Revista Colombiana de Estadística
Vols. 21-22, 1990

ESTIMACION MAXIMO VEROSIMIL DEL TOTAL DE UNA CARACTERISTICA EN INVESTIGACIONES DE MARCOS MULTIPLES CUANDO SE DESCONOCEN LOS TAMAÑOS DE LOS DOMINIOS

David Ospina Botero
Profesor Asociado
Universidad Nacional de Colombia

Resumen. El presente artículo desarrolla un procedimiento general para estimar el total de una característica en situaciones de marcos múltiples. La distribución asintótica del estimador es derivada basándose en el comportamiento asintótico del Estimador Máximo Verosímil de los tamaños de los dominios y del método de muestreo. Algunas estimaciones son obtenidas haciendo uso de la Programación Geométrica Subrogada.

Palabras y frases claves: Distribución binomial; distribución normal bivariada; distribución contagiosa (mixta); distribución multinomial; marco muestral traslapado; programación geométrica subrogada; matriz de varianza-covarianza.

1. Introducción.

La estimación de medias y totales para una población definida por la unión de marcos traslapados fue analizada por Hartley (1962). Los estimadores y las varianzas de los estimadores para las situaciones de tamaños de marco conocidos y tamaños de dominios desconocidos fueron también presentados por Hartley (1962). Lund (1968) sugirió una alteración de estos estimadores consistente en basar los pesos asociados en los tamaños de muestra obtenidos. Los estimadores resultantes tienen una eficiencia igual o mayor que los de Hartley. Fuller y Burmeister (1972), usando los estimadores derivados por ellos para el tamaño del dominio superpuesto, desarrollaron estimadores más eficientes que los de Lund (1968) para muestreo sin reemplazamiento. Ambos artículos tienen que ver con la situación de dos marcos. Cochran (1965) basado en algunos de los resultados de Hartley (1962) desarrolló una teoría completa para los diferentes casos relacionados con cobertura incompleta de alguno de los marcos muestrales pero cobertura total por la unión de dos o tres marcos. Una forma general para los estimadores óptimos y las varianzas de los estimadores no se ha determinado. En este artículo se da una aproximación general a la estimación de totales para una población cuando los tamaños de los marcos se conocen pero los tamaños de los dominios son desconocidos. La aproximación dada se basa en el Principio de Máxima Verosimilitud y en la distribución asintótica de los Estimadores Máximo Verosímiles de los tamaños de los dominios obtenidos anteriormente por Ospina (1985)

2. Caso de Dos Marcos.

Hartley (1962) presentó el estimador de la población total ($\dot{Y}$), la varianza del estimador y las fórmulas para los tamaños de muestra cuando se usa muestreo aleatorio simple en cada uno de los dos marcos. El estimador de marco múltiple del total dado por Hartley fué

$$\dot{Y} = \frac{N_1}{n_1}\{y_1+Py'_{12}\}+\frac{N_2}{n_2}\{y_2+Qy''_{12}\} \qquad (2.1)$$

donde:

N_1 = El número de unidades en el marco 1.

N_2 = El número de unidades en el marco 2.

n_1 = Tamaño de muestra del marco 1.

n_2 = Tamaño de muestra del marco 2.

$$y_1 = \sum_{j=1}^{n_{11}} y_{1j}, \quad y'_{12} = \sum_{j=1}^{n'_{12}} y_{12j}$$

$$y_2 = \sum_{j=1}^{n_{22}} y_{2j}, \quad y''_{12} = \sum_{j=1}^{n''_{12}} y_{12j}$$

$$P+Q = 1$$

n'_{12} = Número de unidades en la muestra extraída del marco 1 que pertenece al dominio traslapado.

n''_{12} = Número de unidades en la muestra extraída del marco 2 que pertenece al dominio traslapado.

y_1, y_2, y'_{12} y y''_{12} se denominan totales muestrales.

Ignorando las correcciones para población finita, la varianza de $\dot{Y}$ se da como:

$$V[\dot{Y}] = \frac{N_1^2}{n_1}\,\frac{\sigma_A^2}{n_1} + \frac{N_2^2}{n_2}\,\frac{\sigma_B^2}{n_2} \tag{2.2}$$

donde:

$$\sigma_A^2 = \frac{1}{N_1}\left[\sum_{j=1}^{\theta_1} (y_{1j}-\bar{Y}_A)^2 + \sum_{j=1}^{\theta_{12}} (Py_{12j}-\bar{Y}_A)^2\right]$$

$$\sigma_B^2 = \frac{1}{N_2}\left[\sum_{j=1}^{\theta_2} (y_{2j}-\bar{Y}_B)^2 + \sum_{j=1}^{\theta_{12}} (Qy_{12j}-\bar{Y}_B)^2\right]$$

$\bar{Y}_A$ y $\bar{Y}_B$ son las medias poblacionales de la característica para los marcos 1 y 2 respectivamente. θ_1, θ_2 y θ_{12} ($N_1 = \theta_1+\theta_{12}$, $N_2 = \theta_{12}+\theta_2$) son el número total de unidades en cada uno de los tres dominios. Cochran (1965) dió una forma más útil para la varianza basada en resultados de muestreo estratificado con afijación proporcional

$$V[\dot{Y}] = \frac{N_1^2}{n_1}\left((1-\alpha)\sigma_1^2 + \alpha P^2\sigma_{12}^2 + \alpha(1-\alpha)(\bar{Y}_1 - P\bar{Y}_{12})^2\right)$$
$$+\frac{N_2^2}{n_2}\left((1-\beta)\sigma_2^2 + \beta Q^2\sigma_{12}^2 + \beta(1-\beta)(\bar{Y}_2 - Q\bar{Y}_{12})^2\right) \tag{2.3}$$

donde:

$n = \frac{nN_1}{N}$, $n_2 = \frac{nN_2}{N}$, $\alpha = \frac{\theta_{12}}{N}$, $\beta = \frac{\theta_{12}}{N}$, y $\bar{Y}_1$, $\bar{Y}_2$, $\bar{Y}_{12}$ son las medias de las características para los dominios 1, 2 y el dominio traslapado respectivamente.

En este artículo diferentes estimadores del total y de la varianza del estimador del total son derivados en base a lo siguiente:

a) Distribución de Y: Para considerar el caso general se asume que Y tiene una función de densidad (o de probabilidad) diferente para cada uno de los tres dominios. Asumiendo que las distribuciones compartirán la misma forma pero con parámetros diferentes tenemos

$Y \sim f_1(y)$ si la observación pertenece al dominio no traslapado en el marco 1.

$Y \sim f_2(y)$ si la observación pertenece al dominio no traslapado en el marco 2.

$Y \sim f_{12}(y)$ si la observación pertenece al dominio traslapado.

