL EN UN MUESTREO POLIETAPICO

FERNANDO ESQUIVEL BOCANEGRA

T E S I S

PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE:
MAESTRO EN CIENCIAS
EN ESTADISTICA EXPERIMENTAL

Universidad Autonoma Agraria
Antonio Narro

PROGRAMA DE GRADUADOS

Buenavista, Saltillo, Coah.
JUNIO DE 1994

Tesis elaborada bajo la supervisión del comité particular de asesoría y aprobada como requisito parcial, para obtener el grado de

MAESTRO EN CIENCIAS
EN
ESTADISTICA EXPERIMENTAL

COMITE PARTICULAR

Asesor Principal : ______________________
M.C. Víctor Cantú Hernández

Asesor : ______________________
Dr. Rolando Cavazos Cadena

Asesor : ______________________
M.C. Félix de J. Sánchez Pérez

Asesor : ______________________
M.C. Mario Cantú Sifuentes

____
Dr. José Manuel Fernández Brondo
Subdirector de Postgrado

Buenavista, Saltillo Coahuila
Junio de 1994

C O M P E N D I O

Estimación de la Media Poblacional en un Muestreo Polietápico

POR

FERNANDO ESQUIVEL BOCANEGRA

MAESTRIA
ESTADISTICA EXPERIMENTAL
UNIVERSIDAD AUTONOMA AGRARIA ANTONIO NARRO
BUENAVISTA, SALTILLO, COAHUILA JUNIO 1994

M.C. Víctor Cantú Hernández - Asesor -

Palabras clave : *población, muestra, parámetro, estimador, estimador insesgado, valor esperado, varianza , esperanza y varianza condicional.*

En el desarrollo teórico que a continuación se presenta, se aborda el problema de estimar la media poblacional y la varianza del estimador, aplicando para tal efecto, diversos diseños o procedimientos de muestreo. En lo que corresponde al Capítulo Uno, la estimación de la

media poblacional y de la varianza del estimador se basa en un muestreo simple aleatorio, sin reemplazo. En el Capítulo Dos, la estimación de tales parámetros se lleva a cabo haciendo uso del muestreo por conglomerados (monoetápico), en donde los conglomerados (de diferente tamaño) son seleccionados mediante un muestreo simple aleatorio, sin reemplazo. Teniendo como base a los dos diseños anteriores, en el Capítulo Tres, la estimación de los parámetros antes mencionados se lleva a cabo considerando un muestreo por conglomerados en dos etapas (submuestreo), en donde las unidades seleccionadas en cada etapa son extraídas mediante un muestreo simple aleatorio, sin reemplazo. Finalmente, en el Capítulo Cuatro, se proponen estimadores (insesgados) tanto para la media poblacional así como para la varianza del estimador, cuyo desarrollo se basa en un muestreo por conglomerados en tres etapas (muestreo polietápico), en donde las unidades muestrales de la primera etapa son seleccionadas con reemplazo y con diferente probabilidad, mientras que las unidades de la segunda y tercera etapa son extraídas usando un muestreo simple aleatorio, sin reemplazo.

ABSTRACT

Estimation of the Population Mean under a Multi-stage Sampling

BY

FERNANDO ESQUIVEL BOCANEGRA

MASTER OF SCIENCE

EXPERIMENTAL STATISTICS

UNIVERSIDAD AUTONOMA AGRARIA ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA. MEXICO. JUN 1994

M.C. Víctor Cantú Hernández - Advisor -

Key words : *population, sample, parameter, estimator, unbiased estimator, expected value, conditional expectation and variance.*

This work concerns the problem of estimation the population mean as well as the variance of the corresponding estimator under several probabilistic

sampling plans. Thus, in Chapter One such estimators are builded under a simple randomized sampling scheme without replacement and these results are extended in Chapter Two to cluster sampling, where the clusters are of (possibly) different size. Next, in Chapter Three the previous constructions are used to study the estimation problems in the context of two-stage sampling schems these -called sub-sampling case- under the assumption that sampling units are chosen according to a simple randomized procedure without replacement. Finally, in Chapter Four the estimation problems are studied in the case of a three-stage sampling method which, essentially, is determined by the following feature: the sampling units are chosen by means of a simple randomized plan with replacement in the first-stage, and without replacement in the remaining of the sampling process.

INDICE DE CONTENIDO

¿No sabe
hacer traba-
jo de otro
trabajo?

CAPITULO TRES

ESTIMACION Y VARIANZA DEL ESTIMADOR DE LA MEDIA POBLACIONAL EN UN MUESTREO POR CONGLOMERADOS EN DOS ETAPAS

CAPITULO CUATRO

ESTIMACION Y VARIANZA DEL ESTIMADOR DE LA MEDIA POBLACIONAL EN UN MUESTREO POR CONGLOMERADOS EN TRES ETAPAS

INTRODUCCION

El principal interés del diseño de encuestas por muestreo es **maximizar** la cantidad de información acerca de un parámetro poblacional a un costo **mínimo.** El *muestreo aleatorio estratificado* es frecuentemente más adecuado para esto que el *muestreo simple aleatorio*, ya que además de reducir el costo por observación, la estratificación puede producir un límite más pequeño para el error de estimación que el que se generaría por una muestra aleatoria del mismo tamaño. Por otra parte, y bajo ciertas condiciones, el *muestreo por conglomerados* es menos costoso que el muestreo estratificado. Sin embargo, un conglomerado usualmente contiene demasiados elementos para obtener una medición de cada uno de ellos. Cuando esta situación ocurre, el experimentador puede seleccionar una muestra de conglomerados y después tomar una muestra de los elementos dentro de cada conglomerado seleccionado. El procedimiento que consiste en primero seleccionar conglomerados y después escoger un número específico de elementos dentro de cada uno de los conglomerados seleccionados es conocido como *submuestreo*. Esto también se conoce como *muestreo en dos etapas*. El procedimiento puede ser generalizado a tres o más etapas, lo cual se conoce como *muestreo polietápico*.

Las ventajas del muestreo por conglomerados en dos o más etapas sobre otros diseños son, entre las más importantes, las que a continuación se mencionan. Primero, un marco que liste todos los elementos de la población puede ser imposible o costoso de obtener, mientras que obtener una lista de todos los conglomerados puede ser fácil. Segundo, el costo por obtener los datos puede incrementarse por los costos de viaje si los elementos muestreados están muy dispersos sobre una gran área geográfica. Por lo tanto, muestrear conglomerados de elementos que se encuentran juntos físicamente suele ser económico.

El objetivo que se persigue en el presente escrito es, **proponer** estimadores (*insesgados*) tanto para la media poblacional así como para la varianza del estimador, cuyo desarrollo teórico, el cual representa la parte culminante de este trabajo, se basa en un muestreo por conglomerados en tres etapas, es decir, en un muestreo polietápico.

CAPITULO UNO

ESTIMACION Y VARIANZA DEL ESTIMADOR DE LA MEDIA POBLACIONAL EN UN MUESTREO SIMPLE ALEATORIO

Introducción

Una encuesta por muestreo tiene como finalidad hacer una inferencia acerca de la población bajo estudio, en base a la información contenida en una muestra. Existen dos factores que distorsionan la cantidad de información contenida en la muestra y, por lo tanto, afectan la precisión de nuestro procedimiento de hacer inferencias. El primero es el tamaño de la muestra seleccionada de la población. El segundo es la cantidad de variación en los datos; la variación frecuentemente puede ser controlada por el método de selección de la muestra. El procedimiento para seleccionar la muestra se denomina **diseño de la encuesta por muestreo**. Para un tamaño de muestra dado, digamos n, se consideran diversos diseños o procedimientos de muestreo para la obtención de las n observaciones en la muestra. El más simple de los métodos del muestreo probabilístico es conocido como *Muestreo Simple Aleatorio* (M.S.A). Aunque este método no es muy usado en la práctica, se acostumbra empezar con él como parte introductoria en la presentación de la teoría de encuestas por muestreo, ya que el conocimiento de esta técnica es necesario en el estudio de otros y más sofisticados métodos muestrales.

En este capítulo haremos uso del muestreo simple aleatorio para obtener estimadores tanto para la media poblacional así como para la varianza del estimador. Propiedades básicas de estos parámetros pueden encontrarse en **Ash, 1970; Koopmans, 1981** y **Bickel, 1982. Feller, 1975; Kendall, 1977** y **Lehmann, 1983** contienen un tratamiento más avanzado.

Conceptos Básicos

El muestreo simple aleatorio es un método en el cual a cada unidad de la población se le asigna la misma probabilidad de ser seleccionada en la muestra, y en donde cada unidad es extraída *una por una* (con o sin reemplazo) hasta seleccionar n unidades, las cuales pueden ser seleccionadas usando *tablas de números aleatorios*.

DEFINICION. Si N es el número de unidades en la población entonces la probabilidad de seleccionar cualquier unidad en la primera extracción es $1/N$. La probabilidad de seleccionar cualquier unidad disponible en la segunda extracción es $1/(N-1)$, y así sucesivamente. A la muestra de tamaño n obtenida bajo este procedimiento se le llama **muestra aleatoria**, (sin reemplazo).

Una propiedad importante del muestreo simple aleatorio se enuncia en el siguiente teorema.

TEOREMA 1. La probabilidad de que una unidad específica de la población sea seleccionada en cualquier extracción dada es igual a la probabilidad de que sea seleccionada en la la primera extracción, es decir

$$P_{ir} = P_{ir=1} = 1/N \quad , \quad \begin{cases} i = 1,2,\ldots,N \\ r = 1,2,\ldots,n \end{cases}$$

en donde

P_{ir} es la probabilidad de que la i-ésima unidad de la población sea seleccionada en la r-ésima extracción.

