UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

" ESTIMADORES DE RAZON Y DE DIFERENCIA MULTIPLE PARA ESTIMACIONES INTERCENSALES POR MEDIO DE MUESTREO. "

T E S I S

Que para obtener el Titulo de:

A C T U A R I O

p r e s e n t a

JORGE L. MONTEJANO ZURITA

MEXICO, D. F. 1971

A MIS PADRES

HILDA Y SALVADOR

A ELVIA (CARI)

A CHAVA Y A TAVO

A MIS FAMILIARES Y AMIGOS

QUIERO DAR LAS GRACIAS A MI MAESTRO Y AMIGO SERGIO VARGAS POR SU INAPRECIABLE AYUDA EN EL LOGRO DE ESTA TESIS. ASI COMO TAMBIEN A MIGUEL CERVERA POR SUS VALIOSOS CONSEJOS.

"ESTIMADORES DE RAZON Y DE DIFERENCIA MULTIPLE

PARA ESTIMACIONES INTERCENSALES POR

MEDIO DE MUESTREO."

CONTENIDO

I N T R O D U C C I O N

En muchas situaciones estamos en posición de aprovechar la información auxiliar, básica o complementaria, relativa a una variable o característica correlacionada con la que se estudia, con el objeto de obtener estimaciones más precisas que las calculadas a partir del muestreo irrestricto aleatorio.

El enfoque principal de la presente tesis es el de utilizar más de una variable auxiliar, para mejorar la estimación de medias y totales de la variable sujeta a estudio, mediante estimadores de razón y de diferencia múltiple, cuando las variables auxiliares son valores conocidos de la misma variable en períodos anteriores.

Este enfoque de muestreo pudo ser antes prohibitivo, debido principalmente a la gran cantidad de operaciones incluídas en los cálculos, en la actualidad esto no constituye ningún obstáculo, debido al gran desarrollo que han tenido las computadoras en los últimos tiempos, valiéndose de su vasta capacidad y alta velocidad para resolver problemas muy complejos fácilmente.

Esta tesis intenta aplicar técnicas específicas a problemas también específicos. El tipo de problemas donde se pueden utilizar estas técnicas, es en aquellas situaciones en las que se llevan a cabo censos periódicamente. La idea fundamental es la de efectuar estimaciones entre censo y censo, para tener información sin tener que esperar el censo siguiente, utilizando información de censos anteriores.

Es importante hacer notar que en esta tesis se hace la suposición de que las unidades de muestreo son las mismas en diferentes períodos de tiempo. Si las muestras son independientes de período en período, la varianza de los estimadores de razón y de diferencia se hace igual a la varianza de una media simple, debido a que los términos de covarianza son cero, resulta entonces inútil la utilización de las variables auxiliares y en consecuencia deberán utilizarse otros métodos de estimación con los que se puedan aprovechar dichas variables, tales como los métodos descritos en Hansen Hurwitz y Madow (1953, pag. 490), los cuales están fuera del enfoque de esta tesis.

Si existe un traslape parcial se pueden aprovechar los métodos aquí descritos; sin embargo como se mencionó anteriormente aquí se supone que el traslape es total.

Supongamos que lo que se quiere estimar es la media poblacional μ_y de una variable y; si se puede encontrar un conjunto de p variables $x_1, x_2, \ldots x_p$, altamente correlacionadas con la variable $\underline{y}$, y si además se conocen las medias poblacionales de dichas variables, las cuales se denotarán por $\mu_1, \mu_2, \ldots\ldots, \mu_p$ entonces se estará en posición de proponer la siguiente expresión general para la estimación de μ_y utilizando p variables auxiliares :

$$\mu_y^* = \sum_{i=1}^{p} w_i \left[m_o + g_i(\mu_i - m_i) \right]$$

Donde w_i es un factor de ponderación para cada variable w_i $(i = 1, 2, \ldots p)$, sujeto a la restricción $\sum_{i=1}^{p} w_i = 1$

m_o es la media muestral (muestreo irrestricto aleatorio) de la variable $\underline{y}$, m_i es la media muestral de la variable x_i .

Las g_i pueden ser consideradas como coeficientes de corrección del estimador simple m_o.

Es interesante hacer notar algunos valores posibles de g_i.

a) Cuando $g_i = 0$ para toda $i = 1, 2, \ldots, p$, el estimador general se reduce al estimador simple.

$$\mu_y^* = m_o$$

b) Si cada g_i es igual al cociente de las medias muestrales m_o y m_i, es decir m_o/m_i, entonces la forma general se convierte en el estimador de razón múltiple, que es como sigue:

$$\mu_y^* = \sum_{i=1}^{p} w_i r_i \mu_i$$

Donde $r_i = \frac{m_o}{m_i}$ para toda i.

c) Si cada g_i es una constante cualquiera k_i $(i = 1, 2, \ldots, p)$, se tendrá lo que se llamará estimador de diferencia múltiple :

$$\mu_y^* = \sum_{i=1}^{p} w_i \left[m_o + k_i (\mu_i - m_i) \right]$$

las k_i se pueden aproximar a partir de encuestas anteriores o de información previa. Si x_i son los valores de la característica y en períodos anteriores, y la naturaleza de las variables no ha cambiado significativamente, el valor de k_i será un valor cercano a 1. (Hansen, Hurwitz y Madow 1953).

d) En el caso en que $g_i = b_i$ $(i = 1, 2, \ldots, p)$, donde b_i es el i-ésimo componente de un vector b de coeficientes de regresión múltiple, como se define usualmente en los libros de estadística matemática, se obtiene un estimador de regresión múltiple. Como b_i es ya un factor de ponderación no es necesario incluir el factor w.

De aquí que $g_i = \frac{b_i}{w_i}$ y si se substituye esto en forma general se tendrá lo siguiente :

$$\overset{*}{\mu}_y = m_0 + \sum_{i=1}^{p} b_i (\mu_i - m_i)$$

b es un vector aleatorio, lo cual significa que varía de muestra a muestra, y por tanto sujeto a errores de muestreo.

Hay que hacer notar que el valor óptimo de g_i sería :

$$g_i = \frac{(\mu_i - m_0)}{(\mu_i - m_i)}$$

por lo que el estimador general se reduciría a :

$$\overset{*}{\mu}_y = \mu_y$$

Lo anterior carece de sentido práctico, puesto que μ_y

es precisamente lo que se quiere estimar.

Se pueden hacer otras definiciones distintas de g_i para obtener otros estimadores de μ_y pero si bien es fácil proponer diversos estimadores, el problema principal radica en evaluar su precisión y sus propiedades.

En esta tesis se llevará a cabo un análisis detallado de los estimadores de razón y de diferencia múltiple, en muestreo irrestricto aleatorio y en muestreo estratificado. Se incluye un tratamiento de los estimadores de razón y de diferencia simples para una mejor comprensión del caso múltiple.

CAPITULO I

ESTIMADORES DE RAZON Y DE DIFERENCIA SIMPLE.

1.1) Estimador de razón simple.

El propósito de esta sección es tratar el método de estimación de razón simple, el cual en muchas situaciones aumenta la precisión de las estimaciones muestrales. El método requiere del uso de una variable auxiliar $\underline{x}$, de la cual tenemos información detallada para ajustar o calibrar nuestras estimaciones de la media de una variable $\underline{y}$ en una población finita, la cual denotaremos por μ_y. Vamos a suponer que la variable de interés $\underline{y}$ está altamenta correlacionada con la variable $\underline{x}$, de la cual conocemos la media poblacional μ_x ; el estimador de la media a partir de la muestra es como sigue :

$$\hat{\mu}_y = \frac{m_y}{m_x}\mu_x = r\mu_x \qquad (1.1.1)$$

Donde m_y y m_x son las medias muestrales de las variables y y x respectivamente. Además suponemos que :

$$E(m_y) = \mu_y \quad y \quad E(m_x) = \mu_x$$

El estimador $\hat{\mu}_y$ será insesgado, si en promedio es igual a μ_y. Si esto es verdadero el estimador de razón producirá la media de la población en promedio.

Desafortunadamente ésto no es siempre verdadero para el estimador de razón, porque el valor promedio de una razón, no es en general igual a la razón de los valores promedio.

En la mayoría de las situaciones de muestreo aplicado, el sesgo como se demostrará más adelante es pequeño, de modo que el error debido al sesgo es más que compensado por una reducción en la varianza del estimador, siempre y cuando y y x estén altamente correlacionadas.

El sesgo tiende a ser pequeño si la correlación entre r y m_x es pequeña, se puede demostrar fácilmente que $\hat{\mu}_y$ es consistente, es decir, el sesgo disminuye al aumentar el tamaño de muestra.