Se conoce que

$P = Pr$ [Una observación en la muestra pertenece al dominio no traslapado en el marco 1].

$$= \frac{\theta_1}{N}$$

$$= \frac{(N_1 - \theta_{12})}{N_1 + N_2 - \theta_{12}}$$

$P_2 = Pr$ [Una observación en la muestra pertenece al dominio no traslapado en el marco 2].

$$= \frac{\theta_2}{N}$$

$$= \frac{(N_2-\theta_{12})}{N_1+N_2-\theta_{12}}$$

$P_{12} = Pr$ [Una observación en la muestra pertenece al dominio traslapado].

$$= \frac{\theta_{12}}{N}$$

$$= \frac{\theta_{12}}{N_1+N_2-\theta_{12}}$$

donde

$$N = \theta_1 + \theta_2 + \theta_{12}. \tag{2.4}$$

Por tanto, la variable aleatoria Y tiene una distribución contagiosa o mixta (Mood, Graybill y Boes, 1974) dada por

$$f(y) = P_1 f_1(y) + P_2 f_2(y) + P_{12} f_{12}(y) \tag{2.5}$$

(2.5) es una función de densidad (o de probabilidad) definida apropiadamente ya que $P_1+P_2+P_{12} = 1$ y $f_1(y)$, $f_2(y)$ y $f_{12}(y)$ son funciones de densidad (o de probabilidad) por sí mismas.

Como estamos asumiendo los tamaños de los dominios desconocidos y P_1, P_2 y P_{12} dependen de θ_{12}, la distribución de Y depende también de θ_{12}.
Por tanto (2.5) puede reescribirse como:

$$f(y/\theta_{12}) = \frac{(N_1-\theta_{12})}{N_1+N_2-\theta_{12}} f_1(y) + \frac{(N_2-\theta_{12})}{N_1+N_2-\theta_{12}} f_2(y) + \tag{2.6}$$

$$+\frac{\theta_{12}}{N_1+N_2-\theta_{12}}\, f_{12}(y).$$

b) Media de la distribución de Y: La esperanza condicional de dado θ_{12} es:

$$E[Y|\theta_{12}] = \frac{(N_1-\theta_{12})}{N_1+N_2-\theta_{12}}\, E_1(Y) + \frac{(N_2-\theta_{12})}{N_1+N_2-\theta_{12}}\, E_2(Y) + \frac{\theta_{12}}{N_1+N_2-\theta_{12}}\, E_{12}(Y) \qquad (2.7)$$

Por tanto:

$$\begin{aligned} E[Y] &= E\{E[Y|\theta_{12}]\} \\ &= E\left[\frac{(N_1-\theta_{12})}{N_1+N_2-\theta_{12}}\right]E_1(Y) + E\left[\frac{(N_2-\theta_{12})}{N_1+N_2-\theta_{12}}\right]E_2(Y) \\ &+ E\left[\frac{\theta_{12}}{N_1+N_2-\theta_{12}}\right]E_{12}(Y) \end{aligned} \qquad (2.8)$$

Reorganizando (2.8) conduce a

$$\bar{Y} = \bar{Y}_1 E\left[\frac{N_1-\theta_{12}}{N_1+N_2-\theta_{12}}\right] + \bar{Y}_2 E\left[\frac{N_2-\theta_{12}}{N_1+N_2-\theta_{12}}\right] + \bar{Y}_{12} E\left[\frac{\theta_{12}}{N_1+N_2-\theta_{12}}\right], \qquad (2.9)$$

donde $\bar{Y}_1$, $\bar{Y}_2$ y $\bar{Y}_{12}$ son las medias poblacionales dentro de los dominios respectivos.

c) EMV de $\bar{Y}$: Sean θ^*_{12}, $\bar{Y}^*_1$, $\bar{Y}^*_2$, $\bar{Y}^*_{12}$ los EMV de θ_{12}, $\bar{Y}_1$, $\bar{Y}_2$ y $\bar{Y}_{12}$ respectivamente. El EMV de $\bar{Y}$, $\bar{Y}^*$, está dado por:

$$\bar{Y}^* = \frac{(N_1-\theta^*_{12})}{N_1+N_2-\theta^*_{12}}\, \bar{Y}^*_1 + \frac{(N_2-\theta^*_{12})}{N_1+N_2-\theta^*_{12}}\, \bar{Y}^*_2 + \frac{\theta^*_{12}}{N_1+N_2-\theta^*_{12}}\, \bar{Y}^*_{12}, \qquad (2.10)$$

ya que la media muestral es el EMV de la media poblacional para la mayoría de las distribuciones más conocidas. (2.10) puede escribirse como:

$$\bar{y}^* = \frac{(N_1-\theta^*_{12})}{N_1+N_2-\theta^*_{12}}\,\bar{y}_1 + \frac{(N_2-\theta^*_{12})}{N_1+N_2-\theta^*_{12}}\,\bar{y}_2 + \frac{\theta^*_{12}}{N_1+N_2-\theta^*_{12}}\,\bar{y}_{12} \tag{2.11}$$

donde $\bar{y}_1$, $\bar{y}_2$ y $\bar{y}_{12}$ son las medias muestrales correspondientes a los dominios 1, 2 y al dominio traslapado. Bajo el supuesto de una muestra grande las distribuciones de las medias muestrales están dadas por

$$\begin{aligned} \bar{y}_1 &\sim N\left(\bar{Y}_1, \frac{\sigma_1^2}{n_{11}}\right) \\ y_2 &\sim N\left(\bar{Y}_2, \frac{\sigma_2^2}{n_{22}}\right) \\ y_1 &\sim N\left(\bar{Y}_{12}, \frac{\sigma_{12}^2}{n_{12}}\right) \end{aligned} \tag{2.12}$$

d) EMV de Y_{TOT} (Total poblacional): El total de una característica puede definirse como:

$$Y_{TOT} = (N_1-\theta_{12})\bar{Y}_1 + (N_2-\theta_{12})\bar{Y}_2 + \theta_{12}\bar{Y}_{12}, \tag{2.13}$$

Entonces, el Y_{TOT} no restringido puede calcularse como:

$$\begin{aligned} Y_{TOT} &= E[Y_{TOT}|\theta_{12}] \\ &= \bar{Y}_1 E[N_1-\theta_{12}] + \bar{Y}_2 E[N_2-\theta_{12}] + \bar{Y}_{12}E[\theta_{12}], \end{aligned} \tag{2.14}$$

Por tanto, el EMV de Y_{TOT}, Y^*, es:

$$Y^* = (N_1-\theta^*_{12})\bar{y}_1 + (N_2-\theta^*_{12})\bar{y}_2 + \theta^*_{12}\bar{y}_{12}, \tag{2.15}$$

Como las muestras tomadas de los dominios son aleatorias, las medias $\bar{y}_1$, $\bar{y}_2$ y $\bar{y}_{12}$ son independientes. Esto implica que dado el tamaño de muestra n_{11} la distribución de $\bar{y}_1$(1.12) no depende de y_2 o y_{12}. La varianza de Y^* puede expresarse como:

$$
\begin{aligned}
V[Y^*] &= V[(N_1-\theta^*_{12})\bar{y}_1] + V[(N_2-\theta^*_{12})\bar{y}_2] \\
&+ V[\theta^*_{12}\bar{y}_{12}] + 2\mathrm{COV}[(N_1-\theta^*_{12})\bar{y}_1,(N_2-\theta^*_{12})\bar{y}_2] \qquad (2.16) \\
&+ 2\mathrm{COV}[(N_1-\theta^*_{12})\bar{y}_1,\theta^*_{12}\bar{y}_{12}] + 2\mathrm{COV}[(N_2-\theta^*_{12})\bar{y}_2,\theta^*_{12}\bar{y}_{12}],
\end{aligned}
$$

Antes de calcular cada uno de los términos de la derecha en (2.16 se mostrará en el siguiente teorema que bajo el supuesto de muestras grandes θ^*_{12} y $\bar{y}_1$ pueden asumirse independientes.