Demostración :

Sea n el número de unidades que forman la muestra. La probabilidad de que la unidad específica $\mathcal{U}_i$ sea seleccionada en la r-ésima extracción es, claramente, el producto de:

(1) la probabilidad de que $\mathcal{U}_i$ no sea seleccionada en cualesquiera de las r - 1 extracciones previas, y

(2) la probabilidad de que $\mathcal{U}_i$ sea seleccionada en la r-ésima extracción bajo la suposición de que no fue seleccionada en cualesquiera de las r - 1 extracciones previas.

La probabilidad de que $\mathcal{U}_i$ no sea seleccionada en la primera extracción es, por definición, $(N - 1)/N$; que no sea seleccionada en la segunda extracción, dado que no fue seleccionada en la primera extracción, es $(N - 2)/(N - 1)$, y así sucesivamente. La probabilidad (1) es, por consiguiente,

$$\frac{N-1}{N} \cdot \frac{N-2}{N-1} \cdot \cdot \cdot \frac{N-r+1}{N-r+2} = \frac{N-r+1}{N}$$

La probabilidad (2) es claramente $1/(N - r + 1)$, por lo tanto

$$P_{ir} = \frac{N-r+1}{N} \cdot \frac{1}{N-r+1} = \frac{1}{N} \qquad \blacksquare$$

Lo anterior demuestra que la probabilidad de que una unidad específica de la población sea seleccionada en la r-ésima extracción es independiente de r, en donde $r = 1,2,...,n$, (**Sukhatme, 1970**).

Como una consecuencia inmediata del teorema anterior, se desprende el siguiente corolario.

COROLARIO 1.1. La probabilidad de que una unidad específica de la población esté incluída en la muestra es igual a n/N.

Demostración :

Ya que la unidad específica puede ser incluída en la muestra en cualesquiera de las n extracciones, se sigue que la probabilidad de que esté incluída en la muestra es la suma de las probabilidades de n eventos mutuamente excluyentes, es decir, la unidad específica puede ser incluída en la muestra en la primera extracción, o en la segunda extracción, . . . , o en la n-ésima extracción. Hemos visto que cada uno de estos eventos tiene la misma probabilidad $1/N$. Por lo tanto, la probabilidad de que la unidad específica se encuentre incluída en la muestra es igual a n/N . ■

Notación

Sea

N : el número de unidades en la población.

n : el tamaño muestral, es decir, el número de unidades en la muestra.

Y : la característica de interés.

Y_i : el valor de la característica de interés para la i-ésima unidad en la población, donde i = 1,2,...,N .

y_i : el valor de la característica bajo estudio para la i - ésima unidad incluída en la muestra,donde i=1,2,...,n.

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i \qquad \text{: la media poblacional.} \qquad (1)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \qquad \text{: la media muestral.} \qquad (2)$$

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} Y_i^2 - N\,\overline{Y}^2\right]$$: el cuadrado medio para la población. (3)

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

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\,\overline{y}^2\right]$$: el cuadrado medio muestral.

(4)

$$V(y) = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \overline{Y})^2$$

$$= \frac{N-1}{N}\,S^2$$: la varianza de la población.

(5)

Con el afán de facilitar el desarrollo que a continuación se expone, supondremos que a cada unidad u_i en la población está asociado un valor variable Y_i (i = 1,2,...,N) para la característica de interés y . Además, supongamos que las n unidades en la muestra aleatoria son u_1, u_2,..., u_n (seleccionadas en ese orden), con los valores de la variable y_1, y_2,..., y_n ,respectivamente.

Estimación De La Media Poblacional

Previamente se ha establecido que el objetivo de la encuesta por muestreo es hacer inferencias acerca de una oblación, a partir de la información contenida en una uestra. Una manera de hacer inferencias es estimar ciertos arámetros de la población, utilizando la información de la uestra. El objetivo de una encuesta por muestreo es, recuentemente, estimar una media poblacional. El siguiente eorema contempla la estimación de este parámetro.

TEOREMA 2. En el M.S.A. sin reemplazo , la media muestral

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

es un estimador insesgado de la media poblacional $\bar{Y}$.

---

emostración :

omo lo hemos indicado, $u_1, u_2, \ldots, u_i, \ldots, u_n$, son n nidades en la muestra con los valores de la variable $_1, y_2, \ldots, y_i, \ldots, y_n$, respectivamente.

a variable aleatoria y_i asociada con la i-ésima elección puede tener cualesquiera de los valores Y_i ($i =$ $, 2, \ldots, N$), cada uno de los cuales tiene una probabilidad $/ N$ (conforme a la definición). Luego , aplicando el oncepto de valor esperado, (**Meyer, 1973**), tenemos

$$E (y_i) = \frac{1}{N} \sum^{N} Y_i = \bar{Y} \qquad (6)$$

Dado

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

se tiene que,

$$E(\bar{y}) = E\left[\frac{1}{n} \sum_{i=1}^{n} y_i \right]$$

aplicando la propiedad que enuncia que el valor esperado de la suma es la suma de los valores esperados, (**Mood, 1974**), tenemos

$$E(\bar{y}) = \frac{1}{n} \sum_{i=1}^{n} E(y_i) \tag{7}$$

sustituyedo (6) en (7) concluímos que

$$E(\bar{y}) = \bar{Y} \tag{8}$$

De este modo, la media muestral $\bar{y}$ es un estimador insesgado de la media poblacional $\bar{Y}$. ■

Según **Cochran, 1977** una manera alternativa de obtener el resultado anterior se puede lograr escribiendo la media muestral en la forma:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{N} a_i Y_i \tag{9}$$

donde

$$a_i = \begin{cases} 1 & \text{si } Y_i \text{ es incluída en la muestra} \\ 0 & \text{de otro modo} \end{cases}$$

En la expresión (9) las a_i son variables aleatorias, mientras que las Y_i forman un conjunto de valores constantes. Además, si $P(a_i = 1)$ representa la probabilidad de que el valor Y_i aparezca en la muestra, entonces, por el corolario 1.1 tenemos

$$P(a_i = 1) = n / N$$

Luego, a_i tiene una distribución de Bernoulli con probabilidad $P = n / N$, (**Cavazos, 1991**); con lo cual

$$E(a_i) = P = n / N$$

Finalmente, considerando el valor esperado de la expresión (9) obtenemos

$$E(\bar{y}) = \frac{1}{n} \sum_{i=1}^{N} Y_i \, E(a_i) = \bar{Y} \qquad \blacksquare$$

Varianza Del Estimador De La Media Poblacional

TEOREMA 3. En el M.S.A. sin remplazo, la varianza de la media muestral, $V(\bar{y})$, está dada por

$$V(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{S^2}{n}$$

$$= (1 - f) \frac{S^2}{n}$$

donde

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^2$$

y f es la fracción de muestreo n/N

Para simplificar la demostración de este teorema, es conveniente demostrar primero el siguiente lema.

LEMA 3.1.

$$E(\bar{y}^2) = \bar{Y}^2 + \left(1 - \frac{n}{N}\right)\frac{S^2}{n}$$

Demostración (lema 3.1.) :
Por analogía con la expresión (6), tenemos

$$E(y_i^2) = \frac{1}{N}\sum_{i=1}^{N} Y_i^2 \tag{10}$$

sumando y restando $\bar{Y}^2$ en el lado derecho de esta igualdad y usando la relación (3), podemos escribir (10) en la forma

$$E(y_i^2) = \bar{Y}^2 + \frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \bar{Y}^2$$

$$= \bar{Y}^2 + \frac{N-1}{N}\left\{\frac{1}{N-1}\left[\sum_{i=1}^{N} Y_i^2 - N\bar{Y}^2\right]\right\}$$

$$= \bar{Y}^2 + \left(1 - \frac{1}{N}\right)S^2 \tag{11}$$

Por otra parte, si y_r y y_s representan los valores de las unidades seleccionadas en la r-ésima y s-ésima extracción, suponiendo que sean Y_i y Y_j , respectivamente, entonces por definición

$$E(y_r\, y_s) = \sum_{i \neq j = 1}^{N} Y_i\, Y_j\, P_{ir} \cdot P_{js|i} \tag{12}$$

en donde $P_{js|i}$ denota la probabilidad de seleccionar Y_j en la s-ésima extracción , dado que Y_i es seleccionada en la r-ésima extracción. Ahora, por el teorema 1, se tiene que

$$P_{ir} = \frac{1}{N} \tag{13}$$

y, como una extensión del mismo resultado,

$$P_{js|i} = \frac{1}{N - 1} \tag{14}$$

sustituyendo (13) y (14) en (12) obtenemos

$$E\left(y_r \; y_s \right) = \frac{1}{N(N-1)} \sum_{i \neq j = 1}^{N} Y_i \; Y_j \tag{15}$$

Si en lugar de y_r y y_s , pensamos en y_i y y_j (con valores Y_i y Y_j , respectivamente), como los valores asociados a las unidades u_i y u_j que han sido seleccionadas, podemos escribir (15) en forma alternativa como

$$E\left(y_i \; y_j \right) = \frac{1}{N(N-1)} \sum_{i \neq j = 1}^{N} Y_i \; Y_j \tag{16}$$

Sabiendo que

$$\left(\sum_{i=1}^{N} Y_i \right)^2 = \sum_{i=1}^{N} Y_i^2 + \sum_{i \neq j = 1}^{N} Y_i \; Y_j \tag{17}$$

a partir de la expresión (16) obtenemos

$$E\,(y_i\, y_j) = \frac{1}{N(N-1)}\left[\left(\sum_{i=1}^{N} Y_i\right)^2 - \sum_{i=1}^{N} Y_i^2\right]$$