La fórmula exacta para la varianza de $\hat{\mu}_y$ incluye una serie infinita y por tanto no es útil en la práctica, pero se pueden usar los primero términos de la serie y llegar a la expresión siguiente :

$$V(\hat{\mu}_Y) = \frac{(1-f)}{n} \mu_Y^2 (C_Y^2 + C_X^2 - 2\rho C_Y C_X)$$

Donde N y n son el tamaño de la población y de la muestra respectivamente, $(1-f) = (1 - \frac{n}{N})$ es el factor de corrección para población finita, C_y y C_x son los coeficientes de variación de las variables y y x, el coeficiente de correlación entre y y x.

La fórmula anterior es la varianza de la distribución muestral y expresa la varianza en términos de parámetros de la población. Como rara vez o nunca se conocen esos parámetros, la fórmula se usa principalmente en trabajo teórico para comprobar varios métodos de estimación.

El estimador de razón no es siempre más preciso que el estimador simple. Se demostrará más adelante que el estimador $\hat{\mu}_Y$ es más preciso sí y sólo si ρ excede $C_x/2C_y$. Por lo tanto el estimador de razón no debe ser usado indistintamente, sino en las circunstancias apropiadas en las que produce un aumento considerable en precisión.

Alguna s veces el propósito del muestreo es estimar la razón misma, más bien que una media o total basados en la razón, ésta última la denotaremos por R ($R = \mu_Y/\mu_X$). Otras aplicaciones importantes del método de razón son la estimación de medias de sub-clases de una población y la estimación de medias en muestreo por conglomerados.

Esperanza y varianza del estimador.-

Se puede ver de (1.1.1.) que $\hat{\mu}_Y$ está en función de r, el cual a su vez es un estimador en que tanto el numerador como el denominador se encuentran sujetos a errores de muestreo, es decir varían de muestra a muestra.

Para hacer un análisis de $\hat{\mu}_Y$, tenemos que hacer primero un análisis de r y ver cual es su repercusión sobre μ_Y.

Si tomamos la esperanza y varianza de $\hat{\mu}_Y$ tenemos :

$$E(\hat{\mu}_Y) = \mu_X E(r) \tag{1.1.2}$$

$$V(\hat{\mu}_Y) = \mu_X^2 V(r) \tag{1.1.3}$$

Para calcular la esperanza de r se puede utilizar el siguiente argumento :

Consideremos las desviaciones de los valores muestrales de los valores poblacionales de las variables m_y y m_x, las cuales denotaremos por ϵ y δ .

Donde :

$$\epsilon = \frac{(m_y - \mu_y)}{\mu_y} \qquad \delta = \frac{(m_x - \mu_x)}{\mu_x}$$

Dichas desviaciones tienen las siguientes propiedades :

i) $E(\epsilon) = 0$

ii) $E(\delta) = 0$

iii) $E(\epsilon^2) = \frac{(1-f)}{n} C_y^2$

iv) $E(\delta^2) = \frac{(1-f)}{n} C_x^2$

v) $E(\epsilon\delta) = \frac{(1-f)}{n} \rho C_y C_x$

Las cuales se pueden demostrar fácilmente.

Por definición $$r = \frac{m_y}{m_x} \qquad (1.1.4)$$

De las expresiones de las desviaciones relativas, despejamos m_y y m_x, quedándonos lo siguiente :

$$m_y = \mu_y (1 + \epsilon) \qquad m_x = \mu_x (1 + \delta)$$

Si ésto lo substituímos en (1.1.4)

Por lo tanto

$$r = \frac{m_y}{m_x} = \frac{\mu_y}{\mu_x} \frac{(1+\epsilon)}{(1+\delta)} = R(1+\epsilon)(1+\delta)^{-1} \quad (1.1.5)$$

Necesitamos suponer que el valor absoluto de μ_x es tan grande, que la probabilidad de que m_x sea un valor cercano a cero, es insignificante y que la desviación relativa en valor absoluto, es suficientemente pequeña como para permitir que todos, excepto unos cuantos términos en el desarrollo de Taylor sean despreciables.

Desarrollaremos $(1+\delta)^{-1}$ en serie de potencias de usando el desarrollo de Taylor.

Al substituir este resultado en (1.1.5) y al efectuar operaciones se tiene que :

$$r = R(1 + \epsilon - \delta + \delta^2 - \epsilon\delta)$$

Además supondremos que la contribución de términos que contienen potencias mayores que la segunda es despreciable.

Por lo tanto consideraremos la siguiente aproximación* de r : $r = R(1 + \epsilon - \delta + \delta^2 - \epsilon\delta)$

---

* Se pueden encontrar otras aproximaciones al utilizar más términos de la serie de Taylor, pero el problema principal radica en evaluar la esperanza de términos mayores, para lo que se requiere de una algebra pesada, una segunda aproximación se puede encontrar en Sukhatme (1954).

Si tomamos la esperanza de la expresión anterior

$$E(r) = R\left[1 + E(\epsilon) - E(\delta) + E(\delta^2) - E(\epsilon\delta)\right]$$

Por las propiedades de las desviaciones relativas vistas anteriormente y al efectuar operaciones obtenemos el siguiente resultado :

$$E(r) = R + R\,\frac{(1-f)}{n}\left(C_x^2 - \rho\, C_y C_x\right) \qquad (1.1.6)$$

Claramente podemos observar que r es un estimador sesgado de R.

$$SESGO\,(r) = R\,\frac{(1-f)}{n}\left(C_x^2 - \rho\, C_y C_x\right) \qquad (1.1.7)$$

Si substituímos (1.1.6) en (1.1.2) obtenemos la esperanza de $\hat{\mu}_y$.

$$E(\hat{\mu}_y) = \mu_y + \mu_y\,\frac{(1-f)}{n}\left(C_x^2 - \rho\, C_y C_x\right)$$

En consecuencia $\hat{\mu}_y$ es un estimador sesgado de μ_y.

Para el cálculo de la varianza de r se puede proceder de la siguiente manera :

Por definición $\quad V(r) = E\left[r - E(r)\right]^2 \qquad (1.1.8)$

Si utilizamos las aproximaciones de r y E (r) obtenidas anteriormente y la substituímos en (1.1.8), al efectuar operaciones y retener términos hasta de segundo grado, tenemos que :

$$V(r) = \frac{(1-f)}{n}\mu_y^2\left[C_y^2 + C_x^2 - 2\rho C_y C_x\right]$$

Geométricamente V(r) es una medida de dispersión de las rectas de regresión muestrales, alrededor de la recta de la población. Uno de los objetos principales de un diseño de muestra es un aumento en la precisión, es decir queremos que la varianza de r sea tan pequeña como sea posible.

En la fórmula anterior podemos observar que C_y, C_x y R son positivos, además el coeficiente de correlación está comprendido entre -1 y 1, por lo tanto la varianza de r será un mínimo cuando $\rho = 1$; ésto quiere decir que el método de razón tendrá varianza mínima cuando exista una correlación positiva perfecta entre y y x.

Otra forma de escribir la varianza de r es :

$$V(r) = \frac{(1-f)}{n}\frac{1}{\mu_x^2}\left[\sum_{i=1}^{N}(y_i - Rx_i)/N-1\right] \qquad (1.1.9)$$

Para encontrar la varianza $\hat{\mu}_y$ substituímos (1.1.9) en (1.1.3).

$$V(\hat{\mu}_y) = \frac{(1-f)}{n}\mu_y^2(C_y^2 + C_x^2 - 2\rho\, C_y C_x) \qquad (1.1.10)$$

Una expresión alternativa a la anterior es :

$$V(\hat{\mu}_y) = \frac{(1-f)}{n}\left[\sum_{i=1}^{N}(y_i - Rx_i)/N-1\right] \quad (1.1.11)$$

Como se verá posteriormente, esta forma simplifica la derivación del estimador de la varianza a partir de la muestra.

Sesgo del estimador de razón.-

En general, r es un estimador sesgado de R, pero es consistente, lo cual puede demostrar fácilmente.

La expresión del sesgo de acuerdo con (1.1.7) es :

$$\text{Sesgo } (r) = R\,\frac{(1-f)}{n}\,(C_x^2 - \rho\, C_y C_x)$$

En el caso en que la regresión de y sobre x sea una línea recta a través del origen, el sesgo será igual a cero; esto se puede verificar de la siguiente forma :

$$\text{Si sesgo } (r) = R\,\frac{(1-f)}{n}\,(C_x^2 - \rho\, C_y C_x) = 0$$

$$\text{Esto implica que : } RS_x = \rho\, S_y \quad (1.1.12)$$

Ahora para demostrar lo anterior supongamos que la recta de regresión pasa a través del origen.

$$Y_i = Bx_i$$

Entonces $$B = \mu_Y / \mu_X = R$$

De lo anterior se concluye que cuando la recta de regresión pasa a través del origen, R = B y por lo tanto el primer miembro de la ecuación (1.1.12) es igual a :

$$RS_x = BS_x = \frac{S_{YX}}{S_x} = \rho S_Y$$

Con lo cual se demuestra lo anterior.