Teorema 2.1. *Si $\bar{y}_1 \sim N(\bar{Y}_1,\sigma_1^2/n_{11})$ entonces θ^*_{12} y $\bar{y}_1$ son estocástimente independientes.*

Demostración. La función de probabilidad de n_{11}, n_{22} es

$$
f(n_{11},n_{22};\theta_{12}) = \binom{n}{n_{11},n_{22}} \left[\frac{N_1(N_1-\theta_{12})}{nN_1}\right]^{n_{11}} \left[\frac{N_2(N_2-\theta_{12})}{nN_2}\right]^{n_{22}} \cdot \left[\frac{(n_1N_2+n_2N_1)\theta_{12}}{nN_1N_2}\right]^{n-n_{11}-n_{22}}, \qquad (2.17)
$$

donde

$$
n = n_1 + n_2, \quad n_{12} = n-n_{11}-n_{22}.
$$

También el Teorema de Lindberg-Levy presentado por Rao (1973), implica que la función de densidad de $\bar{y}_1$ puede aproximarse por una variable aleatoria normal con media $\bar{Y}_1$ y varianza σ_1^2/n_{11} para n_{11} suficientemente grande. Por tanto:

$$
\begin{aligned}
f(\bar{y}_1,\bar{Y}_1,\sigma_1^2/n_{11},n_{22}) &= f(\bar{y}_1,\bar{Y}_1,\sigma_1^2/n_{11}) \\
&= \frac{1}{\sqrt{2\pi}\frac{\sigma_1}{\sqrt{n_{11}}}}\, e^{(-(\bar{y}_1-\bar{Y}_1)^2)/(2\sigma_1^2/n_{11})} \qquad (2.18) \\
&\quad -\infty < \bar{y}_1 < \infty.
\end{aligned}
$$

La función conjunta mixta (discreta y continua) de n_{11}, n_{22} y $\bar{y}_1$ es

$$\begin{aligned}
&f(n_{11},n_{22},y_1;\theta_{12},\bar{Y}_1,\sigma_1^2)\\
&= \binom{n}{n_{11},n_{22}}\left[\frac{N_1(N_1-\theta_{12})}{nN_1}\right]^{n_{11}}\left[\frac{N_2(N_2-\theta_{12})}{nN_2}\right]^{n_{22}} \qquad (2.19)\\
&\cdot\left[\frac{(n_1N_1+n_2N_1)\theta_{12}}{n_2\ N_2}\right]^{n-n_{11}-n_{22}} \frac{1}{\sqrt{2\pi}\frac{\sigma_1}{\sqrt{n_{11}}}}\, e^{(-(\bar{y}_1-\bar{Y}_1)^2)/(2\sigma_1^2/n_{11})}
\end{aligned}$$

$$-\infty < \bar{y}_1 < \infty$$

(2.19) puede ser considerada como la función de verosimilitud conjunta para $\bar{Y}_1$ y θ_{12} (tomando σ_1^2 como una constante conocida) (Mood, Graybill y Boes, 1974).

$$L(\theta_{12},\bar{Y}_1) = f(n_{11},n_{22},\bar{y}_1;\theta_{12},\bar{Y}_1,\sigma_1^2) \qquad (2.20)$$

La distribución conjunta de θ^*_{12}(EMV de θ_{12}) y $\bar{y}_1$(EMV de $\bar{Y}_1$) se distribuye asintóticamente como una distribución normal bivariada (Mood, Graybill y Boes, 1974) con parámetros

$$E[\theta^*_{12}] = \theta_{12}, \qquad (2.21)$$

$$E[\bar{y}_1] = \bar{Y}_1 \qquad (2.22)$$

$$V[\theta^*_{12}] = \frac{(N_1-\theta_{12})(N_2-\theta_{12})\theta_{12}}{n_1(N_2-\theta_{12})+n_2(N_1-\theta_{12})}$$

$$V[\bar{y}_1] = \sigma^2_{11} \qquad (2.24)$$

y,

$$\mathrm{cov}[\theta^*_{12},\ \bar{y}_1] = 0 \qquad (2.25)$$

De (2.25) se concluye que θ^*_{12} y $\bar{y}_1$ son independientes (Hogg y Craig, 1978).

Demostraciones similares prueban la independencia de θ_{12}^* y $\bar{y}_2$ y θ_{12}^* y $\bar{y}_{12}$. Basado en la anterior demostración y algunos teoremas presentados por Mood, Graybill y Boes (1974), el lado derecho de (2.16) llega a ser:

$$V[(N_1-\theta_{12}^*)\bar{y}_1] = (\bar{Y}_1)^2V[\theta_{12}^*] + (N_1-\theta_{12})^2V[\bar{y}_1] + V[\theta_{12}^*]V[\bar{y}_1]$$

$$= (N_1-\theta_{12})^2 \frac{\sigma_1^2}{n_{11}} + V[\theta_{12}^*]\left[(\bar{Y}_1)^2 + \frac{\sigma_1^2}{n_{11}}\right] \qquad (2.26)$$

Similarmente:

$$V[(N_2-\theta_{12}^*)\bar{y}_2] = (N_2-\theta_{12})^2 \frac{\sigma_2^2}{n_{22}} + V[\theta_{12}^*]\left[(\bar{Y}_2)^2 + \frac{\sigma_2^2}{n_{22}}\right] \qquad (2.27)$$

y

$$V[\theta_{12}^*\bar{y}_{12}] = (\theta_{12})^2 \frac{\sigma_{12}^2}{n_{12}} + V[\theta_{12}^*]\left[(\bar{Y}_{12})^2 + \frac{\sigma_{12}^2}{n_{12}}\right] \qquad (2.28)$$