Restando y sumando $N\,\overline{Y}^2$ y usando (1) dentro del corchete en la igualdad anterior, se tiene que

$$E\,(y_i\, y_j) = \frac{1}{N(N-1)}\left[N^2\overline{Y}^{\,2} - N\overline{Y}^{\,2} + N\overline{Y}^{\,2} - \sum_{i=1}^{N} Y_i^2\right]$$

$$E\,(y_i\, y_j) = \frac{1}{N(N-1)}\left[N\overline{Y}^{\,2}(N-1) - \left(\sum_{i=1}^{N} Y_i^2 - N\overline{Y}^{\,2}\right)\right]$$

$$= \overline{Y}^{\,2} - \frac{S^2}{N} \tag{18}$$

De acuerdo a los resultados obtenidos en (11), (17) y (18), encontramos que

$$E\,(\overline{y}^{\,2}) = E\left[\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^2\right]$$

$$= \frac{1}{n^2}\,E\left[\sum_{i=1}^{n} y_i^2 + \sum_{i\neq j=1}^{n} y_i\, y_j\right]$$

$$= \frac{1}{n^2}\left[E\sum_{i=1}^{n} y_i^2 + E\sum_{i\neq j=1}^{n} y_i\, y_j\right]$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{n} E\,(y_i^2) + \sum_{i\neq j=1}^{n} E\,(y_i\, y_j)\right] \tag{19}$$

sustituyendo (11) y (18) en (19) obtenemos

$$E(\bar{y}^2) = \frac{1}{n^2}\left\{n\left[\bar{Y}^2 + \left(1 - \frac{1}{N}\right)S^2\right] + n(n-1)\left[\bar{Y}^2 - \frac{S}{N}\right]\right\}^2$$

y por lo tanto

$$E(\bar{y}^2) = \bar{Y}^2 + \left(1 - \frac{n}{N}\right)\frac{S^2}{n} \qquad (20)$$

■

Una vez ya demostrado el lema 3.1, estamos en condiciones para llevar a cabo la demostración del teorema 3.

Demostración (teorema 3) :

En **Hoog, 1978** podemos ver que

$$V(\bar{y}) = E\left\{\left[\bar{y} - E(\bar{y})\right]^2\right\}$$

$$= E(\bar{y})^2 - \left[E(\bar{y})\right]^2 \qquad (21)$$

sustituyendo las expresiones (8) y (20) en (21), obtenemos

$$V(\bar{y}) = \left(1 - \frac{n}{N}\right)\frac{S^2}{n} = (1 - f)\frac{S^2}{n} \qquad (22)$$

donde $f = n/N$ es la fracción de la muestra, esto es, la fracción de la población que se incluyó en la muestra. ■

OBSERVACION. *La varianza de* y *en la población es, por definición,*

$$V(y) = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$$

De este modo, la varianza de la media muestral podría escribirse de la siguiente manera:

$$V(\bar{y}) = \left[1 - (n - 1) / (N - 1) \right] \frac{V(y)}{n}$$

Si la fracción de la muestra es muy pequeña, de tal forma que n / N y $(n-1) / (N-1)$ son insignificantes con respecto a la unidad, tenemos

$$V(\bar{y}) = \frac{S^2}{n} \doteq \frac{V(y)}{n}$$

Entonces la varianza de $\bar{y}$ depende solamente del tamaño de la muestra y de la varianza de la población y no del tamaño de la población N, (**Raj, 1980**).

Usualmente, el valor de S^2 no será conocido. Esto nos obliga a buscar un estimador para S^2 y, como consecuencia, se tenga un estimador para $V(\bar{y})$. Lo anterior se presenta en el siguiente teorema.

TEOREMA 4. Un estimador insesgado para $V(\bar{y})$, está dado por

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

donde

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

y f es la fracción de muestreo n/N

Demostración :

Considerando el valor esperado del cuadrado medio muestral s^2 , tendremos

$$E(s^2) = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2\right]$$

$$= E\left[\frac{1}{n-1}\left(\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^{\,2}\right)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(y_i^2) - n\,E(\bar{y}^{\,2})\right]$$

usando los resultados (11) y (20), obtenemos

$$E(s^2) = \frac{1}{n-1}\left[n\bar{Y}^{2} + \left(n - \frac{n}{N}\right)S^2 - n\bar{Y}^{2} - \left(1 - \frac{n}{N}\right)S^2\right]$$

$$= S^2 \qquad (23)$$

Esto demuestra que s^2 es un estimador insesgado de S^2, con lo cual

$$\frac{1}{n}\left(1 - \frac{n}{N}\right)E(s^2) = \frac{1}{n}\left(1 - \frac{n}{N}\right)S^2$$

Luego

$$E\left[\left(1 - \frac{n}{N}\right)\frac{s^2}{n}\right] = \left(1 - \frac{n}{N}\right)\frac{S^2}{n}$$

y por lo tanto

$$\hat{V}(\bar{y}) = \left(1 - \frac{n}{N}\right)\frac{s^2}{n} = (1-f)\frac{s^2}{n} \qquad (24)$$

es un estimador insesgado para $V(\bar{y})$. ■

Como podrá observarse, si $f \to 0$,es decir, cuando $N \to \infty$,la expresión (24) queda en la forma

$$\frac{s^2}{n}$$

entonces, el estimador de la varianza de $\bar{y}$ depende solamente del tamaño de la muestra y del cuadrado medio muestral y no del tamaño de la población *N*, (**Sukhatme**, 1970).

CAPITULO DOS

ESTIMACION Y VARIANZA DEL ESTIMADOR DE LA MEDIA POBLACIONAL EN UN MUESTREO POR CONGLOMERADOS

Introducción

Como ya se ha mencionado, el objetivo del diseño de encuestas por muestreo es obtener una cantidad especificada de información acerca de un parámetro poblacional a un costo mínimo. El **muestreo aleatorio estratificado** es frecuentemente más adecuado para esto que el muestreo simple aleatorio, (**Scheaffer, 1986**), debido a que:

(*i*) La estratificación puede producir un límite más pequeño para el error de estimación que el que se generaría por una muestra simple aleatoria del mismo tamaño. Este resultado es particularmente cierto si las mediciones dentro de los estratos son homogéneas.

(*ii*) El costo por observación en la encuesta puede ser reducido mediante la estratificación de los elementos de la población en grupos convenientes.

(*iii*) Se pueden obtener estimaciones de parámetros poblacionales para subgrupos de la población. Los subgrupos deben ser entonces estratos identificables.

Este capítulo introduce un tercer diseño, *muestreo por conglomerados*, el cual en ocasiones proporciona más información por unidad de costo que cualquier otro de los dos diseños mencionados previamente.

Muestra por Conglomerados

Una vez que los conglomerados han sido especificados se debe conformar un marco que liste todos los conglomerados de la población. Entonces se selecciona una muestra aleatoria de conglomerados de este marco mediante el uso, por ejemplo, de tablas de números aleatorios.

DEFINICION.	Una muestra por conglomerados es una muestra aleatoria en la cual cada unidad de muestreo es una colección , o conglomerado, de elementos.

El muestreo por conglomerados es menos costoso que el muestreo aleatorio estratificado, si el costo por obtener un marco que liste todos los elementos poblacionales es muy alto o si el costo por obtener observaciones se incrementa con la distancia que separa los elementos. En la mayoría de los casos, una lista de elementos de la población frecuentemente no está disponible y el uso de un elemento como unidad de muestreo es, por lo tanto, no factible, (**Sukhatme, 1970**). En tales situaciones, es recomendable aplicar el método de muestreo por conglomerados.

Una condición necesaria para la validez del método antes descrito es que cada unidad de la población bajo estudio debe corresponder a uno y sólo uno de los conglomerados, de tal manera que el número total de unidades muestrales en la lista (*marca*) pueda cubrir todas las unidades de la población sin omitir o duplicar alguna de ellas.

Notación

Considere que, una población se divide en N conglomerados de diferente tamaño, es decir, cada grupo tiene diferente número de elementos, y que una muestra de n conglomerados es extraída de esa población mediante el método de muestro simple aleatorio, sin reempalzo.

Sea

N : el número de conglomerados en la población.

n :el número de conglomerados en la muestra.

M_i :el número de elementos en el i-ésimo conglomerado de la pobación, en donde $i = 1,2,...,N$.

m_i :el número de elmentos del i-ésimo conglomerado incluído en la muestra, en donde $i = 1,2,...,n$.

$M_o = \sum_{i=1}^{N} M_i$: el número total de elementos en la población.

$\overline{M} = M_o / N$: el número promedio de elementos por conglomerado.

y : la característica de intrés.