De aquí que cuando la regresión de y sobre x es una línea recta a través del origen, el sesgo de r es igual a cero y por tanto un estimador insesgado de R. En la resolución de problemas prácticos, podemos fácilmente comprobar si los datos satisfacen aproximadamente esta condición.

Otro resultado importante es el debido a Hartley y Ross (1954) quienen proporcionan un límite superior de la razón del sesgo de r al error estandar.

Consideremos la siguiente relación :

$$\begin{aligned} COV(r, m_x) &= E(r m_x) - E(r)E(m_x) \\ &= E(m_y) - E(r)\,E(m_x) \\ &= -\mu_x \; SESGO(r) \qquad (1.1.13) \end{aligned}$$

Por otra parte

$$Cov(r, m_x) = -\rho^* S(r) S_{m_x}$$

De (1.1.12) y (1.1.13)

$$\frac{SESGO(r)}{S(r)} = -\rho^* CV de\ m_x$$

Finalmente

$$\left|\frac{SESGO(r)}{S(r)}\right| \leq CV\ de\ m_x$$

S (r) Error estandar de r. S(r)

ρ * Coeficiente de correlación entre r y m_x.

CV Coeficiente de variación.

El mismo límite se aplica a $\hat{\mu}_y$; de acuerdo a Cochran (1953), si el coeficiente de variación de m_x es menor que 0.1, el sesgo con mucha seguridad puede ser considerado insignificante en relación al error estandar.

Estimación de la varianza a partir de la muestra.-

De la sección 1.2 vimos que la varianza de $\hat{\mu}_y$ fué

$$V(\hat{\mu}_y) = \frac{(1-f)}{n}\left[\Sigma(y_i - Rx_i)/N-1\right]$$

Un estimador muestral de

$$\frac{\sum^{N}(y_i - Rx_i)^2}{N-1}$$

es

$$\frac{\sum^{n}(y_i - Rx_i)^2}{n-1}$$

Por lo tanto

$$v(\hat{\mu}_y) = \frac{(1-f)}{n(n-1)} \sum_{i=1}^{n} (y_i - r x_i)^2 = \frac{(1-f)}{n(n-1)} \left(\sum_{i}^{n} y_i^2 + r^2 \sum_{i}^{n} x_i^2 - 2r \sum_{i}^{n} y_i x_i \right)$$

Siendo esta última más fácil para efectuar los cálculos.

1.2) Estimadores de razón simple en muestreo estratificado.

Existen dos tipos de estimadores de razón en muestreo estratificado; el estimador separado y el estimador combinado. En el primero de ellos es necesario conocer las medias poblacionales de la variable auxiliar en cada estrato, mientras que en el estimador combinado únicamente es necesario conocer la media poblacional total de la variable auxiliar.

a) Estimador separado.

La expresión de este estimador es la siguiente :

$$\mu_y^s = \frac{1}{N} \sum_{h=1}^{H} N_h \hat{\mu}_y^{(h)} = \frac{1}{N} \sum_{h=1}^{H} N_h r^{(h)} \mu_x^{(h)}$$

En donde

$\hat{\mu}_y^{(h)}$ es el estimador de la media poblacional del estrato h.

$\mu_x^{(h)}$ es la media poblacional de la variable auxiliar $\underline{x}$ en el estrato h.

$r^{(h)}$ es la razón muestral en el estrato h.

N_h es el tamaño de la población en el estrato h.

H es el número de estratos.

La expresión para la varianza de μ_Y^s se puede encontrar como sigue :

Por definición

$$V(\hat{\mu}_Y^s) = E(\hat{\mu}_Y^s - \mu_Y)^2$$

Si consideramos la diferencia $(\hat{\mu}_Y^s - \mu_Y)$, la podemos expresar como :

$$\frac{1}{N}\sum_{h=1}^{H} N_h(\hat{\mu}_Y^{(h)} - \mu_Y^{(h)})$$

Y si elevamos al cuadrado este último término tenemos :

$$\frac{1}{N^2}\sum_{h=1}^{H} N_h^2(\hat{\mu}_Y^{(h)} - \mu_Y^{(h)}) + \frac{1}{N^2}\sum_{h=1}^{H} N_h N_\ell(\hat{\mu}_Y^{(h)} - \mu_Y^{(h)})(\hat{\mu}_Y^{(\ell)} - \mu_Y^{(\ell)})$$

Al tomar esperanza se tiene que

$$V(\hat{\mu}_Y^s) = \frac{1}{N^2}\sum_{h=1}^{H} N_h^2 V(\hat{\mu}_Y^{(h)})$$

* Al tomar esperanza del segundo término, éste desaparece porque el muestreo es independiente entre estratos.

Pero $\hat{\mu}_Y^{(h)}$ es un estimador de razón de la media poblacional en el estrato h, con toda validez podemos utilizar la expresión de la varianza de $\hat{\mu}_Y$ pero con el índice h. (h = 1,2,....,H).

Finalmente

$$V(\hat{\mu}_Y^s) = \frac{1}{N^2}\sum_h^H \frac{N_h}{n_h}(1-f_h)\left[S_{Yh}^2 + R_h^2 S_{xh} - 2R_h \rho_h S_{Yh} S_{xh}\right]$$

La fórmula anterior únicamente es válida si el tamaño de muestra en cada estrato es suficientemente grande de tal manera que se pueda aplicar la fórmula de la varianza en cada estrato.

b) Estimador combinado.

La expresión para el estimador combinado es :

$$\hat{\mu}_Y^c = r_c \mu_x = \frac{\hat{m}_Y}{\hat{m}_x}\mu_x$$

Donde

$$\hat{m}_Y = \frac{1}{N}\sum_{h=1}^H N_h m_Y^{(h)} \qquad \hat{m}_x = \frac{1}{N}\sum_{h=1}^H N_h m_x^{(h)}$$

Como habíamos mencionado, este estimador no requiere del conocimiento de las $\mu_x^{(h)}$, sino solamente de μ_x .

La varianza se puede calcular como sigue :

Por definición : $V(\hat{\mu}_Y^c) = E(\hat{\mu}_Y^c - \mu_Y)^2$

Si se considera la expresión dentro del paréntesis, se puede encontrar otra expresión equivalente

$$\begin{aligned}\hat{\mu}_Y^c - \mu_Y &= r_c \mu_x - \mu_Y \\ &= \mu_x (r_c - R) \\ &= \frac{\mu_x}{\hat{m}_x} (\hat{m}_Y - \hat{m}_x R) \\ &\doteq (\hat{m}_Y - \hat{m}_x R) \\ &= \frac{1}{N} \sum_{h=1}^{H} N_h (m_Y^{(h)} - m_x^{(h)} R)\end{aligned}$$

Al elevar esto al cuadrado y al tomar esperanza (el segundo término desaparece), finalmente tenemos la siguiente expresión para la varianza del estimador combinado.

$$V(\hat{\mu}_Y^c) = \frac{1}{N^2} \sum_{h=1}^{H} \frac{N_h}{n_h} (1 - f_h) \left[S_{Yh}^2 + R^2 S_{xh}^2 - 2 R \rho_h S_{Yh} S_{xh} \right]$$

El estimador de razón separado tendrá menor varianza que el combinado si las razones muestrales en el estrato h-ésimo, varían en una forma considerable de estrato a estrato, si se conocen el tamaño de la población en cada estrato y si el número de estratos es grande. Por otra parte el sesgo de cada $r^{(h)}$ puede ser grande comparado con su varianza si el tamaño de muestra dentro de los estratos es reducido, al sumarse estos sesgos pueden formar un sesgo grande. Si además a las anteriores razones agregamos la de disponer de buenos valores N_h/N y de que las razones muestrales en cada estrato no son muy variables en casi todos los casos, esto hace que el estimador combinado sea más utilizado en la práctica.

1.3) Estimador de diferencia simple.

Teóricamente, aunque este tipo de estimador es más sencillo que el de razón, en ciertas situaciones tiene menor varianza que éste.

La forma de este estimador es como sigue :

$$\tilde{\mu}_y = m_y + k(\mu_x - m_x)$$

Donde m_y, m_x, $_x$ y están definidas como en la sección 1.2 k es una constante.