Las covarianzas son derivadas como:

$$\text{cov}[N_1-\theta_{12}^*)\bar{y}_1,(N_2-\theta_{12}^*)\bar{y}_2] = E[(N_1-\theta_{12}^*)\bar{y}_1(N_2-\theta_{12}^*)\bar{y}_2]$$

$$- \{E[(N_1-\theta_{12}^*)\bar{y}_1]E[(N_2-\theta_{12}^*)\bar{y}_2]\}$$

Por tanto:

$$\text{cov}[N_1-\theta_{12}^*)\bar{y}_1,(N_2-\theta_{12}^*)\bar{y}_2] = \bar{Y}_1\bar{Y}_2\{E[(\theta_{12}^*)^2]-[E(\theta_{12}^*)]^2\}$$

$$= \bar{Y}_1\bar{Y}_2V[\theta_{12}^*] \qquad (2.29)$$

Igualmente:

$$\text{cov}[(N_1-\theta_{12}^*)\bar{y}_1,\theta_{12}^*\bar{y}_{12}] = -\bar{Y}_1\bar{Y}_{12}V[\theta_{12}^*] \qquad (2.30)$$

$$\text{cov}[(N_2-\theta_{12}^*)\bar{y}_2,\theta_{12}^*\bar{y}_{12}] = -\bar{Y}_2\bar{Y}_{12}V[\theta_{12}^*] \qquad (2.31)$$

Combinando estos resultados, la forma final de (2.16) se convierte en:

$$V[y^*] = (N_1-\theta_{12})^2\,\frac{\sigma_1^2}{n_{11}} + (N_2-\theta_{12})^2\,\frac{\sigma_2^2}{n_{22}} + (\theta_{12})^2\,\frac{\sigma_{12}^2}{n_{12}}$$

$$+\,\frac{(N_1-\theta_{12})(N_2-\theta_{12})\theta_{12}}{n_1(N_2-\theta_{12})+n_2(N_1-\theta_{12})}\cdot\left[(\bar{Y}_1+\bar{Y}_2-\bar{Y}_{12})^2+\frac{\sigma_1^2}{n_{11}}+\frac{\sigma_2^2}{n_{22}}+\frac{\sigma_{12}^2}{n_{12}}\right] \tag{2.32}$$

Para muestras grandes e ignorando correcciones para población finita:

$$V[y^*] \approx \frac{N_1}{n_1}(N\;-\theta_{12})\sigma_1^2 + \frac{N_2}{n_2}(N_2-\theta_{12})\sigma_2^2 + \frac{N_1N_2\theta_{12}}{n_1N_2+n_2N_1}\sigma_{12}^2$$

$$+\,\frac{(N_1-\theta_{12})(N_2-\theta_{12})\theta_{12}}{n_1(N_2-\theta_{12})+n_2(N_1-\theta_{12})}\left[(\bar{Y}_1+\bar{Y}_2-\bar{Y}_{12})^2+\frac{N_1\sigma_1^2}{n_1(N_1-\theta_{12})}\right.$$

$$\left.+\,\frac{N_2\sigma_2^2}{n_2(N_2-\theta_{12})}+\frac{N_1N_2\sigma_{12}^2}{(n_1N_2+n_2N_1)\theta_{12}}\right] \tag{2.33}$$

El EMV de $V|y^*|$ está dado por

$$V^*[y^*] = (N_1-\theta_{12}^*)^2\,\frac{S_1^2}{n_{11}} + (N_2-\theta_{12}^*)^2\,\frac{S_2^2}{n_{22}} + (\theta_{12}^*)^2\,\frac{S_{12}^2}{n_{12}}$$

$$+\,\frac{(N_1-\theta_{12}^*)(N_2-\theta_{12}^*)\theta_{12}^*}{n_1(N_2-\theta_{12}^*)+n_2(N_1-\theta_{12}^*)}\left[(\bar{y}_1+\bar{y}_2-\bar{y}_{12})+\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}+\frac{S_{12}^2}{n_{12}}\right] \tag{2.34}$$

3. Caso General: M Marcos.

Con M marcos el número máximo de dominios es 2^M-1. Para facilitar la derivación de las fórmulas, denomínense los tamaños de los dominios $\theta_1,\theta_2,\ldots,\theta_T$, donde $T = 2^M-1$. El estimador propuesto del total es

$$y^* = \theta_1^*\bar{y}_1 + \theta_2^*\bar{y}_2+\ldots+\theta_T^*\bar{y}_T \tag{3.1}$$

con varianza

$$V[Y^*] = \sum_{i=1}^{T} V[\theta_i^* \bar{y}_i] + 2\sum_{i<j}^{T} \operatorname{cov}[\theta_i^* \bar{y}_i, \theta_j^* \bar{y}_j] \tag{3.2}$$

Ahora, usando la teoría presentada por Mood, Graybill y Boes (1974):

$$V[\theta_i^* \bar{y}_i] = V[\theta_i^*]V[\bar{y}_i] + V[\theta_i]\bar{Y}_i^2 + \theta_i^2 V[\bar{y}_i] \tag{3.3}$$

$$\begin{aligned} \operatorname{cov}[\theta_i^* \bar{y}_i, \theta_j^* \bar{y}_j] &= E[\theta_i^* \bar{y}_i \theta_j^* \bar{y}_j] - E[\theta_i^* \bar{y}_i]E[\theta_j^* \bar{y}_j] \\ &= \operatorname{cov}[\theta_i^*, \theta_j^*]\bar{Y}_i \bar{Y}_j \end{aligned} \tag{3.4}$$

Reemplazando en (3.2)

$$\begin{aligned} V[Y^*] &= \sum_{i=1}^{T} V[\theta_i^*]V[\bar{y}_i] + \sum_{i=1}^{T} V[\theta_i^*]\bar{Y}_i^2 + \sum_{i=1}^{T} \theta_i^2 V[\bar{y}_i] \\ &\quad + 2\sum_{i<j}^{T} \operatorname{cov}[\theta_i^*, \theta_j^*]\bar{Y}_i \bar{Y}_j \\ &= \sum_{i=1}^{T} V[\theta_i^*]E[\bar{y}_i^2] + 2\sum_{i<j}^{T} [\theta_i^*, \theta_j^*]E[\bar{y}_i \bar{y}_j] + \sum_{i=1}^{T} \theta_i^2 V[\bar{y}_i] \end{aligned} \tag{3.5}$$

que puede expresarse en forma matricial como:

$$V[Y^*] = E\{\underline{\bar{y}}^T V[\underline{\theta}^*]\underline{\bar{y}}\} + \underline{\theta}^T V[\underline{\bar{y}}]\underline{\theta} \tag{3.6}$$

donde ;

$$\underline{\bar{y}} = (\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_T)^T$$

$$\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_T)^T$$

$V[\underline{\theta}^*]$ = Matriz de varianza-covarianza de todos los tamaños de dominio.

$V[\underline{\bar{y}}]$ = Matriz de varianza-covarianza de medias muestrales.

Antes de considerar el teorema siguiente, se presenta alguna notación ya utilizada y otra nueva:

$\underline{\theta}_D = (\theta_1, \theta_2, \ldots, \theta_M)^T$ (Vector de tamaños de dominio dependientes).