Y_{ij} : el valor de la característica bajo estudio para el j-ésimo elemento en el i-ésimo conglomerdo de la población, en donde, $i = 1,2,...,N$; $j = 1,2,...,M_i$.

y_{ij} : el valor de la característica de interés en el j-ésimo elemento del i-ésimo conglomerado incluído en la muestra, en donde $i = 1,2,\ldots,n$; $j = 1,2,\ldots,m_i$.

$$\overline{Y}_{i\cdot} = \frac{1}{M_i}\sum_{i=1}^{M_i} Y_{ij}$$: la media por elemento en el i-ésimo conglomerado de la población, para $i = 1,2,\ldots,N$. (1)

$$\overline{y}_{i\cdot} = \frac{1}{m_i}\sum_{i=1}^{m_i} y_{ij}$$: la media por elemento del i-ésimo conglomerado incluído en la muestra, para $i = 1,2,\ldots,n$. (2)

$$\overline{Y}_{\cdot\cdot} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M_i} Y_{ij}}{\sum_{i=1}^{N} M_i} = \frac{\sum_{i=1}^{N} M_i\,\overline{Y}_{i\cdot}}{M_o} = \frac{\sum_{i=1}^{N} M_i\,\overline{Y}_{i\cdot}}{N\,\overline{M}}$$

$$\overline{Y}_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^{N} u_i\,\overline{Y}_{i\cdot}$$:la media por elemento en la población, en donde, $u_i = M_i / \overline{M}$ (3)

Nuevamente, para facilitar la cosas, supondremos que a cada conglomerado $\mathcal{C}_i$ en la población está asociado un valor variable $M_i\,\overline{Y}_{i\cdot}$ ($i = 1,2,\ldots,N$) que representa el **total**, en el i-ésimo conglomerado,para la característica de interés y. También, supongamos que los n conglomerados en la muestra aleatoria son c_1 , c_2 ,...,c_n (seleccionados en ese orden), con los valores de la variable $m_1\overline{y}_{1\cdot}$, $m_2\overline{y}_{2\cdot}$, ...,$m_n\overline{y}_n$, respectivamente.

Estimación De La Media Poblacional

No intervienen nuevos principios en la obtención de estimadores cuando se ha tomado una muestra probabilística de conglomerados y cada conglomerado de la muestra es enumerado completamente (esto es, no hay submuestreo). Debe considerarse el problema del tamaño óptimo del conglomerado. Este dependerá, por supuesto, del costo de obtener información a partir de conglomerados de diferente tamaño y de la varianza que resulte, (**Raj, 1979**). Empezaremos proporcionando el siguiente teorema.

TEOREMA 1. En el muestreo por conglomerados, la media muestral

$$\bar{y} = \frac{1}{n\,\bar{M}} \sum_{i=1}^{n} m_i \, \bar{y}_{i.}$$

es un estimador insesgado de la media poblacional $\bar{Y}_{..}$.

Demostración :

Sean $c_1, c_2, \ldots, c_i, \ldots, c_n$, n conglomerados en la muestra con los valores de la variable $m_1\bar{y}_{1.}$, $m_2\bar{y}_{2.}, \ldots, m_i\bar{y}_{i.}, \ldots,$ $\ldots$, $m_n\,\bar{y}_{n.}$, respectivamente.

La variable aleatoria $m_i\,\bar{y}_{i.}$ asociada con la i-ésima selección puede tener cualesquiera de los valores $M_i\bar{Y}_{i.}$ (i= 1,2,...,N), cada uno de los cuales tiene una probabilidad 1 / N (ya que el procedimiento de selección es mediante un muestreo simple aleatorio), con lo cual

$$E(m_i \overline{y}_{i.}) = \frac{1}{N} \sum_{i=1}^{N} M_i \overline{Y}_{i.} \tag{4}$$

Ahora bien, dado

$$\overline{y} = \frac{1}{n\,\overline{M}} \sum_{i=1}^{n} m_i\, \overline{y}_{i.}$$

se tiene que

$$E(\overline{y}) = \frac{1}{n\,\overline{M}} \sum_{i=1}^{n} E(m_i\, \overline{y}_{i.}) \tag{5}$$

Sustituyendo (4) en (5) concluímos que

$$E(\overline{y}) = \frac{1}{N\,\overline{M}} \sum_{i=1}^{N} M_i\, \overline{Y}_{i.} = \overline{Y}_{..} \tag{6}$$

De este modo, la media muestral $\overline{y}$ es un estimador insesgado de la media poblacional $\overline{Y}_{..}$. ■

Varianza Del Estimador De la Media Poblacional

TEOREMA 2. En un muestreo por conglomerados, la varianza de la media muestral, $V(\bar{y})$, está dada por

$$V(\bar{y}) = \left(1 - \frac{n}{N}\right)\frac{S_b^2}{n}$$

$$= (1 - f)\frac{S_b^2}{n}$$

en donde

$$S_b^{\,2} = \frac{1}{N-1}\sum_{i=1}^{N}(u_i\,\bar{Y}_{i.} - \bar{Y}_{..})^2 \quad ,u_i = M_i / \bar{M}$$

y f es la fracción de muestreo n/N

Para facilitar los cálculos involucrados en la demostración de este teorema, demostraremos antes el siguiente lema.

LEMA 2.1.

$$E(\bar{y}^{\,2}) = \bar{Y}_{..}^{\,2} + \left(1 - \frac{n}{N}\right)\frac{S_b^2}{n}$$

Demostración (lema 2.1.) :

Por analogía, con la expresión (10) del capítulo uno, tenemos

$$E\left[\left(\frac{m_i}{\overline{M}}\,\bar{y}_{i.}\right)^2\right] = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{M_i}{\overline{M}}\;\overline{Y}_{i.}\right)^2 = \frac{1}{N}\sum_{i=1}^{N}(u_i\overline{Y}_{i.})^2 \tag{7}$$

en donde $\bar{y}_{i.}$ representa el valor medio del conglomerado incluído en la muestra en la i-ésima extracción.

Escribiendo (7) en términos de S_b^2 se obtiene

$$E\left[\left(\frac{m_i}{\overline{M}}\,\bar{y}_{i.}\right)^2\right] = \overline{Y}_{..}^{\,2} + \frac{1}{N}\sum_{i=1}^{N}(\,u_i\overline{Y}_{i.})^2 - \;\overline{Y}_{..}^{\,2}$$

$$= \overline{Y}_{..}^{\,2} + \frac{1}{N}\left[\;\sum_{i=1}^{N}(u_i\overline{Y}_{i.})^2 - N\overline{Y}_{..}^{\,2}\right]$$

$$= \overline{Y}_{..}^{\,2} + \frac{1}{N}\left[\sum_{i=1}^{N}(u_i\overline{Y}_{i.})^2 - 2N\overline{Y}_{..}^{\,2} + N\overline{Y}_{..}^{\,2}\right]$$

$$= \overline{Y}^2_{..} + \frac{1}{N}\left[\sum_{i=1}^{N}(u_i\overline{Y}_{i.})^2 - 2\overline{Y}_{..}\sum_{i=1}^{N}u_i\overline{Y}_{i.} + \sum_{i=1}^{N}\overline{Y}^2_{..}\right]$$

$$= \overline{Y}_{..}^{\,2} + \frac{1}{N}\left[\;\sum_{i=1}^{N}(\;u_i\;\overline{Y}_{i.} - \;\overline{Y}_{..})^2\;\right]$$

$$= \overline{Y}_{..}^{\,2} + \frac{N-1}{N}\left[\frac{1}{N-1}\;\sum_{i=1}^{N}(\;u_i\;\overline{Y}_{i.} - \;\overline{Y}_{..})^2\right]$$

$$= \overline{Y}_{..}^{\,2} + \left(1 - \frac{1}{N}\right)S_b^{\,2} \tag{8}$$

Por otra parte, y siguiendo un procedimiento paralelo al que se usó para obtener la expresión (16) del capítulo uno, tenemos que

$$E\left(\frac{m_i}{\overline{M}}\frac{m_j}{\overline{M}}\;\overline{y}_{i.}\;\overline{y}_{j.}\right) = \frac{1}{N(N-1)}\sum_{i\neq j=1}^{N} u_i\;u_j\;\overline{Y}_{i.}\;\overline{Y}_{j.} \tag{9}$$

luego

$$E\left(\frac{m_i}{\overline{M}}\frac{m_j}{\overline{M}}\;\overline{y}_{i.}\overline{y}_{j.}\right) = \frac{1}{N(N-1)}\left[\left(\sum_{i=1}^{N} u_i\;\overline{Y}_{i.}\right)^2 - \sum_{i=1}^{N}(u_i\;\overline{Y}_{i.})^2\right]$$

$$= \frac{1}{N(N-1)}\left[N^2\overline{Y}_{..}^{\,2} - N\overline{Y}_{..}^{\,2} + N\overline{Y}_{..}^{\,2} - \sum_{i=1}^{N}(u_i\overline{Y}_{i.})^2\right]$$

$$= \frac{1}{N(N-1)}\left\{N\overline{Y}_{..}^{\,2}(N-1) - \left[\sum_{i=1}^{N}(u_i\overline{Y}_{i.})^2 - N\overline{Y}_{..}^{\,2}\right]\right\}$$

$$= \overline{Y}_{..}^{\,2} - \frac{S_b^{\,2}}{N} \tag{10}$$

usando (8) y (10) concluímos que

$$E(\overline{y}^{\,2}) = E\left[\left(\frac{1}{n}\sum_{i=1}^{n}\frac{m_i}{\overline{M}}\;\overline{y}_{i.}\right)^2\right]$$

$$= \frac{1}{n^2}\,E\left[\sum_{i=1}^{n}\left(\frac{m_i}{\overline{M}}\;\overline{y}_{i.}\right)^2 + \sum_{i\neq j=1}^{n}\frac{m_i}{\overline{M}}\frac{m_j}{\overline{M}}\;\overline{y}_{i.}\overline{y}_{j.}\right]$$

$$E(\bar{y}^{2}) = \frac{1}{n^2}\left[\sum_{i=1}^{n} E\left(\frac{m_i}{\bar{M}}\bar{y}_{i.}\right)^2 + \sum_{i\neq j=1}^{n} E\left(\frac{m_i}{\bar{M}}\frac{m_j}{\bar{M}}\bar{y}_{i.}\bar{y}_{j.}\right)\right]$$

$$= \frac{1}{n^2}\left\{n\left[\bar{Y}_{..}^{2} + \left(1-\frac{1}{N}\right)S_b^2\right] + n(n-1)\left[\bar{Y}_{..}^{2} - \frac{S_b^2}{N}\right]\right\}$$

$$= \bar{Y}_{..}^{2} + \left(1-\frac{n}{N}\right)\frac{S_b^2}{n} \tag{11}$$

■

Demostrado el lema 2.1, fácilmente se demuestra el teorema 2.