A este estimador se le denomina de diferencia y es un estimador insesgado y su varianza está dada como sigue :

$$V(\tilde{\mu}_y) = \frac{(1-f)}{n}\left[S_y^2 + k^2 S_x^2 - 2\rho k S_y S_x\right]$$

La cual puede obtenerse de la siguiente forma :

Por definición $V(\tilde{\mu}_y) = E(\tilde{\mu}_y - \mu_y)^2$

Si elevamos al cuadrado y desarrollamos $(\tilde{\mu}_y - \mu_y)$, tenemos

$$(\tilde{\mu}_y - \mu_y)^2 = (m_y - \mu_y)^2 - 2k(m_y - \mu_y)(m_x - \mu_x) + k^2(m_x - \mu_x)^2$$

Al tomar esperanza

$$V(m_y) + k^2 V(m_x) - 2k \operatorname{cov}(m_y, m_x)$$

Por lo tanto

$$V(\tilde{\mu}_y) = \frac{(1-f)}{n}\left[S_y^2 + k^2 S_x^2 - 2\rho k S_y S_x\right]$$

El caso que nos interesa más es cuando $k = 1$, es decir

$$\tilde{\mu}_y = m_y + (\mu_x - m_x)$$

Esto sucede cuando tenemos un dato censal anterior el cual puede ser ajustado con la diferencia de cambio $(m_y - m_x)$ de una muestra, en fórmula esto equivale a :

$$\tilde{\mu}_y = \mu_x + (m_y - m_x)$$

La varianza en este caso estará representada por :

$$V(\tilde{\mu}_y) = \frac{(1-f)}{n}\left[S_y^2 + S_x^2 - 2\rho S_y S_x\right]$$

1.4) Estimadores de diferencia simple en muestreo estratificado.

Al igual que para el estimador de razón tenemos los dos estimadores; el separado y el combinado, los cuales fueron descritos en la sección 1.2.

a) Estimador separado

Su forma es la siguiente :

$$\bar{\mu}_y^s = \frac{1}{N}\sum_{h=1}^{H} N_h \bar{\mu}_y^{(h)} = \frac{1}{N}\sum_{h=1}^{H} N_h \left[m_y^{(h)} + k_h \left(\mu_x^{(h)} - m_x^{(h)}\right)\right]$$

Si tomamos la esperanza de $\bar{\mu}_y^s$ tenemos

$$E(\bar{\mu}_y^s) = \frac{1}{N}\sum_{h=1}^{H} N_h E(\bar{\mu}_y^{(h)}) = \mu_y$$

Lo que demuestra que es un estimador insesgado.

Su varianza está dada por :

$$V(\bar{\mu}_y^s) = \frac{1}{N^2}\sum_{h=1}^{H} \frac{N_h}{n_h}(1-f_h)\left[S_{yh}^2 + k_h^2 S_{xh}^2 - 2k_h\rho_h S_{yh}S_{xh}\right]$$

b) Estimador combinado.

Su forma es como sigue :

$$\bar{\mu}_y^c = \hat{m}_y + k(\mu_x - \hat{m}_x)$$

Donde $\hat{m}_y^c = \frac{1}{N}\sum_{h=1}^{H} N_h m_y^{(h)}$ $\quad \hat{m}_x = \frac{1}{N}\sum^{H} N_h m_x^{(h)}$

Además su varianza está dada por :

$$V(\bar{\mu}_y^c) = \frac{1}{N^2}\sum_{h=1}^{H} \frac{N_h^2}{n_h}(1-f_h)\left[S_{yh}^2 + k^2 S_{xh}^2 - 2k\rho_h S_{yh}S_{xh}\right]$$

1.5) Comparación de los métodos de razón y diferencia con el muestreo irrestricto aleatorio.

Para poder efectuar comparaciones es necesario suponer que el tamaño de muestra es suficientemente grande de modo que la aproximación de la varianza sea válida

Sea :

$$V(m_Y) = \frac{(1-f)}{n} \mu_Y^2 C_Y^2 = \frac{(1-f)}{n} S_Y^2$$

La varianza del estimador simple.

De acuerdo a lo anterior (varianza de razón)

$$V(\hat{\mu}_Y) = \frac{(1-f)}{n} (C_Y^2 + C_X^2 - 2\rho C_Y C_X)$$

Si consideramos el cociente

$$V(\hat{\mu}_Y) / V(m_Y)$$

Tenemos lo siguiente :

$$1 + \frac{C_X^2}{C_Y^2} - 2\rho \frac{C_X}{C_Y} \qquad (1.6.1)$$

Se puede observar que cuando :

$$\frac{C_X^2}{C_Y^2} < 2\rho \frac{C_X}{C_Y}$$

Entonces $\rho > \frac{C_X}{2C_Y}$; lo cual implica que

$$V(\hat{\mu}_Y) < V(m_Y)$$

Para comparar el estimador de diferencia con el irrestricto aleatorio procedemos de una forma análogo al de razón.

De la sección 1.3 sabemos que la varianza del estimador de diferencia es :

$$V(\breve{\mu}_y) = \frac{(1-f)}{n} \left(S_y^2 + k^2 S_x^2 - 2 \rho \, k \, S_x S_y \right)$$

Consideremos el cociente

$$V(\breve{\mu}_y) \, / \, V(m_y)$$

Se tiene que es igual a :

$$1 + k^2 \frac{S_x^2}{S_y^2} - 2 \rho \frac{S_x}{S_y}$$

Si $k^2 \dfrac{S_x^2}{S_y^2} < 2 k \rho \, S_x S_y$; entonces

$$\rho > \frac{k}{2} \, \frac{S_x}{S_y}$$

Lo cual implica que el estimador de diferencia tiene menor varianza que irrestricto aleatorio, si se cumple la condición anterior.

CAPITULO II

ESTIMADORES DE RAZON MULTIPLE.

2.1) Descripción general.

En este capítulo se hará una generalización del método de razón simple, mediante el uso de más de una variable auxiliar para estimar la media μ_0 *.

Las variables que intervienen son : la variable a estimar y y las variables auxiliares x_1 , x_2,....,x_p. Como se demostrará posteriormente mientras más alta sea la correlación entre las variables y y x para toda i = 1,2,....p, este método produce mejores resultados. La notación usada aquí es muy similar al del capítulo anterior.

Denotemos las medias poblacionales por :

$$\mu_1, \mu_2, \ldots\ldots\ldots, \mu_p$$

---

* En este capítulo y en los subsecuentes el índice o se refiere a la variable y.

donde tenemos p medias conocidas y una media desconocida.

Por definición :

$$\mu_o = \frac{1}{N}\sum_{t=1}^{N} y_t \qquad\qquad \mu_i = \frac{1}{N}\sum^{N} x_{it}$$

Las razones poblacionales pueden ser denotadas por

$$R_1, R_2, \ldots\ldots, R_p$$

Donde $\quad R_i = \dfrac{\mu_o}{\mu_i} \quad (i=1,2,\ldots,p);\ \mu_i \neq 0$

La muestra aleatoria simple de tamaño n observada es como sigue :

$(y_t, x_{1t}, \ldots . x_{pt})$ para toda $t = 1, 2, \ldots, n$

y las medidas muestrales son : $m_o, m_1, \ldots, m_o$

Donde $\quad m_o = \dfrac{1}{n}\sum^{n} y_t \qquad m_i = \dfrac{1}{n}\sum^{n} x_{it}$

El estimador de razón múltiple es

$$\hat{\mu}_o = w_1 r_1 \mu_1 + \cdots\cdots + w_p r_p \mu_p$$

Que puede escribirse en una forma compacta como :

$$\hat{\mu}_o = \sum w_i r_i \mu_i$$

El vector de ponderación $w = (w_1, w_2, \ldots, w_p)$

propiedad $\sum w_i = 1$; $r_i = \frac{m_o}{m_i}$

para toda i = 1, 2, ..., p.

Como en el caso simple, la fórmula exacta de la varianza de $\hat{\mu}_o$ está dada por una serie infinita, de la cual se puede obtener una aproximación mediante el uso de términos 1/n, porque los términos mayores o iguales a $1/n^2$ son despreciables para efectos prácticos.

La aproximación de la varianza de μ_o es como sigue :

$$V(\hat{\mu}_o) = \frac{(1-f)}{n} \mu_o^2 \, wAw'$$

A es una matriz cuadrada de orden pxp, cada elemento de la matriz A = (a_{ij}) es definido de la siguiente manera :

C_o coeficiente de variación de la variable y

C_i " " " " " " x_i.

ρ_{ij} " " correlación entre las variables x_i y x_j.

ρ_{oi} " " " " " " y y x_i.

ρ_{oj} " " " " " " y y x_j.

(1-f) es el factor de corrección finita.

w es un factor de ponderación.

w' es el vector transpuesto

Posteriormente explicaremos como se obtuvo la fórmula de la varianza, así como también de que forma se obtiene el valor óptimo del vector de ponderación w de tal manera que la varianza de $\hat{\mu}_0$ sea un mínimo.

2.2) Esperanza y varianza del estimador.

Primeramente pondremos la esperanza y varianza del estimador como una función de la E (r_i) y Cov (r_i, r_j) respectivamente.