$\underline{\theta}_I = (\theta_{M+1}, \theta_{M+2}, \ldots, \theta_T)^T$ (Vector de tamaños de dominio independientes).

$\underline{0} = (0, \ldots, 0)^T$ (Un vector nulo).

$\underline{N} = (N_1, N_2, \ldots, N_M)^T$ (El vector de tamaños de marco).

G una matriz de tamaño M por $T-M(=S)$ tal que $\underline{\theta}_D = [\underline{N} \;\; G]\begin{bmatrix}1\\ \underline{\theta}_I\end{bmatrix}$

$\bar{\underline{y}}_D = (y_1, y_2, \ldots, y_M)^T$ (Vector de medias muestrales dentro de dominios dependientes).

$\bar{\underline{Y}}_D = (\bar{Y}_1, \bar{Y}_2, \ldots, \bar{Y}_M)^T$ (Vector de medias poblacionales dentro de dominios dependientes).

$\bar{\underline{y}}_I = (\bar{y}_{M+1}, \bar{y}_{M+2}, \ldots, \bar{y}_T)^T$ (Vector de medias muestrales dentro de dominios independientes).

$\bar{\underline{Y}}_I = (\bar{Y}_{M+1}, \bar{Y}_{M+2}, \ldots, \bar{Y}_T)^T$ (Vector de medias poblacionales dentro de dominios independientes).

$\bar{\underline{Y}} = (\bar{Y}_1, \bar{Y}_2, \ldots, \bar{Y}_T)^T$ (Vector de medias poblacionales) (3.7)

Teorema 3.1.

$$V[Y^*] = \bar{\underline{Y}}^T V(\underline{\theta}^*)\bar{\underline{Y}} + \underline{\theta}^T V[\bar{\underline{y}}]\underline{\theta} + \underline{d}(V^T(\bar{\underline{y}}_I))\underline{d}(V(\underline{\theta}_I^*))$$
$$+ \underline{d}^T(V(\bar{\underline{y}}_D))\underline{d}(GV(\underline{\theta}_I^*)G^T)$$

donde:

$$\underline{d}(V(\bar{\underline{y}}_D)) = \left(\frac{\sigma_1^2}{n_{11}}, \frac{\sigma_2^2}{n_{22}}, \ldots, \frac{\sigma_M^2}{n_{MM}}\right)^T \tag{3.8}$$

$$\underline{d}(V(\bar{\underline{y}}_I)) = \left(\frac{\sigma_{M+1}^2}{n_{M+1,M+1}}, \frac{\sigma_{M+2}^2}{n_{M+2,M+2}}, \ldots, \frac{\sigma_T^2}{n_{TT}}\right)^T \tag{3.9}$$

$$\underline{d}(V(\underline{\theta}_I^*)) = (V(\theta_{M+1}^*), V(\theta_{M+2}^*), \ldots,$$

$$\underline{d}(V(\underline{\theta}_I^*)) = \left(\sigma^2_{\theta^*_{M+1}}, \sigma^2_{\theta^*_{M+2}}, \ldots, \sigma^2_{\theta^*_T}\right)^T \tag{3.10}$$

$\underline{d}(GV(\underline{\theta}^*)G^T)$ = El vector cuyos elementos son la diagonal de la matriz $GV(\underline{\theta}_I^*)G^T$. (3.11)

Demostración.

$$V[\underline{\theta}^*] = V\begin{bmatrix} \underline{\theta}_D^* \\ \underline{\theta}_I^* \end{bmatrix}$$

$$= V\begin{bmatrix} \underline{N} & G \\ \underline{0} & I \end{bmatrix}\begin{bmatrix} 1 \\ \underline{\theta}_I \end{bmatrix}$$ donde I = Una matriz idéntica de tamaño $(T-M)$ por $(T-M)$.

$$= \begin{bmatrix} GV(\underline{\theta}_I^*)G^T & GV(\underline{\theta}_I^*) \\ V(\underline{\theta}_I^*)G^T & V(\underline{\theta}_I^*) \end{bmatrix} \tag{3.12}$$

Ahora usando (2.6):

$$V[y^*] = E\left[\underline{\bar{y}}^T\begin{bmatrix} GV(\underline{\theta}_I^*)G^T & GV(\theta_I^*) \\ V(\underline{\theta}_I^*)G^T & V(\theta_I^*) \end{bmatrix}\underline{\bar{y}}\right] + \underline{\theta}^T V[\underline{\bar{y}}]\underline{\theta} \tag{3.13}$$

$$= E\{\underline{\bar{y}}_D^T GV(\underline{\theta}_I^*)G^T\underline{\bar{y}}_D\} + 2E\{\underline{y}_D^T GV(\underline{\theta}_I^*)\underline{\bar{y}}_I\} + E\{\underline{\bar{y}}_I V(\underline{\theta}_I^*)\underline{\bar{y}}_I\} + \underline{\theta}^T V[\underline{\bar{y}}]\underline{\theta}$$

pero,

$$E\{\underline{\bar{y}}_D^T GV(\underline{\theta}_I^*)G^T\underline{\bar{y}}_D\} = \underline{\bar{Y}}_D^T GV(\underline{\theta}_I^*)G^T\underline{\bar{Y}}_D + d^T(V(\underline{\bar{y}}_D))\underline{d}(GV(\underline{\theta}_I^*)G^T) \tag{3.14}$$

En una manera similar puede demostrarse

$$E\{\underline{\bar{y}}_D^T GV(\underline{\theta}_I^*)\underline{\bar{y}}_I\} = \underline{\bar{Y}}_D^T GV(\underline{\theta}_I^*)\underline{\bar{Y}}_I \tag{3.15}$$

$$E\{\underline{\bar{y}}_I^T V(\underline{\theta}_I^*)\underline{\bar{y}}_I\} = \underline{\bar{Y}}_I^T V(\underline{\theta}_I^*)\underline{\bar{Y}}_I + \underline{d}^T(V(\underline{\bar{y}}_I))\underline{d}(V(\underline{\theta}_I^*)) \tag{3.16}$$

Reemplazando en (3.13)

$$V[y^*] = \underline{\bar{y}}_D^T GV(\underline{\theta}_I^*)G^T\underline{\bar{y}}_D + \underline{d}^T(V(\underline{\bar{y}}_D^T))\underline{d}(GV(\underline{\theta}_I^*)G^T) + 2\underline{\bar{y}}_D^T GV(\underline{\theta}_I^*)\underline{\bar{y}}_I$$

$$+ \underline{\bar{y}}_I^T V(\underline{\theta}_I^*)\underline{\bar{y}}_I + \underline{d}^T(V(\underline{\bar{y}}_I))\underline{d}(V(\underline{\theta}_I^*)) + \underline{\theta}^T V(\underline{\bar{y}})\underline{\theta}$$

$$= \underline{\bar{y}}^T V(\underline{\theta}^*)\underline{\bar{y}} + \underline{\theta}^T V(\underline{\bar{y}})\underline{\theta} + \underline{d}^T(V(\underline{\bar{y}}_I))\underline{d}(V(\underline{\theta}_I^*))$$

$$+ \underline{d}^T(V(\underline{\bar{y}}_D))\underline{d}(GV)\underline{\theta}_I^*)G^T) \qquad (3.17)$$

Para muestras grandes una fórmula aproximada para $V[y^*]$ puede obtenerse reemplazando los tamaños de muestra verdaderos correspondientes a cada dominio en (3.8) y (3.9) por sus "valores esperados" respectivos. La distribución asintótica de θ_I^* fué obtenida por Ospina (1985).