Demostración (teorema 2.) :

Por definición,

$$V(\bar{y}) = E\left\{\left[\bar{y} - E(\bar{y})\right]^2\right\}$$

$$= E(\bar{y}^{2}) - \left[E(\bar{y})\right]^2 \tag{12}$$

sustituyendo las expresiones (6) y (11) en (12) obtenemos

$$V(\bar{y}) = \left(1-\frac{n}{N}\right)\frac{S_b^2}{n} \tag{13}$$

■

Para poblaciones muy grandes ($N \to \infty$), observe que (13) se transforma en

$$V(\bar{y}) = \frac{S_b^2}{n} \tag{14}$$

Nuevamente, el valor de S_b^2 es desconocido. La estimación de S_b^2 nos conduce directamente a un estimador para $V(\bar{y})$.

> TEOREMA 3. Un estimador insesgado para $V(\bar{y})$, está dado por
>
> $$\hat{V}(\bar{y}) = (1 - f)\frac{s_b^2}{n}$$
>
> en donde
>
> $$s_b^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{m_i}{\bar{M}}\bar{y}_{i.} - \bar{y}\right)^2$$
>
> y f es la fracción de muestreo n/N

Demostración :

Considerando el valor esperado de s_b^2, vemos que

$$E(s_b^2) = E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{m_i}{\bar{M}}\bar{y}_{i.} - \bar{y}\right)^2\right]$$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n}\left(\frac{m_i}{\bar{M}}\bar{y}_{i.}\right)^2 - n\,\bar{y}^2\right]$$

$$E(s_b^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n} E\left(\frac{m_i}{\overline{M}}\,\overline{y}_{i\cdot}\right)^2 - n\,E(\overline{y})^2\right]$$

usando (8) y (11) obtenemos

$$(s_b^{\,2}) = \frac{1}{n-1}\left[n\overline{Y}_{\cdot\cdot}^{\,2} + \left(n - \frac{n}{N}\right) S_b^{\,2} - n\overline{Y}_{\cdot\cdot}^{\,2} - \left(1 - \frac{n}{N}\right) S_b^{\,2}\right]$$

$$= S_b^{\,2} \tag{15}$$

sto demuestra que $s_b^{\,2}$ es un estimador insesgado de $S_b^{\,2}$, on lo cual

$$\frac{1}{n}\left(1 - \frac{n}{N}\right) E(s_b^2) = \frac{1}{n}\left(1 - \frac{n}{N}\right) S_b^2$$

uego

$$E\left[\left(1 - \frac{n}{N}\right)\frac{s_b^2}{n}\right] = \left(1 - \frac{n}{N}\right)\frac{S_b^2}{n}$$

por lo tanto

$$\hat{V}(\overline{y}) = \left(1 - \frac{n}{N}\right)\frac{s_b^2}{n} = (1 - f)\frac{s_b^2}{n} \tag{16}$$

s un estimador insesgado para $V(\overline{y})$. ■

De nuevo, si $N \to \infty$, entonces (16) adquiere la forma

$$\hat{V}(\overline{y}) = \frac{s_b^{\,2}}{n} \tag{17}$$

Los conglomerados que forman las unidades de muestreo en la primera etapa se llaman *unidades primarias*, y los elementos o grupos de elementos que se encuentran dentro de los conglomerados y que forman las unidades de muestreo en la segunda etapa se conocen como *unidades secundarias*. El procedimiento puede ser generalizado a tres o más etapas, lo cual se conoce como *muestreo polietápico*, (**Sukhatme, 1970**).

Notación

Suponga que las unidades de la primera etapa son de diferente tamaño, y que un muestreo simple aleatorio (sin reemplazo) es empleado en cada etapa.

Sea

N : el número de conglomerados en la población.

n : el número de conglomerados en la muestra.

M_i : el número de unidades secundarias en la i-ésima unidad primaria; $i = 1,2,\ldots,N$.

m_i : el número de unidades secundarias seleccionadas de la i-ésima unidad primaria incluída en la muestra.

$M_o = \sum_{i=1}^{N} M_i$: el número total de unidades secundarias en la pobalción.

$\overline{M} = M_o / N$: el número promedio de elementos por conglomerado.

$m_o = \sum_{i=1}^{n} m_i$: el número de unidades secundarias en la muestra.

y : la característica de interés.

Y_{ij} : el valor de la característica de interés para la j-ésima unidad secundaria de la i-ésima unidad primaria. $i = 1,2,\ldots,N$; $j = 1,2,\ldots,M_i$.

y_{ij} : el valor de la característica de interés para la j-ésima unidad seleccionada de la i-ésima unidad primaria incluída en la muestra.
$i = 1,2,\ldots,n$; $j = 1,2,\ldots,m_i$.

$\overline{Y}_{i.} = \dfrac{1}{M_i} \sum_{j=1}^{M_i} Y_{ij}$: la media de las observaciones en la i-ésima unidad primaria. (1)

$\overline{y}_{i.} = \dfrac{1}{m_i} \sum_{j=1}^{m_i} y_{ij}$: la media de las m_i unidades secundarias seleccionadas de la i-ésima unidad primaria en la muestra (2)

$$\overline{Y}_{..} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M_i} Y_{ij}}{\sum_{i=1}^{N} M_i} = \frac{\sum_{i=1}^{N} M_i \overline{Y}_{i.}}{M_o} = \frac{\sum_{i=1}^{N} M_i \overline{Y}_{i.}}{N \overline{M}}$$

$= \dfrac{1}{N} \sum_{i=1}^{N} u_i \overline{Y}_{i.}$: la media por elemento en la población, en donde, $u_i = M_i / \overline{M}$ (3)

En el siguiente desarrollo, supondremos una población dividida en N conglomerados de donde una muestra de tamaño n es seleccionada, y que los conglomerados seleccionados tienen respectivamente M_1, M_2, . . . , M_n unidades secundarias, de las cuales se tomaron muestras aleatorias de m_1, m_2, . . . , m_n unidades, (**Raj, 1980**).

Estimación De La Media Poblacional

TEOREMA 1. En un muestreo por conglomerados en dos etapas, la media muestral $\bar{y}$ está dada por

$$\bar{y} = \frac{1}{n\,\bar{M}} \sum_{i=1}^{n} M_i\, \bar{y}_{i.}$$

es un estimador insesgado de la media poblacional $\bar{Y}_{..}$.

Demostración :

Dado

$$\bar{y} = \frac{1}{n\,\bar{M}} \sum_{i=1}^{n} M_i\, \bar{y}_{i.}$$

En **Mood, 1974** podemos encontrar que

$$E(\bar{y}) = E\left(\frac{1}{n\,\bar{M}} \sum_{i=1}^{n} M_i\, \bar{y}_{i.} \right)$$

$$= E\left(\frac{1}{n\,\bar{M}} \sum_{i=1}^{n} M_i\, E(\bar{y}_{i.} \mid i_1, \ldots, i_n) \right) \qquad (4)$$

en donde, la parte que corresponde a $i_1,...,i_n$ significa que la i-ésima unidad primaria es incluída en una muestra de tamaño *n*. Para abreviar, simplemente diremos $E(\bar{y}_{i.} \mid i)$.

Considerando específicamente la i-ésima unidad primaria, y de acuerdo con el teorema 2 (capítulo uno), se deduce que

$$E\,(\,\bar{y}_{i.} \mid i\,) = \bar{Y}_{i.} \qquad (5)$$

sustituyendo (5) en la expresión (4), tenemos

$$E\,(\,\bar{y}\,) = E\left(\frac{1}{n\,\bar{M}} \sum_{i=1}^{n} M_i\,\bar{Y}_{i.}\right)$$

Finalmente, aplicando el concepto de esperanza matemática, se concluye que

$$E\,(\,\bar{y}\,) = \frac{1}{N\,\bar{M}} \sum_{i=1}^{N} M_i\,\bar{Y}_{i.} = \frac{1}{N}\sum_{i=1}^{N} u_i\,\bar{Y}_{i.} = \bar{Y}_{..} \qquad (6)$$

■

Varianza Del Estimador De La Media Poblacional

TEOREMA 2. En el muestreo por conglomerados en dos etapas, la varianza de la media muestral, $V\,(\,\bar{y}\,)$, está dada por

$$V\,(\,\bar{y}\,) = \left(\frac{1}{n} - \frac{1}{N}\right) S_b^{\,2} + \frac{1}{n\,N}\sum_{i=1}^{N} u_i^{\,2}\left(\frac{1}{m_i} - \frac{1}{M_i}\right) S_i^{\,2}$$

en donde

$$S_i^{\,2} = \frac{1}{M_i - 1}\sum_{j=1}^{M_i}(\,Y_{ij} - \bar{Y}_{i.})^2 ,$$

$$S_b^{\,2} = \frac{1}{N-1}\sum_{i=1}^{N}(u_i\bar{Y}_{i.} - \bar{Y}_{..})^{2} , u_i = M_i/\bar{M}$$

Demostración :

