Revista Colombiana de Estadística
No. 4 - 1981

EL PROBLEMA DE LA ESTIMACION DE PARAMETROS CUANDO SE TRABAJA CON MARCOS DE MUESTRA SUPERPUESTOS

David Ospina Botero
Profesor Asociado
Universidad Nacional

Abstract

When the domain sizes in a multiple frame survey are not known the research worker can either spend some of the project budget to determine the domain sizes, and thus be able to use one estimator or do not determine the domain sizes and use another estimator. This paper looks at the comparison of these two procedures and make some recommendations.

Resumen

Cuando en una investigación los tamaños de los dominios con marcos múltiples de muestra se desconocen, se puede pensar en invertir parte del presupuesto del proyecto en la determinación de dichos tamaños y utilizar un estimador directo, o no especificarlos y entonces hacer uso de otro estimador. Este artículo presenta algunas recomendaciones relativas a cada uno de estos procedimientos.

Introducción - Conceptos

Dos conceptos que se deben distinguir en la teoría y práctica de poblaciones finitas de muestreo son :

1. Población de unidades
2. Marco para el muestreo de las unidades.

El concepto 1 involucra una clara definición de las unidades que pertenecen a la población. El concepto 2 especifica las facilidades físicas, tales como las listas de unidades que se emplean para tener acceso a las unidades de muestreo individuales. En muchas situaciones es difícil designar un marco de referencia único por alguna razón. En esos casos el investigador tiene que realizar un sondeo basado en una multiplicidad de marcos de muestra para obtener una cobertura total. (Este problema fue atacado primeramente por H.O. Hartley (1962) y más tarde por R. Cochran (1965). Los conceptos y notación usados en este artículo son esencialmente los mismos que ellos utilizaron en sus trabajos). Una vez se ha llevado a cabo el muestreo de un marco puede ser ventajoso separar las unidades muestreadas en grupos, particularmente cuando se desean hacer estimaciones de sus medias y sus totales. Si el número de elementos que pertenecen a cada grupo en la población no se conoce, dichos grupos se denominan DOMINIOS.

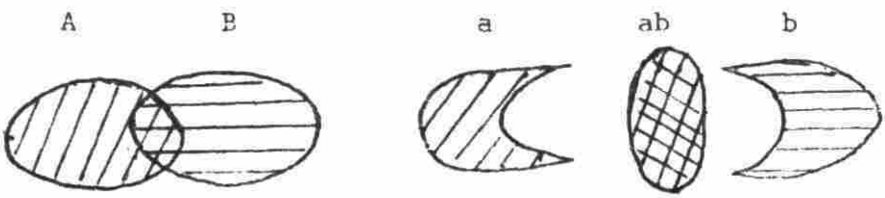
Para fijar la idea, considérense dos marcos A y B de los cuales se han extraído sendas muestras aleatorias. Los diseños de éstas pueden ser diferentes pero deben cumplir los siguientes supuestos :

1. Toda unidad en la población de interés pertenece como mínimo a uno de los marcos.

2. Es posible registrar para cada unidad en la muestra si ella pertenece o no al otro marco.

Con base en lo anterior podemos dividir las unidades de la muestra en 3 dominios.(Para k marcos, tendremos 2^k-1 dominios).

ANTES DE MUESTREAR DESPUES DE MUESTREAR



Dominio "a" - La unidad pertenece únicamente al marco A

Dominio "b" - La unidad pertenece únicamente al marco B

Dominio "ab"- La unidad pertenece a ambos marcos

NOTACION UTILIZADA	MARCOS		DOMINIOS		
	A	B	a	b	ab
Número en la población	N_A	N_B	N_a	N_b	N_{ab}
Número en la muestra	n_A	n_B	n_a	n_b	n'_{ab}, n''_{ab}
Total poblacional	Y_A	Y_B	Y_a	Y_b	Y_{ab}
Media poblacional	$\bar{Y}_A$	$\bar{Y}_B$	$\bar{Y}_a$	$\bar{Y}_b$	$\bar{Y}_{ab}$
Total muestral	y_A	y_B	y_a	y_b	y'_{ab} , y''_{ab}
Media muestral	$\bar{y}_A$	$\bar{y}_B$	$\bar{y}_a$	$\bar{y}_b$	$\bar{y}'_{ab}$, $\bar{y}''_{ab}$

Nótese que n'_{ab} y n''_{ab} denotan las submuestras de n_A y n_B respectivamente, que caen en el dominio superpuesto ab. Las medias correspondientes $\bar{y}'_{ab}$ y $\bar{y}''_{ab}$ pueden calcularse solamente si $n'_{ab} > 0$ y $n''_{ab} > 0$.