Ejemplo 3.2.

$M = 3$ (Tres marcos)

$\underline{\theta} = (\theta_1,\theta_2,\theta_3,\theta_{12},\theta_{13},\theta_{23},\theta_{123})^T$

$\quad = (\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6,\theta_7)^T$

$\underline{\theta}_D = (\theta_1,\theta_2,\theta_3)^T$

$\underline{\theta}_I = (\theta_4,\theta_5,\theta_6,\theta_7)^T$

$\underline{N} = (N_1,N_2,N_3)^T$

$\underline{y}_D = (\bar{y}_1,\bar{y}_2,\bar{y}_3)^T$

$\underline{\bar{y}}_I = (\bar{y}_4,\bar{y}_5,\bar{y}_6,\bar{y}_7)^T(=(\bar{y}_{12},\bar{y}_{13},\bar{y}_{23},\bar{y}_{123})^T)$

$\underline{y} = (\bar{y}_1,\bar{y}_2,\bar{y}_3,\bar{y}_4,\bar{y}_5,\bar{y}_6,\bar{y}_7)^T$

$\underline{\bar{Y}}_D = (\bar{Y}_1,\bar{Y}_2,\bar{Y}_3)^T$

$\underline{\bar{Y}}_I = (\bar{Y}_4,\bar{Y}_5,\bar{Y}_6,\bar{Y}_7)^T(=(\bar{Y}_{12},\bar{Y}_{13},\bar{Y}_{23},\bar{Y}_{123})^T)$

$\underline{\bar{Y}} = (\bar{Y}_1,\bar{Y}_2,\bar{Y}_3,\bar{Y}_4,\bar{Y}_5,\bar{Y}_6,\bar{Y}_7)^T$

Pero,

$$\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \\ \theta_7 \end{bmatrix} = \begin{bmatrix} N_1 & -1 & -1 & 0 & -1 \\ N_2 & -1 & 0 & -1 & -1 \\ N_3 & 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ \theta_4 \\ \theta_5 \\ \theta_6 \\ \theta_7 \end{bmatrix}$$

Por tanto:

$$G = \begin{bmatrix} -1 & -1 & 0 & -1 \\ -1 & 0 & -1 & -1 \\ 0 & -1 & -1 & -1 \end{bmatrix}$$

$$V[\underline{\bar{y}}] = \begin{bmatrix} V[\underline{\bar{y}}_D] & \underline{0} \\ \underline{0} & V[\underline{\bar{y}}_I] \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\sigma_1^2}{n_{11}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{\sigma_2^2}{n_{22}} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\sigma_3^2}{n_{33}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\sigma_4^2}{n_{44}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{\sigma_5^2}{n_{55}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{\sigma_6^2}{n_{66}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{\sigma_7^2}{n_{77}} \end{bmatrix}$$

donde:

$$\frac{\sigma_4^2}{n_{44}} = \frac{\sigma_{12}^2}{n_{12}}, \ \frac{\sigma_5^2}{n_{55}} = \frac{\sigma_{13}^2}{n_{13}}, \ \frac{\sigma_6^2}{n_{66}} = \frac{\sigma_{23}^2}{n_{23}}, \ \frac{\sigma_7^2}{n_{77}} = \frac{\sigma_{123}^2}{n_{123}} \qquad (3.18)$$

Sea $c_{ii} = V(\theta_i^*)$, $\quad i = 1,\ldots,7$

$$c_{ij} = \mathrm{COV}(\theta_i^*,\theta_j^*), \quad i,j = 1,\ldots,7;\ i \neq j \tag{3.19}$$

Por tanto:

$$V[\underline{\theta}_I^*] = \begin{bmatrix} c_{44} & c_{45} & c_{46} & c_{47} \\ c_{45} & c_{55} & c_{56} & c_{57} \\ c_{46} & c_{56} & c_{66} & c_{67} \\ c_{47} & c_{57} & c_{67} & c_{77} \end{bmatrix}$$

$$GV(\underline{\theta}_I^*) = \begin{bmatrix} -1 & -1 & 0 & -1 \\ -1 & 0 & -1 & -1 \\ 0 & -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} c_{44} & c_{45} & c_{46} & c_{47} \\ c_{45} & c_{55} & c_{56} & c_{57} \\ c_{46} & c_{56} & c_{66} & c_{67} \\ c_{47} & c_{57} & c_{67} & c_{77} \end{bmatrix}$$

$$= \begin{bmatrix} c_{44}\ c_{45}\ c_{47} & c_{55}\ c_{45}\ c_{57} & c_{46}\ c_{56}\ c_{67} & c_{77}\ c_{47}\ c_{57} \\ c_{44}\ c_{46}\ c_{47} & c_{45}\ c_{56}\ c_{57} & c_{66}\ c_{46}\ c_{67} & c_{77}\ c_{47}\ c_{67} \\ c_{45}\ c_{46}\ c_{47} & c_{55}\ c_{56}\ c_{57} & c_{66}\ c_{56}\ c_{67} & c_{77}\ c_{57}\ c_{67} \end{bmatrix}(-1)$$

$$= [(h_{ij})],\ i = 1,2,3,\ j = 1,2,3,4 \tag{3.20}$$

$$GV(\theta_I^*)G^T = [(g_{ij})], \quad i = 1,2,3;\ j = 1,2,3 \tag{3.21}$$

$$= \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{bmatrix}$$

$$g_{11} = c_{44}+c_{55}+c_{77}+2c_{45}+2c_{47}+2c_{57}$$

$$g_{22} = c_{44}+c_{66}+c_{77}+2c_{46}+2c_{47}+2c_{67}$$

$$g_{33} = c_{55}+c_{66}+c_{77}+2c_{56}+2c_{57}+2c_{67}$$

$$g_{12} = c_{44}+c_{77}+c_{45}+c_{46}+2c_{47}+c_{56}+c_{57}+c_{67}$$

$$g_{13} = c_{55}+c_{77}+c_{45}+c_{46}+c_{47}+c_{56}+2c_{57}+c_{67}$$

$g_{23} = c_{66}+c_{77}+c_{45}+c_{46}+c_{47}+c_{56}+c_{57}+2c_{67}$

$g_{21} = g_{12}$

$g_{31} = g_{13}$

$g_{32} = g_{23}$

$$\underline{\bar{y}}_D^T GV(\underline{\theta}_I^*)G^T\underline{\bar{y}}_D = (\bar{y}_1,\bar{y}_2,\bar{y}_3)\begin{bmatrix} g_{11} & g_{12} & g_{13} & \bar{y}_1 \\ g_{21} & g_{22} & g_{23} & \bar{y}_2 \\ g_{31} & g_{32} & g_{33} & \bar{y}_3 \end{bmatrix}$$