Por una propiedad conocida, (**Mood, 1974**), tenemos que

$$V(\bar{y}) = V\Big[E(\bar{y} \mid i)\Big] + E\Big[V(\bar{y} \mid i)\Big]$$

con lo cual,

$$V(\bar{y}) = V\left[\frac{1}{n}\sum_{i=1}^{n} u_i \; E(\bar{y}_{i.} \mid i)\right] + E\left[\frac{1}{n^2}\sum_{i=1}^{n} u_i^{\,2} \; V(\bar{y}_{i.} \mid i)\right]$$

sustituyendo la expresión (5) en el primer término de la relación anterior, obtenemos

$$V(\bar{y}) = V\left[\frac{1}{n}\sum_{i=1}^{n} u_i \; \bar{Y}_{i.}\right] + E\left[\frac{1}{n^2}\sum_{i=1}^{n} u_i^{\,2} \; V(\bar{y}_{i.} \mid i)\right] \tag{7}$$

Ahora, si consideramos exclusivamente la i-ésima unidad primaria, a partir del teorema 3 (capítulo uno), se concluye que

$$V(\bar{y}_{i.} \mid i) = \left(\frac{1}{m_i} - \frac{1}{M_i}\right) S_i^{\,2} \tag{8}$$

en donde

$$S_i^{\,2} = \frac{1}{M_i - 1}\sum_{j=1}^{M_i} (Y_{ij} - \bar{Y}_{i.})^2$$

luego, al sustituir (8) en el segundo término de la expresión (7), se tiene

$$V(\bar{y}) = V\left[\frac{1}{n}\sum_{i=1}^{n} u_i \bar{Y}_{i.}\right] + E\left[\frac{1}{n^2}\sum_{i=1}^{n} u_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_i^2\right]$$

Aplicando el teorema 2 (capítulo dos), y por definición de valor esperado, concluímos que

$$V(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right) S_b^2 + \frac{1}{n\,N}\sum_{i=1}^{N} u_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right) S_i^2 \tag{9}$$

■

TEOREMA 3. Un estimador insesgado para $V(\bar{y})$, está dado por

$$\hat{V}(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right) s_b^2 + \frac{1}{n\,N}\sum_{i=1}^{n} u_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right) s_i^2$$

en donde

$$s_i^2 = \frac{1}{m_i - 1}\sum_{j=1}^{m_i} (y_{ij} - \bar{y}_{i.})^2 ,$$

$$s_b^2 = \frac{1}{n-1}\sum_{i=1}^{n} (u_i \bar{y}_{i.} - \bar{y})^2 , \quad u_i = M_i / \bar{M}$$

Demostración :

Considerando el cuadrado medio s_b^2 definido por

$$s_b^2 = \frac{1}{n-1} \sum_{i=1}^{n} (u_i \bar{y}_{i.} - \bar{y})^2$$

entonces, se puede mostrar que

$$E (s_b^2) = S_b^2 + \frac{1}{N} \sum_{i=1}^{N} u_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i} \right) S_i^2 \qquad (10)$$

También

$$E \left[\frac{1}{n} \sum_{i=1}^{n} u_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i} \right) s_i^2 \right] = \frac{1}{N} \sum_{i=1}^{N} u_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i} \right) S_i^2 \qquad (11)$$

luego, un estimador insesgado de S_b^2 está dado por

$$\hat{S}_b^2 = s_b^2 - \frac{1}{n} \sum_{i=1}^{n} u_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i} \right) s_i^2 \qquad (12)$$

Ya que (11) y (12) proporcionan estimadores insesgados de las dos componentes en (9), concluímos

$$\hat{V} (\bar{y}) = \left(\frac{1}{n} - \frac{1}{N} \right) s_b^2 + \frac{1}{n\ N} \sum_{i=1}^{n} u_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i} \right) s_i^2 \qquad (13)$$

■

Para N suficientemente grande, tenemos

$$\hat{V} (\bar{y}) = \frac{s_b^2}{n} \qquad (14)$$

C A P I T U L O   C U A T R O

Estimacion Y Varianza Del Estimador De La Media Poblacional En Un Muestreo Por Conglomerados En Tres Etapas

Introducción

En la mayoría de las encuestas que se realizan a gran escala, las técnicas muestrales que frecuentemente se aplican están diseñadas sobre un muestreo polietápico, en el que se ven inmiscuidas tres, cuatro o aun cinco etapas, y en donde las unidades que corresponden a las diferentes etapas son, en cuanto al tamaño, desiguales, (**Sukhatme, 1970**). Es, por lo tanto, necesario extender la teoría que se ha desarrollado para un muestreo en dos etapas a un muestreo por conglomerados en tres o más etapas.

En este último capítulo, enfocaremos nuestra atención al desarrollo teórico que involucra la estimación de la media poblacional en un muestreo por conglomerados en *tres etapas*, con la advertencia de que por ningún motivo, debe interpretarse como un tratamiento exhaustivo del tema en cuestión.

Muestra Por Conglomerados En Tres Etapas

DEFINICION.	Una muestra por conglomerados en tres etapas se obtiene seleccionando primero una muestra de conglomerados (unidades primarias), posteriormente, de cada conglomerado seleccionado se extrae una muestra de unidades secundarias, y finalmente, se extrae una muestra de los elementos (unidades terciarias) de cada unidad de la segunda etapa seleccionada.

En la estimación de la media poblacional a partir del muestreo por conglomerados en dos etapas, se usaron probabilidades iguales de selección en cada etapa de muestreo. Cuando las unidades de la primera etapa son muy extensas y varían considerablemente en sus tamaños, este sistema de submuestreo es rara vez tan eficiente como muestrear con probabilidades diferentes,(**Sukhatme, 1970**).

En el desarrollo teórico que enseguida se expone, supondremos que las unidades de la primera etapa son seleccionadas con <u>diferente probabilidad</u> y <u>con reemplazo</u>, mientras que las unidades de la segunda y tercera etapa serán extraídas mediante un muestreo simple aleatorio sin reemplazo.

Notación

Sea

N : el número de unidades primarias en la población.

n : el número de unidades primarias en la muestra.

M_i : el número de unidades secundarias en la i-ésima unidad primaria. $i = 1,2,...,N$.

m_i : el número de unidades secundarias seleccionadas de la i-ésima unidad primaria. $i = 1,2,...,n$.

B_{ij} : el número de unidades terciarias en la j-ésima unidad secundaria de la i-ésima unidad primaria. $i = 1,2,...,N$; $j = 1,2,...,M_i$.

$Q_i = \sum_{j=1}^{M_i} B_{ij}$: el número de unidades terciarias en la i-ésima unidad primaria.

$N\,\overline{Q} = \sum_{i=1}^{N} Q_i$: el número total de unidades terciarias en la población.

b_{ij} : el número de unidades terciarias seleccionadas de la j-ésima unidad secundaria en la i-ésima unidad primaria. $i = 1,2,...,n$; $j = 1,2,...,m_i$.

P_i : la probabilidad de selección asignada a la i-ésima unidad primaria, tal que

$$\sum_{i=1}^{N} P_i = 1$$

y : la característica de interés.

Y_{ijk} : el valor de la característica bajo estudio de la k-ésima unidad terciaria en la j-ésima unidad secundaria de la i-ésima unidad primaria. $i = 1,2,...,N$; $j = 1,2,...,M_i$; $k = 1,2,...,B_{ij}$.

y_{ijk} : el valor de la característica de interés de la k-ésima unidad terciaria seleccionada de la j-ésima unidad secundaria en la i-ésima unidad primaria incluídas en la muestra. $i = 1,2,\ldots,n$; $j = 1,2,\ldots,m_i$, $k = 1,2,\ldots,b_{ij}$.

$\overline{Y}_{ij.} = \dfrac{1}{B_{ij}} \sum_{k=1}^{B_{ij}} Y_{ijk}$: la media de las observaciones en la j-ésima unidad secundaria de la i-ésima unidad primaria. (1)

$\overline{y}_{ij.} = \dfrac{1}{b_{ij}} \sum_{k=1}^{b_{ij}} y_{ijk}$: la media de las b_{ij} unidades terciarias seleccionadas de la j-ésima unidad secundaria en la i-ésima unidad primaria. (2)

$$\overline{Y}_{i..} = \frac{\sum_{j=1}^{M_i} \sum_{k=1}^{B_{ij}} Y_{ijk}}{\sum_{j=1}^{M_i} B_{ij}} = \frac{\sum_{j=1}^{M_i} B_{ij}\, \overline{Y}_{ij.}}{M_i\, \overline{B}_{i.}}$$

$= \dfrac{1}{M_i} \sum_{j=1}^{M_i} v_{ij}\, \overline{Y}_{ij.}$: la media de las observaciones en la i-ésima unidad primaria, en donde

$$v_{ij} = \frac{B_{ij}}{\overline{B}_{i.}} \quad , \quad \overline{B}_{i.} = \frac{1}{M_i} \sum_{j=1}^{M_i} B_{ij} \qquad (3)$$

$\overline{y}_{i..} = \dfrac{1}{m_i} \sum_{j=1}^{m_i} v_{ij}\, \overline{y}_{ij.}$: la media muestral para la i-ésima unidad primaria. (4)

$\overline{Y}_{...} = \sum_{i=1}^{N} P_i\, \overline{Y}_{i..}$: la media poblacional. (5)

Dado que P_i , la probabilidad de selección asignada a la i-ésima unidad primaria, es arbitraria (siempre y cuando $\sum_{i=1}^{N} P_i = 1$), podemos entonces considerar que P_i sea proporcional al tamaño de la i-ésima unidad primaria, es decir

$$P_i = \frac{\sum_{j=1}^{M_i} B_{ij}}{\sum_{i=1}^{N} \sum_{j=1}^{M_i} B_{ij}} = \frac{Q_i}{N\,\overline{Q}} \tag{6}$$

con lo cual

$$\overline{Y}_{...} = \frac{1}{N} \sum_{i=1}^{N} \frac{Q_i}{\overline{Q}}\, \overline{Y}_{i..} = \frac{1}{N} \sum_{i=1}^{N} w_i\, \overline{Y}_{i..} \tag{7}$$

en donde $w_i = Q_i / \overline{Q}$.