I. ESTIMADORES DE TOTALES Y SUS VARIANZAS.

Cuatro casos diferentes pueden considerarse dependiendo del conocimiento del número total de unidades en los marcos y en los dominios y de nuestra habilidad para asignar el número de elementos a seleccionar (tamaños muestrales) de cada dominio.

1. Todos los tamaños de los dominios se conocen. Es factible asignar tamaños de muestra a los dominios.

Dominios = Estratos.

2. Todos los tamaños de los dominios se conocen. Los tamaños de muestra pueden asignarse únicamente a los marcos.

Dominios = Post-estratos.

3. Los tamaños de los dominios no se conocen pero los tamaños de marco sí se conocen. Los tamaños de muestra pueden asignarse solamente a los marcos.

Dominios = Dominios propios.

4. Tanto los tamaños de los dominios como los tamaños de los marcos no se conocen, pero las magnitudes relativas de los marcos se conocen. Los tamaños de muestra sólo pueden asignarse a los marcos.

Dominios = Dominios en poblaciones de tamaño desconocido.

En el caso 1, el problema de estimación se reduce a la metodología estándar para muestreo estratificado. En el caso 4, sólo será posible estimar medias poblacionales. Nos limitaremos aquí a los casos 2 y 3. Hartley (1962) ha sugerido que a los valores de la variable, para estas unidades muestreadas localizadas en más de un marco, les sea asignado un coeficiente ponderado que dependa del marco del cual se muestreó. Además, la suma de estos coeficientes debe ser igual a la unidad. Esta característica es valiosa para obtener estimaciones insesgadas de medias y totales poblacionales. Este método realmente crea estratos no superpuestos de marcos superpuestos. Para dos marcos, sea y_i el valor de la característica y de la i-ésima unidad de muestreo.

Defínase:

$$u_i = \begin{cases} y_i & \text{si la i-ésima unidad está en a} \\ py_i & \text{si la i-ésima unidad está en ab} \end{cases} \tag{1.1}$$

cuando se muestrea del marco A, y

$$u_i = \begin{cases} qy_i & \text{si la i-ésima unidad está en ab} \\ y_i & \text{si la i-ésima unidad está en b} \end{cases} \tag{1.2}$$

cuando se muestrea del marco B.

Claramente Y, el total de los y_i, para la población original de $N = N_A + N_B$ unidades, es igual al total U de las u_i para la muestra población de $N^* = N_a + 2N_{ab} + N_b$ unidades, provisto $p + q = 1$.

$$Y = Y_a + Y_{ab} + Y_b = Y_a + pY_{ab} + qY_{ab} + Y_b = U$$

Este artículo centra toda su atención en aquel caso en que ninguno de los dos marcos cubre completamente la población de interés, pero su unión sí la cubre. Dos situaciones se considerarán. La primera es aquella en la cual el número de unidades en la intersección de los dos marcos, N_{ab}, se conocen, pero es imposible identificar el grupo al cual pertenece cada unidad antes del muestreo. La segunda es aquella en la cual el número de unidades en la intersección, N_{ab}, no se conoce. Hartley (1962) ha propuesto los estimadores y varianzas para estas dos situaciones.

1.1. NUMERO EN LA INTERSECCION CONOCIDO (Muestreo aleatorio simple en ambos marcos).

El estimador post-estratificado del total está dado por :

$$\hat{Y} = N_a\bar{y}_a + N_{ab}(p\bar{y}'_{ab} + q\bar{y}''_{ab}) + N_b\,\bar{y}'_b \qquad (1.3)$$

donde cada término ya se definió anteriormente. W. Cochran (1963) ha mostrado que si la muestra es grande y las varianzas dentro de los post-estratos (dominios) no difieren sensiblemente, entonces la varianza de la estimación post-estratificada para el total, ignorando las correcciones para poblaciones finitas, sería :

$$V[\hat{Y}] = N_a^2\,V[\bar{y}_a] + N_{ab}^2\,p^2 V[\bar{y}_{ab}] + N_b^2\,V[\bar{y}_b] + N_{ab}^2 q^2\,V[\,y''_{ab}\,]$$

$$= \frac{N_a^2\,\sigma_a^2}{n_a} + \frac{N_{ab}^2\,p^2\,\sigma_{ab}^2}{n'_{ab}} + \frac{N_b^2\,\sigma_b^2}{n_b} + \frac{N_{ab}^2\,q^2\,\sigma_{ab}^2}{n''_{ab}}$$