$$= \sum_{i=1}^{3}\sum_{j=1}^{3}\bar{y}_i\bar{y}_j g_{ij} \tag{3.22}$$

$$\underline{\bar{y}}_D^T GV(\underline{\theta}_I^*)\underline{\bar{y}}_I = (\bar{y}_1,\bar{y}_2,\bar{y}_3)\begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \end{bmatrix}\begin{bmatrix} \bar{y}_4 \\ \bar{y}_5 \\ \bar{y}_6 \\ \bar{y}_7 \end{bmatrix}$$

$$= \sum_{i=1}^{3}\sum_{j=4}^{7}\bar{y}_i\bar{y}_j h_{i(j-3)} \tag{3.23}$$

$$\underline{\bar{y}}_I^T V(\underline{\theta}_I^*)\underline{\bar{y}}_I = (\bar{y}_4,\bar{y}_5,\bar{y}_6,\bar{y}_7)\, V(\underline{\theta}_I^*)\begin{bmatrix} \bar{y}_4 \\ \bar{y}_5 \\ \bar{y}_6 \\ \bar{y}_7 \end{bmatrix}$$

$$= \sum_{i=4}^{7}\sum_{j=4}^{7}\bar{y}_i\bar{y}_j c_{ij} \tag{3.24}$$

$$\underline{d}^T(V(\underline{\bar{y}}_D))\underline{d}(GV(\underline{\theta}_I^*)G^T) = \left(\frac{\sigma_1^2}{n_{11}}, \frac{\sigma_2^2}{n_{22}}, \frac{\sigma_3^2}{n_{33}}\right)\begin{bmatrix} g_{11} \\ g_{22} \\ g_{33} \end{bmatrix}$$

$$= \sum_{i=1}^{3}\frac{\sigma_i^2}{n_{ii}}\, g_{ii} \tag{3.25}$$

$$\underline{d}^T(V(\bar{\underline{y}}_I))\underline{d}(V(\underline{\theta}_I^*)) = \left(\frac{\sigma_4^2}{n_{44}}, \frac{\sigma_5^2}{n_{55}}, \frac{\sigma_6^2}{n_{66}}, \frac{\sigma_7^2}{n_{77}}\right)\begin{bmatrix} c_{44} \\ c_{55} \\ c_{66} \\ c_{77} \end{bmatrix}$$

$$= \sum_{i=4} \frac{\sigma_i^2}{n_{ii}} c_{ii} \tag{3.26}$$

$$\underline{\theta}^T V(\bar{\underline{y}})\underline{\theta} = (\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6,\theta_7)V|\bar{\underline{y}}|\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \\ \theta_7 \end{bmatrix}$$

$$= \sum_{i=1}^{7} (\theta_i) \frac{\sigma_i^2}{n_{ii}} \tag{3.27}$$

Ahora, reemplazando en (3.17):

$$V[y^*] = \sum_{i=1}^{3}\sum_{j=1}^{3} \bar{y}_i\bar{y}_j g_{ij} + \sum_{i=1}^{3} \frac{\sigma_i^2}{n_{ii}} g_{ii} + 2\sum_{i=1}^{3}\sum_{j=4}^{7} \bar{y}_i\bar{y}_j h_{i(j-3)}$$

$$+ \sum_{i=4}^{7}\sum_{j=4}^{7} \bar{y}_i\bar{y}_j c_{ij} + \sum_{i=4}^{7} \frac{\sigma_i^2}{n_{ii}} c_{ii} + \sum_{i=1}^{7} \theta_i^2 \frac{\sigma_i^2}{n_{ii}} \tag{3.28}$$

La fórmula generla para $V[y^*]$ dada por (3.17) puede escribirse alternativamente como

$$V[y^*] = \sum_{i=1}^{M}\sum_{j=1}^{M} \bar{y}_i\bar{y}_j g_{ij} + \sum_{i=1}^{M} \frac{\sigma_i^2}{n_{ii}} g_{ii} + 2\sum_{i=1}^{M}\sum_{j=M+1}^{T} \bar{y}_i\bar{y}_j h_{i(j-M)}$$

$$\sum_{i=M+1}^{T}\sum_{j=M+1}^{T} \bar{y}_i\bar{y}_j c_{ij} + \sum_{i=M+1}^{T} \frac{\sigma_i^2}{n_{ii}} c_{ii} + \sum_{i=1}^{T} \theta_i^2 \frac{\sigma_i^2}{n_{ii}} \tag{3.29}$$

donde c_{ij}, g_{ij} y h_{ij} se definen en forma similar a (3.19), (3.20) y (3.21).

El EMV de $V[Y^*]$, dado por (3.29), puede obtenerse reemplazando cada parámetro por su EMV como sigue:

$$V^*[Y^*] = \sum_{i=1}^{M} \sum_{j=1}^{M} \bar{y}_i \bar{y}_j g^*_{ij} + \sum_{i=1}^{M} \frac{S_i^2}{n_{ii}} g^*_{ii} + 2 \sum_{i=1}^{M} \sum_{j=M+1}^{T} \bar{y}_i \bar{y}_j h^*_{i(j-M)}$$

$$+ \sum_{i=M+1}^{T} \sum_{j=M+1}^{T} \bar{y}_i \bar{y}_j c^*_{ij} + \sum_{i=M+1}^{T} \frac{S_i^2}{n_{ii}} c^*_{ii} + \sum_{i=1}^{T} \theta_i^{*2} \frac{S_i^2}{n_{ii}} \quad (3.30)$$

donde:

$$\bar{y}_i = \sum_{j=1}^{n_{ii}} \frac{y_{ij}}{n_{ii}} \text{ (MLE of } \bar{Y}_i)$$

$$S_i^2 = \sum_{j=1}^{n_{ii}} \frac{(y_{ij}-\bar{y}_i)^2}{n_{ii}} \text{ (MLE of } \sigma_i^2)$$

c^*_{ij} es el elemento (i,j) de la matriz $V^*(\underline{\theta}^*_I)$ la cual es el EMV de $V(\underline{\theta}^*_I)$.
g^*_{ij} es el elemento (i,j) de la matriz $GV^*(\underline{\theta}^*_I)G^T$.
h^*_{ij} es el elemento (i,j) de la matriz $GV^*(\underline{\theta}^*_I)$.
(3.30) puede escribirse también en forma matricial.