El estimador propuesto

$$\hat{\mu}_0 = \sum_{i=1}^{p} w_i r_i \mu_i$$

Se puede escribir como

$$\hat{\mu}_0 = \sum_{i=1}^{p} w_i r_i (\mu_0 / R_i)$$

Por lo tanto tenemos finalmente

$$\hat{\mu}_0 = \mu_0 \sum_{i=1}^{p} (w_i / R_i) r_i$$

Al tomar esperanza de la expresión anterior tenemos que:

$$E(\hat{\mu}_o) = \mu_o \sum_{i}^{p} (w_i/R_i)\, E(r_i) \qquad (2.2.1)$$

Por otra parte la varianza de $\hat{\mu}_o$ se puede escribir como

$$V(\hat{\mu}_o) = \mu_o^2 \sum_{i}^{p} (w_i w_j / R_i R_j)\, cov(r_i, r_j)$$

ya que por definición:

$$V(\hat{\mu}_o) = E(\hat{\mu}_o - E(\hat{\mu}_o))^2$$

La expresión dentro del paréntesis rectangular se puede poner en la forma:

$$\begin{aligned} \mu_o - E(\hat{\mu}_o) &= \sum_{i}^{p} w_i r_i \mu_i - \sum_{i}^{p} w_i \mu_i E(r_i) \\ &= \sum_{i}^{p} w_i \mu_i [r_i - E(r_i)] \end{aligned}$$

Si llamamos $z_i = w_i \mu_i [r_i - E(r_i)]$

Entonces $[\hat{\mu}_o - E(\hat{\mu}_o)]^2 = \left(\sum_{i=1}^{p} z_i\right)^2 = \sum_{i}^{p}\sum_{j}^{p} z_i z_j$

al tomar la esperanza de la expresión anterior,

$$E[\hat{\mu}_o - E(\hat{\mu}_o)]^2 = \sum_{i}^{p}\sum_{j}^{p} w_i w_j \mu_i \mu_j\, cov(r_i, r_j)$$

Como $\mu_i = \frac{\mu_o}{R_i}$ y $\mu_j = \frac{\mu_o}{R_j}$ finalmente tenemos que:

$$V(\hat{\mu}_o) = \mu_o^2 \sum_i^p \sum_j^p (w_i w_j / R_i R_j) \operatorname{cov}(r_i, r_j) \quad (2.2.2)$$

De lo anterior se puede observar que la esperanza y varianza del estimador están en función de E (r_i) y Cov (r_i, r_j) respectivamente.

Para el cálculo de la esperanza de $\hat{\mu}_o$ se puede utilizar un argumento semejante al del primer capítulo.

Por definición

$$r_i = \frac{m_o}{m_i}$$

Podemos escribir

$$\delta_i = \frac{m_i - \mu_i}{\mu_i} \qquad \epsilon_o = \frac{m_o - \mu_o}{\mu_o}$$

Para tal efecto necesitamos suponer que el valor de μ_i es tal que :

1) Las x_i son positivas y n es suficiente grande

2) El error relativo $|\delta_i|$ es suficientemente pequeño de tal manera que se puedan despreciar términos del desarrollo de Taylor.

De las expresiones anteriores tenemos que :

$$r_i = \frac{m_o}{m_i} = \frac{\mu_o}{\mu_i}\frac{(1+\epsilon_o)}{(1+\delta_i)} = R_i(1+\epsilon_o)(1+\delta_i)^{-1}$$

Si utilizamos el desarrollo de Taylor para el término y despreciamos términos de orden $1/n^2$ en adelante, se obtiene la siguiente aproximación :

$$r_i = R_i(1+\epsilon_o-\delta_i+\delta_i^2-\epsilon_o\delta_i)$$

Al tomar esperanza de la expresión anterior se obtiene

$$E(r_i) = R_i\left[1+\frac{(1-f)}{n}(C_i^2-\rho_{oi}C_oC_i)\right]$$

Si llamamos $b_i = C_i^2-\rho_{oi}C_oC_i$

Finalmente tenemos

$$E(\hat{\mu}_o) = \mu_o\sum w_i\left[1+\frac{(1-f)}{n}b_i\right]$$

lo cual lo podemos expresar en forma matricial

$$E(\hat{\mu}_o) = \mu_o\left[1+\frac{(1-f)}{n}wb'\right]$$

Donde b y w son vectores de un renglón por p columnas

Para encontrar la Cov (r_i, r_j), utilizaremos una aproximación de la manera siguiente :

Por definición $\operatorname{cov}(r_i, r_j) = E(r_ir_j)-E(r_i)E(r_j)$

Para encontrar el producto E (r_i) E (r_j) es necesario utilizar las siguientes aproximaciones:

$$E(r_i) = R_i \left[1 + \frac{(1-f)}{n} (C_i^2 - \rho_i\, C_o C_i) \right]$$

$$E(r_j) = R_j \left[1 + \frac{(1-f)}{n} (C_j^2 - \rho_j\, C_o C_j) \right]$$

Al efectuar operaciones obtenemos lo siguiente :

$$E(r_i r_j) = R_i R_j \left[1 + \frac{(1-f)}{n} (b_i + b_j) \right]$$

Ahora para encontrar el término E $(r_i\, r_j)$, primero efectuamos el producto entre r_i y r_j (usando las aproximaciones anteriores) y tomamos esperanza :

$$E(r_i r_j) = R_i R_j \left[1 + \frac{(1-f)}{n} (C_o^2 - \rho_{oi}\, C_o C_i) + \frac{(1-f)}{n} (C_o^2 - \rho_{oj}\, C_o C_j) - \frac{(1-f)}{n} (C_o^2 - \rho_{oi}\, C_o C_i - \rho_{oj}\, C_o C_j) \right]$$

Si llamamos

$$a_{ij} = (C_o^2 - \rho_{oi}\, C_o C_i - \rho_{oj} C_o C_j + \rho_{ij} C_i C_j)$$

Tenemos que

$$E(r_i r_j) = R_i R_j \left[1 + \frac{(1-f)}{n} (b_i + b_j) + \frac{(1-f)}{n} a_{ij} \right]$$

Si ahora substituimos (2.2.5) y (2.2.6) en (2.2.4), obtenemos una aproximación de la covarianza entre r_i y r_j:

$$cov(r_i, r_j) = R_i R_j \frac{(1-f)}{n} a_{ij}$$

La cual la podemos reemplazar en (2.2.2), quedándonos lo siguiente :

$$V(\hat{\mu}_o) = \frac{(1-f)}{n} \mu_o^2 \sum_{i=1}^{p} \sum_{j}^{p} w_i w_j a_{ij}$$

Finalmente tenemos una expresión matricial para la varianza de $\hat{\mu}_o$

$$V(\hat{\mu}_o) = \frac{(1-f)}{n} \mu_o^2 \, wAw' \qquad (2.2.7)$$

Donde $A = (a_{ij})$ es una matriz simétrica de orden $p \times p$.

w es un vector renglón de $1 \times p$.

Otra forma de escribir la matriz A es como el producto matricial TCT', donde C y T son definidas de acuerdo a lo siguiente :

$C = (C_{ij})$ es una matriz de varianza-covarianza relativa de orden $(p+1) \times (p+1)$

$$C_{ij} = \frac{S_{ij}}{\mu_i \mu_j}$$

Por ejemplo S_{oo} es la varianza de $\underline{y}$, S_{oi} es la covarianza entre $\underline{y}$ y x_i.

Vamos a suponer que C es positiva definida, por otra parte definimos T como la matriz de orden p x (p + 1), la cual es de la siguiente forma :

$$T = \begin{bmatrix} 1 & -1 & 0 & \dots 0 \\ 1 & 0 & -1 & \dots 0 \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ 1 & 0 & \cdot & \dots -1 \end{bmatrix}$$

T' es la matriz transpuesta de T.

Como C es positiva definida y T es de rango p se concluye que A es positiva definida. Entonces se puede usar ésta otra forma alternativa de A.

2.3) Factores de Ponderación.