Para muestras grandes : $n_a \simeq n_A \frac{N_a}{N_A}$, $n_b \simeq n_B \frac{N_b}{N_B}$,

$$n'_{ab} \simeq n_A \frac{N_{ab}}{N_A} , \quad n''_{ab} \simeq n_B \frac{N_{ab}}{N_B}$$

Utilizando estas definiciones :

$$V[\hat{Y}] = \frac{N_A}{n_A} \{N_a \sigma_a^2 + N_{ab} p^2 \sigma_{ab}^2\} + \frac{N_B}{n_B} \{N_b \sigma_b^2 + N_{ab} q^2 \sigma_{ab}^2\} \qquad (1.4)$$

Esta fórmula no es otra que la de la varianza del estimador total en muestreo estratificado con afijación óptima, es decir, cuando

$$n_h = \frac{N_h}{N} n$$

donde n_h = tamaño de muestra del estrato h

N_h = tamaño de la población del estrato h.

Haciendo $\alpha = \frac{N_{ab}}{N_A}$, $\beta = \frac{N_{ab}}{N_B}$, obtenemos con la ayuda del álgebra :

$$V[\hat{Y}] = \frac{N_A^2}{n_A} \{(1-\alpha) \sigma_a^2 + \alpha p^2 \sigma_{ab}^2\} + \frac{N_B^2}{n_B} \{(1-\beta) \sigma_b^2 + \beta q^2 \sigma_{ab}^2\} \qquad (1.5)$$

Este resultado fue dado por Hartley.

En algunas ocasiones puede ser ventajoso considerar una función de costo :

$$C = C_A n_A + C_B n_B \qquad (1.6)$$

donde C es el costo total del muestreo, C_A es el costo de obtener una observación del marco A y C_B el de una observación del marco B, es posible encontrar los valores de n_A, n_B y p que minimicen la varianza cuando el costo de muestreo es fijo, o que minimicen el costo cuando el valor de la varianza se determina de antemano. Usando el costo como una condición lateral, la técnica clásica de los multiplicadores de Lagrange puede utilizarse. Después de algún trabajo, el valor óptimo de p es dado por Hartley (1962) como una de las soluciones de la ecuación :

$$\frac{C_A}{C_B} \cdot \frac{p^2}{q^2} = \frac{\sigma_a^2(1-\alpha) + \alpha\, p^2\, \sigma_{ab}^2}{\sigma_b^2(1-\beta) + \beta\, p^2\, \sigma_{ab}^2}$$

Las raíces de esta ecuación pueden hallarse usando un procedimiento exacto como el de Ferrari. Debido a que p y q son ambos positivos y tales que $p+q = 1$, la raíz de esta expresión cuyo valor esté entre cero y uno debe escogerse.

R. Cochram (1965) ha encontrado los valores óptimos para n_A y n_B de :

$$n_A = \theta\, N_A \sigma_{ab} \left\{ \frac{\phi_A(1-\alpha) + \alpha\, p^2}{C_A} \right\}^{1/2}$$

$$n_B = \theta\, N_B \sigma_{ab} \left\{ \frac{\phi_B\,(1-\beta) + \beta q^2}{C_B} \right\}^{1/2}$$

donde

$$\phi_A = \frac{\sigma_a^2}{\sigma_{ab}^2} \quad , \quad \phi_B = \frac{\sigma_b^2}{\sigma_{ab}^2}$$

y

$$\theta = \frac{C}{\sigma_{ab}\left[\sqrt{C_A}\, N_A\{\phi_A(1-\alpha)+\alpha\, p^2\}^{1/2}+\sqrt{C_B}\, N_B\{\phi_B(1-\beta)+\beta q^2\}^{1/2}\right]}$$

R. Cochran también ha trabajado este problema en dos formas diferentes : Primero, tomando una muestra aleatoria simple de tamaño n_A del marco A y observaciones al azar del marco B con un número predeterminado n_b para obtenerse del grupo b; y, segundo, teniendo un tamaño de muestra fijo del marco B más que fijando un número observado en el área no superpuesta del marco B.