$$V^*[Y^*] = \underline{\bar{y}}_D^T GV^*(\underline{\theta}^*_I)G^T \underline{\bar{y}}_D + \underline{d}^T(V^*(\underline{\bar{y}}_D))\underline{d}(GV^*(\underline{\theta}^*_I)G^T)$$

$$+ 2\underline{\bar{y}}_D^T GV^*(\underline{\theta}^*_I)\underline{\bar{y}}_I + \underline{y}_I^T V^*(\underline{\theta}^*_I)\underline{\bar{y}}_I + \underline{d}^T(V^*(\underline{\bar{y}}_I))\underline{d}(V^*(\underline{\theta}^*_I))$$

$$+ \underline{\theta}^{*T} V^*(\underline{\bar{y}})\underline{\theta}^* \quad (3.31)$$

4. Proceso de Simulación.

Los resultados teóricos derivados en las Secciones 2 y 3 se validaron mediante simulaciones para las situaciones de dos y tres marcos. El proceso de simulación consistió de varios pasos: (1) Se generaron muestras aleatorias de distribuciones binomiales (dos marcos) y multinomiales (tres marcos) haciendo uso de las subrutinas GGBN y GGMTN del IMSL (Institute of Mathematical Statistics Library, 1980). (2) Se registró el número de unidades en la muestra total que pertenecían a cada dominio y se escribieron dos subrutinas para obtener las estimaciones máximo-verosímiles para cada tamaño de dominio. Se utilizó el procedimiento conocido como Programación Geométrica Subrogada presentado por Cooke (1983) para obtener las estimaciones máximo-verosímiles y se halló que trabaja muy eficientemente en esta situación. (3) Se utilizó la subrutina GGNML (Normal Random Deviate Generator) del IMSL (1980) para generar observaciones normales para cada una de las poblaciones que corresponde a cada uno de los dominios usando los tamaños de muestra obtenidos en el paso (2). (4) Se registraron las estimaciones máximo-verosímiles para el total usando los resultados de los pasos (2) y (3). (5) El test para normalidad dado por Ryan y Joiner el cual es equivalente al test de Shapiro-Wilk (1965) se aplicó aquí para probar la bondad de la aproximación en diferentes niveles de significación.

Los tamaños de los marcos se tomaron grandes para ser consistentes con la aproximación multinomial (el tamaño mínimo igual a 10.000) El tamaño del dominio traslapado se consideró al menos el 10% y como máximo el 90% de cualquiera de los marcos. Los tamaños de muestra fueron diferentes en cada caso pero nunca menores de 50 para obtener una buena aproximación al supuesto de muestras grandes. Se tomaron 75 muestras diferentes para decidir acerca de qué tan adecuada es la aproximación normal para la distribución del EMV del total de la característica Y bajo consideración. Se asignaron diferentes valores cada vez a los tamaños traslapados. El valor observado del esdístico R_p (probability plot correlation coefficient) se comparó con los valores críticos dados por Ryan y Joiner. La hipótesis de normalidad no fué rechazada en ningún caso para la situación de dos marcos. En la situación de tres marcos el rechazo sólo ocurrió en 1 de los 15 casos considerados. Estos rechazos iniciales no deben preocupar ya que las "correlaciones límites" observadas al incrementar las muestras a 1000 mostraron que el supuesto normal era correcto. Se consideraron situaciones diferentes con tamaños de marco iguales y diferentes, así como afijaciones muestrales proporcionales y no proporcionales. Se tomaron en cuenta también casos especiales con medias y varianzas diferentes y no muy diferentes. El sesgo relativo negativo del EMV de la varianza del EMV del total fué la cuestión más importante para resaltar aquí.

5. Conclusiones.

Un procedimiento general usado para estimar totales y medias poblaciones de características específicas se ha presentado. Esta aproximación general se ha tomado usando el criterio de máxima-verosimilitud. La distribución normal asintótica para estos estimadores fué también identificada haciendo uso de los resultados teoricos obtenidos por Ospina y Cochran (1985) para estimaciones de tamaños de dominio.

Esta aproximación máximo-verosímil es muy útil en aquellos casos donde los tamaños de marco originales son tan grandes que determinar los tamaños exactos de los traslapes es poco práctico. La aproximación normal de los EMV de los totales trabaja bien siempre y cuando cada dominio traslapado cubra al menos el 5% de los marcos a los cuales pertenece. También parece dar resultados satisfactorios independientemente de la afijación muestral a los marcos o las medias y varianzas verdaderas dentro de cada dominio. El proceso de simulación no sugirió restricciones en este aspecto.

Un comentario final debe hacerse con respecto al EMV de la varianza del EMV para el total de una característica. El EMV subestimó consistentemente la varianza verdadera. Una corrección leve del estimador propuesto podrá mejorar este sesgo negativo. Sin embargo esto requiere un análisis más cuidadoso.

*

BIBLIOGRAFIA

Cochran, R.S. (1965). "Theory and Application of Multiple Frame Surveys". Iowa State University, Ames, Iowa. Tesis de Doctorado.

Cooke, W.P. (1983). "Surrogate geometric programming estimates of restricted multinomial proportions," Communications in Statistics, 12, 291-305.

Fuller, W.A, y L.F. Burmeister (1972). "Estimators for samples selected from two overlapping frames," Proceedings of the Social Statistics Section of the American Statistical Association, 245-249.

Hartley, H.O. (1962). "Multiple frame surveys," Proceedings of Social Science Section of American Statistical Association Meetings, Minneapolis, Minnesota.

Hogg, R.V. y A.T. Craig (1978). Introduction to Mathematical Statistics, 4ª Edition, MacMillan Publishing Co., Inc New York, 120.

IMSL, Institute of Mathematical Statistics Library (1980), Houston, Texas.

Lund, R.E. (1968). "Estimators in multiple frame surveys," Proceedings of Social Science Section of American Statistical Association, 282-288.

Mood, A.M., F.A. Graybill, y D.C. Boes (1974). Introduction to the Theory of Statistics, 3ª Edition, McGraw-Hill Book Company, New York, 175-181, 279, 315-321, 331-338, 358-362.

Ospina, D. (1985). "A Maximum Likelihood Approach in Multiple Frame Surveys", University of Wyoming, Laramie, Wyoming. Tesis de Doctorado.

———— y Cochran (August 1985). "An Asymptotic Distribution for the MLE of Domain Sizes in Multiple Frame Situations, artículo presentado en el encuentro Conjunto de ASA-ENAR-WANAR-IMS en Las Vegas.

Rao, C.R. (1973). Linear Statistical Inference and its Aplications, 2ª Edition, John Wiley and Sons, New York, 351-366.

Ryan, T.A., Jr., B.L. Joiner. "Normal probability plots and tests for normality", Reporte Técnico, Statistics Department, Pennsylvania State University.

Shapiro, S.S. y M.B. Wilk (1965). "An analysis of variance test for normality (complete samples), "Biometrika, 52, (1965), 591.

* *