En esta sección vamos a determinar el valor óptimo de w que hace mínima la varianza de $\hat{\mu}_o$ con la condición de que $\sum_{i=1}^{k} w_i = 1$

Para encontrar el extremo restringido (máximo ó mínimo) de función de varias variable, se puede proceder de la siguiente manera : podemos hacer uso de la desigualdad generalizada de Cauchy ó bien mediante el empleo de multiplicadores de Lagrange. En esta sección vamos a usar este último, el cual se puede enunciar como sigue :

Para encontrar el máximo ó mínimo de una función de p variables $f(x_1, x_2, \ldots, x_p)$ condicionada por un sistema de ecuaciones de la forma

$$g_i(x_1, x_2, \ldots x_p) = 0$$
$$i = 1, 2, \ldots, r \quad (r < p)$$

Se puede construir la siguiente función

$$\Psi = f(x_1, x_2, \ldots, x_p) + \sum_{i=1}^{r} \lambda_i g_i$$

Si derivamos parcialmente Ψ con respecto a $x_1, x_2, \ldots x_p$ e igualamos a cero, obtenemos un sistema de p ecuaciones y si consideramos lar r condiciones para encontrar los valores de $p + r$ incógnitas de este sistema, llegamos a lo que se desea

INCOGNITAS

$$x_1, x_2, \ldots, x_p, \lambda_1, \lambda_2 \ldots, \lambda_r$$

En forma matricial nuestro problema lo podemos plantear de la siguiente manera

$$\text{Min } V(\hat{\mu}_0) = \frac{\mu_0^2}{n} wAw'$$

Sujeta a la condición

$$\sum w_i = 1 = we'$$

Sea

$$\psi = V(\hat{\mu}_0) + \lambda(we'-1)$$

Si tomamos la derivada de ψ con respecto a w tenemos:

$$\frac{\partial \psi}{\partial w} = 2\frac{\mu_0^2}{n} wAw'$$

Si lo anterior lo igualamos a cero y despejamos a w.

$$w = -\frac{\lambda e A^{-1}}{(2\mu_0^2/n)} \qquad (2.3.1)$$

De acuerdo a la restricción

$$we' = 1$$

Al multiplicar la expresión (2.3.1) por e' obtenemos

$$we' = \frac{\lambda(eA^{-1}e')}{(2\mu_0^2/n)} = 1$$

De donde al despejar

$$\lambda = -\frac{(2\mu_0^2/n)}{eA^{-1}e'}$$

y si substituímos esto en (2.3.1) obtenemos la expresión requerida para w.

$$w_{opt} = \frac{eA^{-1}}{eA^{-1}e'}$$

El cual hace mínima la varianza de $\hat{\mu}_0$

Si reemplazamos esta expresión para en (2.2.3) y (2.2.7) llegamos al siguiente resultado.

$$E(\hat{\mu}_0) = \mu_0 \left(1 + \frac{1}{n} \frac{e A^{-1} b'}{e A^{-1} e'}\right) \qquad (2.2.3)$$

$$V(\hat{\mu}_0) = \frac{\mu_0^2}{n} \cdot \frac{1}{e A^{-1} e'} \qquad (2.2.4)$$

De (2.2.3) se observa fácilmente que el sesgo es eliminado si y sólo si $e A^{-1} b' = 0$ lo cual implica que b = 0.

Es decir $C_i = \rho_{0i} C_0$

Esto ocurre cuando la regresión es tomada individualmente a través del origen. Hay que hacer notar que los pesos serán uniformes sí y sólo si las sumas de las columnas de A son iguales.

Un ejemplo en donde los pesos son uniformes es en el caso en que:

$$C_0 = C_1 = C_2 = \cdots = C_p$$

$$\rho_{0i} = \rho_{0j} = \rho_{ij} = \rho$$

La varianza en este caso esta dada por

$$V(\hat{\mu}_o) = \frac{(1-f)}{n} \mu_o^2 C^2 (1-\rho) \frac{(p+1)}{p}$$

2.4) Estimador de la varianza a partir de la muestra.

En el capítulo anterior (estimador simple) vimos que

$$V(\hat{\mu}_y) = \frac{1}{N} \sum_{i=1}^{N} (y_i - Rx_i)^2 / (N-1)$$

y su varianza a partir de la muestra es

$$v(\hat{\mu}_y) = \frac{1}{n} \sum_{i=1}^{n} (y_i - rx_i)^2 / n-1$$

En el caso multivariado es conveniente escribir la matriz A en una forma similar a la expresión de $V(\hat{\mu}_y)$ e ignorar el factor de corrección finita (1-f).

$$a_{ij} = C_o^2 - \rho_{oi} C_o C_i - \rho_{oj} C_o C_j + \rho_{ij} C_i C_j = \frac{S_{oo}}{\mu_o^2} - \frac{S_{oi}}{\mu_o \mu_i} - \frac{S_{oj}}{\mu_o \mu_j} + \frac{S_{ij}}{\mu_i \mu_j}$$

$$= \frac{1}{(N-1)\mu_o^2} \Big[\big(\sum_t y_t^2 - N\mu_o^2\big) - R_i \big(\sum_t y_t x_{it} - N\mu_o \mu_i\big) - R_j \big(\sum_t y_t x_{jt} - N\mu_o \mu_j\big)$$

$$+ R_i R_j \big(\sum_t x_{it} x_{jt} - N\mu_i \mu_j\big) \Big]$$

$$= \frac{1}{\mu_o^2} \frac{\sum_t (y_t - R_i x_{it})(y_t - R_j x_{jt})}{N-1}$$

Por lo tanto $\hat{a}_{ij} = a_{ij} \sum_{t}^{n} \frac{(y_t - r_i x_{it})(y_t - r_j x_{jt})}{n-1} = \mu_\circ^2 a_{ij}$

De aquí que la $V(\hat{\mu}_*)$ puede ser estimado por

$$v(\hat{\mu}_\circ) = \frac{1}{n} \cdot \frac{1}{e \hat{A}^{-1} e'}$$

2.5) Efecto al aumentar el número de variables.

El objetivo principal de esta sección es analizar que pasa cuando utilizamos dos conjuntos de variables auxiliares del siguiente tipo :

$x_1, x_2, \ldots, x_p$ y $x_1, x_2, \ldots, x_p, \ldots, x_s$

Donde $(s > p)$

Para tal efecto consideremos las varianzas al utilizar p y s variables (con afijación óptima)

Sean :

$$V(\hat{\mu}_\circ / p) = \frac{\mu_\circ^2}{n} \cdot \frac{1}{e A_p^{-1} e'}$$

$$V(\hat{\mu}_\circ / s) = \frac{\mu_\circ^2}{n} \cdot \frac{1}{e A_s^{-1} e'}$$

La matriz A_s la podemos expresar como sigue :

$$A_s = \begin{bmatrix} A_p & \beta \\ \beta' & \gamma \end{bmatrix}$$

Además por medio de particiones tenemos A_s^{-1}

$$A_s^{-1} = \begin{bmatrix} A_p^{-1} + \delta\eta\delta' & -\delta\eta \\ -\eta\delta' & \eta \end{bmatrix}$$

Donde : $\eta = (\gamma - \beta A_p^{-1} \beta')^{-1}$ $\qquad \delta = A_p^{-1}\beta'$

Considerese como es la diferencia.

$$V(\hat{\mu}_0/s) - V(\hat{\mu}_0/p)$$

Con respecto a cero (< = >)

Supongamos que :

$$V(\hat{\mu}_0/s) - V(\hat{\mu}_0/p) \leq 0$$

Es decir

$$\frac{1}{\eta} \cdot \frac{1}{e A_s^{-1} e'} - \frac{1}{n} \cdot \frac{1}{e A_p^{-1} e'} \leq 0$$

Esto es equivalente

$$\frac{1}{e A_s^{-1} e'} \leq \frac{1}{e A_p^{-1} e'}$$

De aquí que :

$$e A_s^{-1} e' \leq e A_p^{-1} e'$$

Por lo tanto

$$e\, A_s^{-1} e' - e\, A_p^{-1} e' \geq 0$$

$e\, A_s^{-1} e' - e\, A_s^{-1} e'$ lo podemos expresar como

$$e \begin{pmatrix} S \\ -I \end{pmatrix} \eta \,(S' - I)\, e'$$

Lo anterior es mayor o igual a cero debido a que es positiva definida de tal manera que $x \eta x' > 0$. Por lo que nuestra suposición se confirma.

2.6) Estimadores de razón múltiple en muestreo estratificado.

a) Estimador separado.

Sea $\hat{\mu}_0^{(h)} = w_1^{(h)} r_1^{(h)} \mu_1^{(h)} + \ldots\ldots + w_p^{(h)} r_p^{(h)} \mu_p^{(h)}$

$h = 1, 2, \ldots, H$

Donde

$\mu_0^{(h)}$ estimador de razón múltiple de $\mu_0^{(h)}$

Por lo tanto el estimador separado está dado por :

$$\hat{\mu}_0^{s} = \frac{N_1}{N} \hat{\mu}_0^{(1)} + \frac{N_2}{N} \hat{\mu}_0^{(2)} + \cdots + \frac{N_p}{N} \hat{\mu}_0^{(H)}$$

$$= \frac{1}{N} \sum_{h=1}^{H} N_h \hat{\mu}_0^{(h)}$$

La expresión para la varianza es :

$$V(\hat{\mu}_0^s) = \frac{1}{N^2}\sum_{h=1}^{H} N_h^2\, V(\hat{\mu}_0^{(h)})$$

b) Estimador combinado

El estimador es como sigue :

$$\hat{\mu}_0^c = l_1 \frac{\hat{m}_0}{\hat{m}_1}\mu_1 + \cdots + l_p \frac{\hat{m}_0}{\hat{m}_p}\mu_p \qquad \sum_{i=1}^{p} l_i = 1$$

En este caso l_j son los factores de ponderación.