1.2 NUMERO EN LOS DOMINIOS DESCONOCIDO.

En este caso el número de unidades en cada marco se conoce pero el número en los dominios no. Hartley (1962) nos da el estimador de Y :

$$\dot{Y} = \frac{N_A}{n_A}\{y_a + py'_{ab}\} + \frac{N_B}{n_B}\{y_b + qy''_{ab}\} \tag{1.7}$$

donde y_a, y_b, y'_{ab} y y''_{ab} son los totales muestrales para los distintos dominios.

Una forma útil de la varianza de $\dot{Y}$ se obtiene aplicando la fórmula estándar para la comparación de muestreo aleatorio con muestreo aleatorio estratificado con afijación proporcional, donde

$$V_{aleatoria} = V_{proporcional} + \sum_{h}^{H} \frac{N_h(\bar{Y}_h - \bar{Y})^2}{N_n} \text{ ,(para la media)}$$

$$= V_{proporcional} + N \sum_{h}^{H} \frac{N_h(\bar{Y}_h - \bar{Y})^2}{n}, \text{ (para el total)}$$

donde $V_{proporcional} = V\left[\hat{Y}\right]$ (La varianza cuando los tamaños de los dominios se conocen).

Se necesita calcular el segundo factor, el incremento cuando no se conocen los tamaños de los dominios. En este caso :

$$N \sum_{h}^{H} \frac{N_h(\bar{Y}_h - \bar{Y})^2}{n} = \frac{N_A N_a}{n_A}\left(\bar{Y}_a - \frac{N_a \bar{Y}_a + p\, N_{ab} \bar{Y}_{ab}}{N_A}\right)^2$$

$$+ \frac{N_A N_{ab}}{n_A}\left(p\, \bar{Y}_{ab} - \frac{N_a \bar{Y}_a + p\, N_{ab} \bar{Y}_{ab}}{N_A}\right)^2$$

$$+ \frac{N_B N_{ab}}{n_B}\left(q\, \bar{Y}_{ab} - \frac{N_b \bar{Y}_b + q\, N_{ab} \bar{Y}_{ab}}{N_B}\right)^2$$

$$+ \frac{N_B N_b}{n_B}\left(\bar{Y}_b - \frac{N_b \bar{Y}_b + q\, N_{ab} \bar{Y}_{ab}}{N_B}\right)^2 \qquad (1.8)$$

ahora $\bar{Y}_a - \dfrac{N_a \bar{Y}_a + p\, N_{ab} Y_{ab}}{} = \dfrac{N_{ab}}{N_A}(\bar{Y}_a - p\, \bar{Y}_{ab})$

En forma similar,

$$p\, \bar{Y}_{ab} - \frac{N_a \bar{Y}_a + p\, N_{ab} \bar{Y}_{ab}}{N_A} = \frac{N_a}{N_A}(p\, \bar{Y}_{ab} - \bar{Y}_a)$$

$$q\, \bar{Y}_{ab} - \frac{N_b \bar{Y}_b + q\, N_{ab} \bar{Y}_{ab}}{N_B} = \frac{N_b}{N_B}(p\, \bar{Y}_{ab} - \bar{Y}_b)$$

$$\bar{Y}_b - \frac{N_b \bar{Y}_b + q\, N_{ab} \bar{Y}_{ab}}{N_B} = \frac{N_{ab}}{N_B}(\bar{Y}_b - q\, \bar{Y}_{ab})$$

Sustituyendo en (1.8) se obtiene :

$$N \sum_{h}^{H} \frac{N_h(\bar{Y}_h - \bar{Y})^2}{n} = \frac{N_A^2}{n_A} \{\alpha(1-\alpha)(\bar{Y}_a - p\bar{Y}_{ab})^2\} + \frac{N_B}{n_B}\{\beta(1-\beta)(\bar{Y}_b - q\bar{Y}_{ab})^2\}$$