Estimación De La Media Poblacional

TEOREMA 1. En un muestreo por conglomerados en tres etapas, la media muestral

$$\overline{y} = \frac{1}{n} \sum_{i=1}^{n} \overline{y}_{i..}$$

es un estimador insesgado de la media poblacional $\overline{Y}_{...}$.

Demostración :

Sea

$$\overline{y} = \frac{1}{n} \sum_{i=1}^{n} \overline{y}_{i..}$$

calculando $E(\overline{y})$, se tiene que

$$E(\bar{y}) = E\left(\frac{1}{n}\sum_{i=1}^{n}\bar{y}_{i..}\right) = E\left[\frac{1}{n}\sum_{i=1}^{n}E(\bar{y}_{i..} \mid i)\right] \quad (8)$$

Si $\bar{y}_{i..}$ es el valor de la media muestral, en la i-ésima unidad primaria, obtenido a partir de un muestreo por conglomerados en dos etapas; entonces, por el teorema 1 (capítulo tres), el valor esperado de $\bar{y}_{i..}$ dado que la i-ésima unidad primaria es incluída en la muestra, está dado por

$$E(\bar{y}_{i..} \mid i) = \bar{Y}_{i..} \quad (9)$$

Sustituyendo (9) en (8) obtenemos

$$E(\bar{y}) = E\left(\frac{1}{n}\sum_{i=1}^{n}\bar{Y}_{i..}\right) = \frac{1}{n}\sum_{i=1}^{n}E(\bar{Y}_{i..})$$

en donde la expectativa de $\bar{Y}_{i..}$ es

$$E(\bar{Y}_{i..}) = \sum_{i=1}^{N}P_i\,\bar{Y}_{i..}\,,$$ donde P_i está dada por (6).

Por lo tanto

$$E(\bar{y}) = \frac{1}{N}\sum_{i=1}^{N}\frac{Q_i}{\bar{Q}}\bar{Y}_{i..} = \frac{1}{N}\sum_{i=1}^{N}w_i\,\bar{Y}_{i..} = \bar{Y}_{...} \quad (10)$$

■

Varianza Del Estimador De La Media Poblacional

TEOREMA 2. En el muestreo por conglomerados en tres etapas, la varianza de la media muestral, $V(\bar{y})$, se define por

$$V(\bar{y}) = \frac{N-1}{N} \cdot \frac{\sigma_b^2}{n} + \frac{1}{nN} \sum_{i=1}^{N} w_i \left(\frac{1}{m_i} - \frac{1}{M_i} \right) S_i'^2$$

$$+ \frac{1}{nN} \sum_{i=1}^{N} \frac{w_i}{m_i M_i} \sum_{j=1}^{M_i} v_{ij}^2 \left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}} \right) S_w^2$$

en donde

$$S_w^2 = \frac{1}{B_{ij} - 1} \sum_{k=1}^{B_{ij}} (Y_{ijk} - \bar{Y}_{ij.})^2$$

$$S_i'^2 = \frac{1}{M_i - 1} \sum_{j=1}^{M_i} (v_{ij} \bar{Y}_{ij.} - \bar{Y}_{i..})^2$$

$$\sigma_b^2 = \frac{1}{N-1} \sum_{i=1}^{N} w_i (\bar{Y}_{i..} - \bar{Y}_{...})^2$$

$$= \frac{1}{N-1} \left[\sum_{i=1}^{N} w_i \bar{Y}_{i..}^2 - N \bar{Y}_{...}^2 \right]$$

Para facilitar la demostración de este teorema, demostraremos antes el siguiente lema.

LEMA 2.1.

$$V \left[\frac{1}{n} \sum_{i=1}^{n} \bar{Y}_{i..} \right] = \frac{N-1}{N} \cdot \frac{\sigma_b^2}{n}$$

Varianza Del Estimador De La Media Poblacional

TEOREMA 2. En el muestreo por conglomerados en tres etapas, la varianza de la media muestral, $V(\bar{y})$, se define por

$$V(\bar{y}) = \frac{N-1}{N} \cdot \frac{\sigma_b^2}{n} + \frac{1}{nN} \sum_{i=1}^{N} w_i \left(\frac{1}{m_i} - \frac{1}{M_i} \right) S_i'^2$$

$$+ \frac{1}{nN} \sum_{i=1}^{N} \frac{w_i}{m_i M_i} \sum_{j=1}^{M_i} v_{ij}^2 \left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}} \right) S_w^2$$

en donde

$$S_w^2 = \frac{1}{B_{ij} - 1} \sum_{k=1}^{B_{ij}} (Y_{ijk} - \bar{Y}_{ij.})^2$$

$$S_i'^2 = \frac{1}{M_i - 1} \sum_{j=1}^{M_i} (v_{ij} \bar{Y}_{ij.} - \bar{Y}_{i..})^2$$

$$\sigma_b^2 = \frac{1}{N-1} \sum_{i=1}^{N} w_i (\bar{Y}_{i..} - \bar{Y}_{...})^2$$

$$= \frac{1}{N-1} \left[\sum_{i=1}^{N} w_i \bar{Y}_{i..}^2 - N \bar{Y}_{...}^2 \right]$$

Para facilitar la demostración de este teorema, demostraremos antes el siguiente lema.

LEMA 2.1.

$$V\left[\frac{1}{n} \sum_{i=1}^{n} \bar{Y}_{i..} \right] = \frac{N-1}{N} \cdot \frac{\sigma_b^2}{n}$$

Demostración (lema 2.1.) :
Tenemos que

$$V\left[\frac{1}{n}\sum_{i=1}^{n}\bar{Y}_{i..}\right] = \frac{1}{n^2}\sum_{i=1}^{n} V\left(\bar{Y}_{i..}\right)$$

$$= \frac{1}{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[E\left(\bar{Y}_{i..} - \bar{Y}_{...}\right)^2\right]\right\}$$

$$= \frac{1}{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\sum_{i=1}^{N} P_i\left(\bar{Y}_{i..} - \bar{Y}_{...}\right)^2\right]\right\}$$

$$= \frac{1}{n\,N}\sum_{i=1}^{N} w_i\left(\bar{Y}_{i..} - \bar{Y}_{...}\right)^2$$

$$= \frac{N-1}{N}\cdot\frac{\sigma_b^2}{n} \qquad (11)$$

■

Enseguida demostraremos el teorema 2.

Demostración :
Dado que

$$V(\bar{y}) = V\left[E(\bar{y} \mid i)\right] + E\left[V(\bar{y} \mid i)\right]$$

$$= V\left[\frac{1}{n}\sum_{i=1}^{n} E\left(\bar{y}_{i..} \mid i\right)\right]$$

$$+ E\left[\frac{1}{n^2}\sum_{i=1}^{n} V\left(\bar{y}_{i..} \mid i\right)\right] \qquad (12)$$

sustituyendo la expresión (9) en el primer término de la relación (12) obtenemos

$$V(\bar{y}) = V\left[\frac{1}{n}\sum_{i=1}^{n}\bar{Y}_{i..}\right] + E\left[\frac{1}{n^2}\sum_{i=1}^{n}V(\bar{y}_{i..}|i)\right] \qquad (13)$$

Nuevamente, si $\bar{y}_{i..}$ es un valor obtenido a partir de un muestreo por conglomerados en dos etapas, luego, por el teorema 2 (capítulo tres) tenemos que

$$V(\bar{y}_{i..}|i) = \left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_i'^2 + \frac{1}{m_i M_i}\sum_{j=1}^{M_i} v_{ij}^2\left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}}\right)S_w^2 \qquad (14)$$

en donde

$$S_w^2 = \frac{1}{B_{ij} - 1}\sum_{k=1}^{B_{ij}}(Y_{ijk} - \bar{Y}_{ij.})^2 ,$$

$$S_i'^2 = \frac{1}{M_i - 1}\sum_{j=1}^{M_i}(v_{ij}\bar{Y}_{ij.} - \bar{Y}_{i..})^2$$

sustituyendo (14) en el segundo término de la expresión (13), se tiene

$$V(\bar{y}) = V\left[\frac{1}{n}\sum_{i=1}^{n}\bar{Y}_{i..}\right] + E\left\{\frac{1}{n^2}\sum_{i=1}^{n}\left[\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_i'^2 + \frac{1}{m_i M_i}\sum_{j=1}^{M_i} v_{ij}^2\left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}}\right)S_w^2\right]\right\} \qquad (15)$$

Desarrollando el valor esperado que aparece en el segundo término de la expresión (15), se sigue que

$$V(\bar{y}) = V\left[\frac{1}{n}\sum_{i=1}^{n}\bar{Y}_{i..}\right] + \frac{1}{nN}\sum_{i=1}^{N} w_i\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S'^2_i$$

$$+ \frac{1}{nN}\sum_{i=1}^{N}\frac{w_i}{m_i\, M_i}\sum_{j=1}^{M_i} v^2_{ij}\left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}}\right)S^2_w$$