Se puede demostrar que la varianza de $\hat{\mu}_0^c$ es como sigue :

$$V(\hat{\mu}_0^c) = \frac{(1-f)}{n}\sum_{i=1}^{p}\sum_{j=1}^{p} l_i l_j a_{ij}$$

Donde:

$$l = \frac{e A^{-1}}{e A^{-1} e'}$$

Este problema se ha reducido a la estructura del problema original de tal manera que la teoría anterior permanece válida.

NOTA.- $\frac{\hat{m}_0}{\hat{m}_i} = x_i \quad \forall i = 1, 2, \ldots, p.$

CAPITULO III

ESTIMADORES DE DIFERENCIA MULTIPLE

3.1) Descripción General.

En este capítulo se describirá lo que se ha llamado estimador de diferencia múltiple, el cual es una generalización del estimador de diferencia simple.

Tal vez la aplicación más simple de este estimador es en estimaciones intercensales, en que las variables auxiliares son datos de censos anteriores, razón por la cual representa la parte principal de la presente tesis.

El estimador general que se propone es de la siguiente forma :

$$\tilde{\mu}_0 = \sum u_i \left[m_0 + k_i (\mu_i - m_i) \right] \qquad (3.1.1)$$

Con la condición de que : $\sum_{i=1}^{p} u_i = 1$

m_0, m_i y μ_i están definidos como en el capítulo anterior, u en este caso es el vector de ponderación para el estimador y k_i es una constante para cada variable x_i .

La expresión general para la varianza es :

$$V(\tilde{\mu}_0) = \frac{(1-f)}{n} \sum_{i=1}^{p} \sum_{j=1}^{p} u_i u_j z_{ij}$$

Donde :

$$z_{ij} = S_0^2 - k_i S_{0i} - k_j S_{0j} + k_i k_j S_{ij}$$

S_o Desviación estandar unitaria de la variable y.

S_{oi} Covarianza unitaria entre la variable y y x_i.

S_{ij} " " " " " x_i y x_j

u vector renglón de 1 x p.

k_i es una constante (i = 1, 2. . . ., p).

En forma matricial se puede expresar como

$$V(\tilde{\mu}_o) = \frac{(1-f)}{n} u Z u'$$

Cuando las variables auxiliares son datos de censos anteriores

$$z_{ij} = S_o^2 - S_{oi} - S_{oj} + S_{ij}$$

En el desarrollo de este capítulo encontraremos la función de ponderación y mostraremos el estimador separado y combinado en muestreo estratificado.

3.2) Esperanza y varianza del estimador.

Sin pérdida de generalidad podemos encontrar la esperanza y varianza de $\tilde{\mu}_o$ considerando que $k_i = 1$, sí x_i (i = j . . . p) es el valor de la característica y en un tiempo anterior y la naturaleza de las variables no ha sufrido un cambio significativo durante ese tiempo, el valor de k_i será igual a la unidad.

Por lo tanto

$$\tilde{\mu}_0 = \sum_{i=1}^{p} u_i \left[m_0 + (\mu_i - m_i) \right]$$

De lo siguiente se ve que es un estimador in sesgado.

$$E(\tilde{\mu}_0) = \sum_{i=1}^{p} u_i \left[E(m_0) + (\mu_i - E(m_i)) \right] = \mu_0$$

La varianza del estimador la podemos deducir como sigue :

Por definición

$$V(\tilde{\mu}_0) = E(\tilde{\mu}_0 - \mu_0)^2 \qquad (3.2.1)$$

Si consideramos la expresión en el paréntesis y elevamos al cuadrado.

$$(\tilde{\mu}_0 - \mu_0)^2 = \left\{ \sum_{i=1}^{p} u_i \left[m_0 + (\mu_i - m_i) \right] - \mu_0 \right\}^2$$

Si $d_i = m_0 + (\mu_i - m_i)$

Entonces

$$(\tilde{\mu} - \mu_0)^2 = \left(\sum_{i=1}^{p} u_i d_i - \mu_0 \right)^2 = \left(\sum u_i d_i \right)^2 - 2\mu_0 \sum_{i=1}^{p} u_i d_i + \mu_0^2 \qquad (3.2.2)$$

El primer término lo podemos expresar como sigue :

$$\left(\sum_{i=1}^{p} u_i d_i \right)^2 = \sum_{i=1}^{p} u_i^2 d_i^2 + \sum_{i \neq j}^{p} u_i u_j d_i d_j$$

Por lo tanto (3.2.2) es como sigue :

$$(\tilde{\mu}_0 - \mu_0)^2 = \sum_{i=1}^{p} u_i^2 d_i^2 + \sum_{i \neq j} u_i u_j d_i d_j - 2\mu_0 \sum u_i d_i + \mu_0^2$$

Al tomar la experanza de la expresión anterior tenemos que :

$$E(\tilde{\mu}_0-\mu_0)^2=\sum_{i=1}^{p} u_i^2 E(d_i^2)+\sum_{i\neq j}^{p} u_i u_j E(d_i d_j) - 2\mu_0 \sum u_i E(d_i) + \mu_0^2 \qquad (3.2.3)$$

Se puede demostrar fácilmente que :

i) $E(d_i)=\mu_0$

ii) $E(d_i^2)=\frac{(1-f)}{n}\left(S_0^2-2S_{0i}+S_i^2\right)+\mu_0^2$

iii) $E(d_i d_j)=\frac{(1-f)}{n}\left(S_0^2-S_{0i}-S_{0j}+S_{ij}\right)$

Si substituímos i), ii), iii) en (3.2.3) y efectuamos operaciones tenemos

$$V(\tilde{\mu}_0)=\frac{(1-f)}{n}\Big[\sum u_i^2\left(S_0^2-2S_{0i}+S_i^2\right) + \sum_{i\neq j} u_i u_j \left(S_0^2-S_{0i}-S_{0j}+S_{ij}\right) \qquad (3.2.4)$$

$$=\frac{(1-f)}{n}\Big[\sum_i^p\sum_j^p u_i u_j\left(S_0^2-S_{0i}-S_{0j}+S_{ij}\right)\Big] \qquad (3.2.5)$$

Si llamamos

$$z_{ij} = S_o^2 - S_{oi} - S_{oj} \quad S_{ij}$$

Por lo tanto

$$V(\bar{\mu}_o) = \frac{(1-f)}{n} \sum_{i}^{p} \sum_{j}^{p} u_{ij} \, z_{ij} \qquad (3.2.6)$$

Finalmente esto lo podemos expresar en forma matricial

$$V(\bar{\mu}_o) = \frac{(1-f)}{n} u \, Z \, u' \qquad (3.2.7)$$

3.3) Factores de Ponderación.

Para encontrar el vector de ponderación $\underline{u}$ que minimiza $V(\bar{\mu}_o)$ vamos a utilizar la desigualdad general de Cauchy,

$$(xy') \leq (xMx')\ (yMy')$$

Donde M es una matriz positiva definida y simétrica, la igualdad se sostiene sí y sólo si $x\,M = \theta$ y donde $\theta \neq 0$ (escalar)

Sea $e = (1, 1\ldots, 1)$ un vector suma de $1 \times p$.

Si hacemos corresponder $x = u$, $y = e$ y $M = Z$

$$1 = (ue')^2 \leq (u Z u')(e Z^{-1} e')$$

La igualdad, como se puede ver, se sostiene si y sólo

si $\quad uZ = \theta e \quad ; \quad u = \theta e Z^{-1}$

Pero por la restricción $ue' = 1$ se sigue que

$$\theta = \frac{1}{e Z^{-1} e'}$$

De aquí que la óptima u es

$$u = \frac{e Z^{-1}}{e Z^{-1} e'}$$

En consecuencia la varianza está dada por :

$$V(\tilde{\mu}_o) = \frac{(1-f)}{n} \frac{1}{e Z^{-1} e'}$$

Un caso particular ocurre cuando los factores de ponderación son uniformes, por ejemplo $u_i = 1/p$ para $i = 1, \ldots, p$.