Entonces,

$$V[\dot{Y}] = V[\hat{Y}] + \frac{N_A^2}{n_A} \{\alpha(1-\alpha)(\bar{Y}_a - p\bar{Y}_{ab})^2\} + \frac{N_B^2}{n_B} \{\beta(1-\beta)(\bar{Y}_b - q\,\bar{Y}_{ab})^2\}$$

y finalmente:

$$V[\dot{Y}] = \frac{N_A^2}{n_A} \{(1-\alpha)\sigma_a^2 + \alpha\, p^2\, \sigma_{ab}^2\} + \frac{N_B^2}{n_B} \{(1-\beta)\sigma_b^2 + \beta\, q^2\, \sigma_{ab}^2\}$$

$$+ \frac{N_A^2}{n_A} \{\alpha(1-\alpha)(\bar{Y}_a - p\,\bar{Y}_{ab})^2 + \frac{N_B^2}{n_B} \{\beta(1-\beta)\,(\bar{Y}_b - q\,\bar{Y}_{ab})^2\,\}$$

El incremento en la varianza causado por el desconocimiento del tamaño del área supuesta (ab) puede expresarse como :

$$V[\dot{Y}] - V[Y] = \frac{N_a\, N_{ab}}{n_A} (\bar{Y}_a - p\,\bar{Y}_{ab})^2 + \frac{N_b N_{ab}}{n_B} (\bar{Y}_b - q\,\bar{Y}_{ab})^2 \quad (1.10)$$

Se debe notar que esta expresión aumenta cuando el tamaño del área superpuesta amenta . El valor óptimo de p ha sido hallado por R. Cochran (1965).

II. FUNCION DE COSTO ASOCIADA CON EL PROCEDIMIENTO DE MUESTREO - REGLA DE DECISION.

La situación de tamaño de dominio conocidos se presenta raras veces excepto cuando un dominio es un subconjunto de otro. Esto es, un marco contiene todas las unidades en la población, supóngase un marco de área (o mapa), y otro, que puede ser menos costoso de muestrear, por ejemplo una lista de nombres y números de teléfono no contiene todas las unidades. El caso usual es cuando hay algunas unidades en un marco, otras en otro y unas más en ambos y los tamaños de población en estos tres dominios son desconocidos. Esta es la situación que conduce al estimador $\dot{Y}$ discutido anteriormente. Puesto que $\dot{Y}$ obviamente tiene una varianza mayor que el estimador $\hat{Y}$, basado en tamaños de dominio desconocidos, la pregunta que sale a flote se relaciona con la posibilidad de emplear algunos de los recursos existentes para hallar los tamaños de dominio y entonces utilizar $\hat{Y}$. La investigación presente es un intento para establecer alguna guía para la escogencia entre obtener $\dot{Y}$ utilizando todos los recursos, y gastar algunos de estos para determinar los tamaños de dominio y obtener $\hat{Y}$ con una muestra reducida. El incremento en la varianza del estimador del total es el más importante criterio a considerar.

Sea

$$C = C_A n_A + C_B n_B + C_2 \qquad (2.1)$$

donde :

C_A = costo para conseguir información por unidad del marco A.

C_B = costo para conseguir información por unidad del marco B.

C_2 = costo de obtener los tamaños de los dominios.

Cuando los tamaños de los dominios son conocidos :

$$V[\hat{Y}] = \frac{N_A^2}{n_A}\{\sigma_a^2(1-\alpha) + \alpha p^2\sigma_{ab}^2\} + \frac{N_B^2}{n_B}\{\sigma_b^2(1-\beta) + \beta\, q^2\sigma_{ab}^2\}$$

y cuando los tamaños de los dominios son desconocidos :

$$V[\dot{Y}] = \frac{N_A^2}{n_A'}\{\sigma_a^2(1-\alpha) + \alpha p^2\sigma_{ab}^2 + \alpha(1-\alpha)(\bar{Y}_a - p\,\bar{Y}_{ab})^2\}$$

$$+ \frac{N_B^2}{n_B'}\{\sigma_b^2(1-\beta) + \beta\, q^2\sigma_{ab}^2 + \beta(1-\beta)(\bar{Y}_b - q\,\bar{Y}_{ab})^2\}$$

donde $n_A \neq n_A'$ y $n_B \neq n_B'$ ya que

$C = n_A C_A + n_B C_B + C_2$ (cuando los tamaños de dominio son conocidos)

$C = n_A' C_A + n_B' C_B$ (cuando los tamaños de dominio con desconocidos).

Se considerarán diferentes situaciones empezando con la más simple.