Finalmente, y aplicando el lema 2.1 de este capítulo, concluímos que

$$V(\bar{y}) = \frac{N-1}{N}\cdot\frac{\sigma^2_b}{n} + \frac{1}{nN}\sum_{i=1}^{N} w_i\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S'^2_i$$

$$+ \frac{1}{nN}\sum_{i=1}^{N}\frac{w_i}{m_i\, M_i}\sum_{j=1}^{M_i} v^2_{ij}\left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}}\right)S^2_w \qquad (16)$$

■

TEOREMA 3. Un estimador insesgado de $V(\bar{y})$, esta dado por

$$\hat{V}(\bar{y}) = \frac{\sigma'^2_b}{n}$$

en donde

$$\sigma'^2_b = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{y}_{i..} - \bar{y})^2$$

Demostración :

Sea

$$\sigma'^2_b = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{y}_{i..} - \bar{y})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}\bar{y}^2_{i..} - n\,\bar{y}^2\right] \qquad (17)$$

Considerando el valor esperado en (17), tenemos

$$E(\sigma_b'^2) = \frac{1}{n-1}\left[E \sum_{i=1}^{n} \bar{y}_{i..}^2 - n\,E(\bar{y}^2) \right]$$

$$= \frac{1}{n-1}\left[E \sum_{i=1}^{n} E(\bar{y}_{i..}^2 \mid i) - n\,E(\bar{y}^2) \right] \tag{18}$$

Observe que

$$E(\bar{y}_{i..}^2 \mid i) = V(\bar{y}_{i..} \mid i) + \left[E(\bar{y}_{i..} \mid i) \right]^2 \tag{19}$$

así como

$$E(\bar{y}^2) = V(\bar{y}) + [E(\bar{y})]^2 \tag{20}$$

Por las expresiones (9) y (14), la relación (19) queda en la forma

$$E(\bar{y}_{i..}^2 \mid i) = \left(\frac{1}{m_i} - \frac{1}{M_i} \right) S_i'^2 + \frac{1}{m_i M_i} \sum_{j=1}^{M_i} v_{ij}^2 \left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}} \right) S_w^2 + \bar{Y}_{i..}^2 \tag{21}$$

Además, sustituyendo (10) y (16) en (20) obtenemos

$$E(\bar{y}^2) = \frac{N-1}{N} \cdot \frac{\sigma_b^2}{n} + \frac{1}{nN} \sum_{i=1}^{N} w_i \left(\frac{1}{m_i} - \frac{1}{M_i} \right) S_i'^2 + \frac{1}{nN} \sum_{i=1}^{N} \frac{w_i}{m_i M_i} \sum_{j=1}^{M_i} v_{ij}^2 \left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}} \right) S_w^2 + \bar{Y}_{...}^2 \tag{22}$$

A partir de la expresión (21), se tiene que

$$E \sum_{i=1}^{n} E\left(\bar{y}_{i..}^{2} \mid i \right) = n\left[\frac{1}{N} \sum_{i=1}^{N} w_i \left(\frac{1}{m_i} - \frac{1}{M_i} \right) S_i'^{2} \right.$$

$$+ \frac{1}{N} \sum_{i=1}^{N} \frac{w_i}{m_i \ M_i} \sum_{j=1}^{M_i} v_{ij}^{2} \left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}} \right) S_w^{2}$$

$$\left. + \frac{1}{N} \sum_{i=1}^{N} w_i \ \bar{Y}_{i..}^{2} \right] \qquad (23)$$

Sustituyendo (22) y (23) en (18), tenemos

$$E\left(\sigma_b'^{2} \right) = \frac{n}{n-1} \left[\frac{1}{N} \sum_{i=1}^{N} w_i \left(\frac{1}{m_i} - \frac{1}{M_i} \right) S_i'^{2} \right.$$

$$+ \frac{1}{N} \sum_{i=1}^{N} \frac{w_i}{m_i \ M_i} \sum_{j=1}^{M_i} v_{ij}^{2} \left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}} \right) S_w^{2}$$

$$+ \frac{1}{N} \sum_{i=1}^{N} w_i \ \bar{Y}_{i..}^{2} - \frac{N-1}{N} \cdot \frac{\sigma_b^{2}}{n}$$

$$- \frac{1}{n \ N} \sum_{i=1}^{N} w_i \left(\frac{1}{m_i} - \frac{1}{M_i} \right) S_i'^{2}$$

$$\left. - \frac{1}{n \ N} \sum_{i=1}^{N} \frac{w_i}{m_i \ M_i} \sum_{j=1}^{M_i} v_{ij}^{2} \left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}} \right) S_w^{2} - \bar{Y}_{...}^{2} \right]$$

$$E(\sigma_b'^2) = \frac{n}{n-1}\left\{ \frac{N-1}{N}\left[\frac{1}{N-1}\left(\sum_{i=1}^{N} w_i\,\bar{Y}_{i..}^{2} - N\,\bar{Y}_{...}^{2}\right]\right\}\right.$$

$$- \frac{N-1}{N}\cdot\frac{\sigma_b^2}{n} + \frac{1}{N}\sum_{i=1}^{N} w_i\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_i'^2$$

$$- \frac{1}{n\,N}\sum_{i=1}^{N} w_i\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_i'^2$$

$$+ \frac{1}{N}\sum_{i=1}^{N}\frac{w_i}{m_i\,M_i}\sum_{j=1}^{M_i} v_{ij}^2\left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}}\right)S_w^2$$

$$\left. - \frac{1}{n\,N}\sum_{i=1}^{N}\frac{w_i}{m_i\,M_i}\sum_{j=1}^{M_i} v_{ij}^2\left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}}\right)S_w^2\right\}$$

$$= \frac{n}{n-1}\left\{\left(1 - \frac{1}{n}\right)\left[\frac{N-1}{N}\cdot\sigma_b^2\right]\right.$$

$$+ \left(1 - \frac{1}{n}\right)\left[\frac{1}{N}\sum_{i=1}^{N} w_i\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_i'^2\right]$$

$$+ \left(1 - \frac{1}{n}\right)\left[\frac{1}{N}\sum_{i=1}^{N}\frac{w_i}{m_i\,M_i}\sum_{j=1}^{M_i} v_{ij}^2\left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}}\right)S_w^2\right]$$

$$= \frac{N-1}{N}\cdot\sigma_b^2 + \frac{1}{N}\sum_{i=1}^{N} w_i\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_i'^2$$

$$+ \frac{1}{N}\sum_{i=1}^{N}\frac{w_i}{m_i\,M_i}\sum_{j=1}^{M_i} v_{ij}^2\left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}}\right)S_w^2$$

Por lo tanto, un estimador insesgado de V $(\bar{y})$ está dado por

$$\hat{V}(\bar{y}) = \frac{\sigma'^{2}_{b}}{n} \qquad \blacksquare$$

R E S U M E N

Para finalizar este trabajo, se presenta enseguida un cuadro en el que aparecen los estimadores (*insesgados*) tanto para la media poblacional así como para la varianza del estimador, cuyo desarrollo teórico, el cual ha representado la parte medular de este documento, se basa en un muestreo por conglomerados en tres etapas, en donde las unidades de la primera etapa tienen diferente tamaño y son seleccionadas con reemplazo y diferente probabilidad (proporcional al tamaño del conglomerado), mientras que las unidades de la segunda y tercera etapa son extraídas mediante un muestreo simple aleatorio, sin reemplazo.

ESTIMADOR	PARAMETRO
$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_{i..}$	$\bar{Y}_{...}$
$\hat{V}(\bar{y}) = \frac{\sigma'^{2}_{b}}{n}$	$V(\bar{y})$

LITERATURA CITADA

Ash, R.B., 1970. *Basic Probability Theory*, Wiley, New York.

Bickel, P. J. & Doksum K.A., 1982. *Mathematical Statistics*, Holden Day, San Francisco.

Cavazos, R., 1991. *Fundamentos de Estadística - Parte I*, Centro de Investigación y de Estudios Avanzados del I.P.N., México, D.F.

Cochran, W.G., 1977. *Sampling Techniques*, Wiley, New York.

Feller, W., 1975.*Introducción a la Teoría de Probabilidades y sus Aplicaciones*, Vol.I, Limusa, México, D.F.

Hogg, R.V. & Craig A.T., 1978. *Introduction to Mathematical Statistics*, Mc. Millan, New York.

Koopmans, L.H., 1981. *An Introduction to Contemporary Statistics*, Duxbury, Boston.

Kendall, M.G. & Stuart A., 1977. *The Advanced Theory of Statistics*, Vol. I, Hafner, New York.

Lehmann, E.L., 1983. *Theory of Point Estimation*, Wiley, New York.

Meyer, P.L.,1973. *Probabilidad y Aplicaciones Estadísticas*, Fondo Educativo Interamericano, México D.F.

Mood, A.M., Graybill F.A. & Boes D.C.,1974. *Introduction to the Theory of statistics*, Mc Graw Hill, Singapore.

Raj, D., 1979. *La Estructura de las Encuestas por Muestreo*, Fondo de Cultura Económica, México, D.F.

Raj, D., 1980. *Teoría del Muestreo*, Fondo de Cultura Económica, México, D.F.

Scheaffer, R.V., Mendenhall W. & Ott L., 1986. *Elementos de Estadística*, Iberoamericana, México, D.F.

Sukhatme, P. V. & Sukhatme B.V., 1970. *Sampling theory of Surveys with Applications*, Indian Society of Agricultural Statistics, New Delhi, India, and the Iowa State University Press, Ames, Iowa, U.S.A.