1. Asúmase,

$$n_A = n_B\ ,\ n_A' = n_B'$$

$$C_A = C_B = C_1.$$

Reemplazando se obtiene :

$$V[\hat{Y}] = \frac{2C_1}{C-C_2}\{N_A^2[\sigma_a^2(1-\alpha)+\alpha p^2\sigma_{ab}^2] + N_B^2[\sigma_b^2(1-\beta) + \beta q^2\sigma_{ab}^2]\}$$

$$V[\dot{Y}] = \frac{2C_1}{C}\{N_A^2\ [\sigma_a^2(1-\alpha)+\alpha p^2\sigma_{ab}^2+ \alpha(1-\alpha)\ (\bar{Y}_a - p\ \bar{Y}_{ab})^2\ ]$$

$$+ N_B^2\ [\sigma_b^2\ (1-\beta) + \beta q^2\sigma_{ab}^2 + \beta(1-\beta)\ (\bar{Y}_b - q\ \bar{Y}_{ab})^2\ ]\ \}$$

y,

$$V\ [\hat{Y}] - V[\dot{Y}] = \frac{1}{C - C_2}\ [V_n] - \frac{1}{C}\ [V_n + U\ ]$$

donde,

$$V_n = 2C_1\{N_A^2[\sigma_a^2(1-\alpha)+\ \alpha p^2\sigma_{ab}^2] + N_B^2\ [\sigma_b^2(1-\beta)+\beta q^2\sigma_{ab}^2\ ]\ \}\quad (2.2)$$

y,

$$U = 2C_1\{\ N_A^2[\alpha(1-\alpha)(\bar{Y}_a-p\bar{Y}_{ab})^2]+N_B^2\ [\beta(1-\beta)(\ \bar{Y}_b-q\ \bar{Y}_{ab})^2]\}\quad (2.3)$$

2C, puede eliminarse sin que la función de decisión se altere.

Ahora, si $V[\hat{Y}]-V[\dot{Y}] \geq 0$, la varianza del estimador del total es más grande cuando se conocen los tamaños de los dominios que cuando no se conocen. En este caso no se debe gastar C_2 para determinar los tamaños de los dominios.

Un caso interesante ocurre con la situación contraria :

$$V[\dot{Y}] - V[\hat{Y}] > 0$$

que puede escribirse como :

$$\frac{1}{C}[V_n + U] - \frac{1}{C - C_2}[V_n] > 0$$

$$V_n\left[\frac{1}{C} - \frac{1}{C-C_2}\right] + \frac{U}{C} > 0$$

$$\frac{U}{C} + V_n \frac{C-C_2 - C}{C(- C_2)} > 0$$

$$U - \frac{C_2}{C - C_2}[V_n] > 0$$

$$U(C- C_2) - C_2 V_n > 0$$

$$C - C_2\left[\frac{U + V_n}{U}\right] > 0$$

Entonces, si

$$\frac{C_2}{C} < \frac{U}{V_n + U}$$

se deben hallar los tamaños de los dominios a un costo de C_2.

Finalmente,

$$\frac{C_2}{C} > \frac{U}{V_n + U}, \tag{2.4}$$

es la situación donde no vale la pena gastar C_2 para hallar los tamaños de los dominios.

2. Asúmase :

$$N_A = K\, n_B,\ n'_A = K' n'_B \qquad (K,\ K' > 0\ )$$
$$C_A = C_B = C'$$

En este caso la afijación a los marcos es proporcional a los tamaños de ellos.

Para mostrar que $K=K'$, considérese lo siguiente :

$$n_A + n_B = n \qquad\qquad n'_A + n'_B = n'$$

$$n_B\ (K + 1) = n \qquad\qquad n'_B\ (K' + 1) = n'$$

si

$$n_B = \frac{n\ N_B}{N_A + N_B} \quad \text{y} \quad n'_B = \frac{n' N_B}{N_A + N_B}$$

entonces

$$\frac{n\ N_B}{N_A + N_B}\,(K+1) = n \qquad \text{y} \qquad \frac{n'\ N_B}{N_A + N_B}\,(K'+1) = n'$$

$$\frac{N_A + N_B}{N_B} = K + 1 \qquad \text{y} \qquad \frac{N_A + N_B}{N_B} = K' \neq 1$$

de donde $K = K'$.

Sustituyendo por los tamaños de muestra en las expresiones originales, se obtiene :

$$V[\hat{Y}] = \frac{(K+1)C_1}{C - C_1} \{ \frac{N_A^2}{K} [\sigma_a^2(1 -\alpha)+\alpha\ p^2\sigma_{ab}^2]+ N_B^2[\sigma_b^2(1-\beta)+\beta q^2\sigma_{ab}^2]\}$$

$$V[\dot{Y}] = \frac{(K+1)C_1}{C} \{\frac{N_A^2}{K} [\sigma_a^2(1-\alpha)+ \alpha p^2\sigma_{ab}^2+\alpha(1-\alpha)\ (\bar{Y}_a-p\ \bar{Y}_{ab})^2]$$

$$+ N_B^2\ [\sigma_b^2(1-\beta) + \beta q^2\sigma_{ab}^2 + \beta(1-\beta)(\bar{Y}_b - q\ \bar{Y}_{ab})^2]\ \}$$

y,

$$V[\dot{Y}]- V[\hat{Y}] = \frac{1}{C}\ [V_n' + U'\ ] - \frac{1}{C - C_2}\ [V_n']$$

Aplicando la misma técnica que en el primer caso, se obtiene la siguiente regla de decisión :

Si $\frac{C_2}{C} < \frac{U'}{V_n' + U'}$ **, hallar los tamaños de los dominios a un costo de** C_2

$\frac{C_2}{C} > \frac{U'}{V_n' +U'}$, no pagar C_2 para determinar los tamaños de los dominios (2.5)

donde,

$$V_n' = \frac{N_A^2}{K}\ [\sigma_a^2(1-\alpha)+\alpha\ p^2\sigma_{ab}^2]+ N_B^2\ [\sigma_b^2\ (1-\beta)+ \beta q^2\sigma_{ab}^2] \quad (2.6)$$

$$U'= \frac{N_A^2}{K}\ \alpha(1-\alpha)(\bar{Y}_a- p\ \bar{Y}_{ab})^2+ N_B^2\ \beta(1-\beta)(\bar{Y}_b-q\ \bar{Y}_{ab})^2, \quad (2.7)$$

3. Otro caso interesante para considerar es cuando los costos unitarios para obtener la información de los dos marcos no son iguales. Puesto que ambos costos son positivos, ellos pueden expresarse uno como una función del otro.

Para este caso :

$$n_A = K\, n_B \,, \quad n'_A = Kn'_B$$

$$C_A = MC_B \qquad\qquad (K,\ M > 0)$$

con estos supuestos obtenemos :

$$V[\hat{Y}] = \frac{(MK+1)C_B}{C - C_2}\left\{\frac{N_A^2}{K}\left[\sigma_a^2(1-\alpha)+\alpha p^2\sigma_{ab}^2\right]+N_B^2\left[\sigma_b^2(1-\beta)+\beta q^2\sigma_{ab}^2\right]\right\}$$

$$V[\dot{Y}] = \frac{(MK+1)}{C}\, C_B\left\{\frac{N_A^2}{K}\left[\sigma_a^2(1-\alpha)+\alpha p^2\sigma_{ab}^2+\alpha(1-\alpha)(\bar{Y}_a-p\,\bar{Y}_{ab})^2\right]\right.$$

$$\left. + N_B^2\left[\sigma_b^2(1-\beta)+\beta q^2\,\sigma_{ab}^2+\beta(1-\beta)(\bar{Y}_b-q\,\bar{Y}_{ab})^2\right]\right\}$$

Después de hacer las simplificaciones correspondientes se obtiene EXACTAMENTE la misma regla de decisión del segundo caso. Esto permite concluir que la decisión final no está influenciada por la relación entre los costos de muestreo para los dos marcos.

NOTA FINAL : En una entrega posterior se analizarán los problemas anteriores considerando distribuciones y supuestos diferentes.

BIBLIOGRAFIA

Cochran, R.S. Theory and Application of Multiple Frame Surveys, Unpublished Ph.D. dissertation, Department of Statistics, Iowa State University, Ames, Iowa, 1965.

Cochran, W.G. Sampling Techniques, Second Edition, New York, N.Y., John Wiley and Sons, Inc., 1963.

Hartley, H.O. Multiple Frame Surveys. Proceedings of Social Science Section of American Statistical Association Meetings, Minneapolis, Minnesota, 1962.