Lo anterior se logra haciendo la suposición de que :

$$S = S_o = S_i = S_j \quad \text{y} \quad \rho = \rho_{oi} = \rho_{oj} = \rho_{ij}$$

Si substituímos esto en la fórmula de la varianza (3.2.4) y al efectuar operaciones

$$V(\tilde{\mu}_o) = \frac{(1-f)}{n} S^2 (1-\rho) \left[2 \sum u_i^2 + \sum_{i \neq j}^{p} u_i u_j \right] \quad (3.3.1)$$

Se puede demostrar que

$$\sum_{i \neq j}^{p} u_i u_j = \frac{p-1}{p}$$

Por otra parte

$$1 = \left(\sum u_i \right)^2 = \sum_{i=1}^{p} u_i^2 - \sum_{i=1}^{p} u_i u_j$$

Por lo tanto

$$\sum_{i=1}^{p} u_i^2 = 1 - \frac{p-1}{p} = \frac{1}{p}$$

Al substituir (3.2.3) y (3.3.3) en (3.3.1) y efectuando operaciones, se llega finalmente a :

$$V(\tilde{\mu}_o) = \frac{(1-f)}{n} S^2 (1-\rho) \frac{(p+1)}{p}$$

3.4) Estimador de la varianza a partir de la muestra.

Si consideramos primero el caso en que utilizamos la constante k_i tenemos :

$$z_{ij} = S_o^2 - k_i\, S_{oi} - k_j\, S_{oj} + k_i\, k_j\, S_{ij}$$

$$= \frac{1}{N-1}\left[\sum_{t=1}^{N} y_t^2 - N\mu_o^2\right] - \frac{k_i}{N-1}\left[\sum y_t\, x_{it} - N\mu_o\mu_i\right] - \frac{k_j}{N-1}\left[\sum y_t\, x_{jt} - N\mu_o\mu_j\right] + \frac{k_i k_j}{N-1}\left[\sum_{t=1}^{N} x_{it}\, x_{jt} - N\mu_i\mu_j\right]$$

Si agrupamos tenemos que :

$$z_{ij} = \frac{\sum_{t}^{N}(y_t - k_i x_{it})(y_t - k_j x_{jt})}{N-1}$$

Lo cual se puede estimar con :

$$\frac{\sum_{t}^{n}(y_t - k_i x_{it})(y_t - k_j x_{jt})}{n-1}$$

Por lo tanto esto permite estimar la varianza como :

$$v(\tilde{\mu}_o) = \frac{(1-f)}{n}\,\frac{1}{e\,\tilde{Z}^{-1}e'}$$

3.5) Estimadores de diferencia múltiple en muestreo estratificado.

a) Estimador separado.

Su forma es la siguiente :

$$\tilde{\mu}_o^{s} = \frac{N_1}{N}\,\tilde{\mu}_o^{(1)} + \cdots\cdots + \frac{N_H}{N}\,\tilde{\mu}_o^{(H)}$$

Donde :

$$\mu_0^{(h)} = u_1^{(h)}\left[m_0^{(h)} + (\mu_1^{(h)} - \hat{m}_1^{(h)})\right] + \cdots\cdots + u_p^{(h)}\left[m_0^{(h)} + (\mu_p^{(h)} - m_p^{(h)})\right]$$

h = 1, 2, ... , H

La varianza del estimador se puede expresar como sigue :

$$V(\breve{\mu}_0^{s}) = \frac{1}{N^2} \sum_{h=1}^{H} N_h^2 V(\breve{\mu}_0^{(h)})$$

Para $V(\breve{\mu}_0^{(h)})$ con toda validez se usa $V(\breve{\mu}_0)$ pero con el índice h.

b) Estimador combinado.

$$\breve{\mu}_0^{c} = \sum_{i=1}^{p} u_i\left[\hat{m}_0 + (\mu_i - \hat{m}_i)\right]$$

Donde $\hat{m}_0$ y $\hat{m}_i$ están definidos como en las secciones anteriores.

Su varianza esta dada por :

$$V(\breve{\mu}_0^{c}) = \frac{1}{N^2} \sum_{h=1}^{H} \frac{N_h^2}{n_h}(1-f_h)\left[S_0^{2(h)} - \sum_{i=1}^{p} u_i S_{0i}^{(h)} - \sum_{j=1}^{p} u_j S_{0j}^{(h)} + \sum_{i \neq j} u_i u_j S_{ij}^{(h)}\right]$$

CAPITULO IV

COMPARACION DE LOS METODOS DE RAZON Y DIFERENCIA

Las varianzas de los estimadores de razón y diferencia simple son :

$$V(\hat{\mu}_y) = \frac{1}{n}(S_y^2 + R^2 S_x^2 - 2 R S_{yx})$$
$$V(\breve{\mu}_y) = \frac{1}{n}(S_y^2 + S_x^2 - 2 S_{yx})$$

Consideremos la diferencia

$$V(\breve{\mu}_y) - V(\hat{\mu}_y) = \frac{1}{n}(S_x^2 - 2 S_{xy} - R^2 S_x^2 + 2 R S_{xy})$$

Nos interesa conocer las condiciones bajo las cuales, el estimador de diferencia tiene menor varianza que el de razón; esto ocurre cuando

$$S_x^2 - 2 S_{xy} - R^2 S_x^2 + 2 R S_{xy} < 0$$

O sea

$$S_x^{\,2}\,(1-R^2)^2 < 2\;S_{xy}\cdot(1-R) \qquad (5.1.1)$$

Es conveniente analizar cada uno de los tres casos siguientes :

1) $R < 1$

2) $R = 1$

3) $R > 1$

En el primer caso $R = 1$; y esto implica que $(1-R^2)$ es positivo; por lo tanto (5.1.1) se puede expresar como :

$$\frac{(1-R^2)}{(1-R)} < \frac{2\,S_{YX}}{S_x^2}$$

$$\frac{(1-R)(1+R)}{(1-R)} < 2\,b$$

(b Coeficiente de regresión)

De aquí que

$$R < 2b - 1$$

En el segundo caso, si $R = 0$ es fácil ver que :

$$V(\hat{\mu}_Y) = V(\tilde{\mu}_Y)$$

Por último si $R > 1$ entonces $(1 - R^2)$ es negativo y al pasarlo al primer miembro en la expresión (5.1.1) cambia el signo de la desigualdad.

Por lo tanto

$$\frac{(1-R^2)}{(1-R)} > 2\frac{S_{yx}}{S_x^2} = 2b$$

Y la condición para que el estimador de diferencia será mejor que el de razón es que :

$$R > 2b - 1$$

De lo anterior podemos concluir que el estimador de diferencia tendrá menor varianza que el de razón si :

a.- $R < 1$ y además $R < 2b - 1$

b.- $R > 1$ y además $R > 2b - 1$

CAPITULO V

EJEMPLOS Y APLICACIONES

Las posibles aplicaciones de los estimadores de razón y de diferencia, múltiple que se proponen en la presente tesis se pueden encontrar en aquellas actividades en las que se llevan a cabo censos en forma períodica. Es así que la aplicación de dichos métodos es posible en las actividades industriales, agrícolas, comerciales y de población para las que se efectúan censos períodicamente con un desglose, nivel, ámbito y alcance determinados por los mismos censos.

En el sector privado las posibilidades son también bastante grandes, debido a que la mayoría de las empresas tienen registros históricos de su desarrollo que permitan la aplicación de éstas técnicas. Se puede mencionar la estimación de cuentas que afectan el régimen de encaje legal o depósito obligatorio en una Institución de crédito.

En todos estos campos existe una serie de problemas que se presentan a la mitad de los censos, cuya resolución podría llevarse a cabo mediante la aplicación de las técnicas anteriores, sin tener que efectuar un censo completo, a un costo menor y en muchas situaciones con mejores resultados.

Ejemplo 1.-

Este primer ejemplo, nos muestra numéricamente que el estimador de diferencia tiene menor varianza que el de razón, bajo condiciones mencionadas en el capítulo anterior.

Supongamos que la muestra obtenida de una cierta población junto con la variable auxiliar es como sigue :

x	y
1	26
2	32
3	38
4	44
5	50
15	190

$$R = 190 / 15 = 12.67 \quad (R > 1)$$

$$\text{Factor} = \frac{2 S_{yx}}{S_x^2} - 1 = 11$$

$R > 11$ lo cual debe implicar que :

$S_x^2 = 2 \qquad S_x = 1.414$

$S_y^2 = 72 \qquad S_y = 8.4852$

$S_{xy} = 12$

$$V(\hat{\mu}_y) = 1/5(72 + 2\,(12.7)^2 - 2\,(12.7)(12)) = \frac{88.97}{5} = \underline{17.79}$$

$$V(\tilde{\mu}_y) = 1/5(72 + 2 - 2 \times 12) = \underline{10}$$

Ejemplo 2.-

Como segundo ejemplo consideraremos el expuesto por Olkin (1968) quien solo utiliza el estimador de razón múltiple. Se presenta este ejemplo para tener una idea más clara de los diferentes resultados que se pueden obtener por medio del método de razón múltiple y el método de diferencia múltiple.

El problema consiste en estimar el número promedio de habitantes en 200 de las ciudades más grandes en los Estados Unidos, excluyendo las 5 más grandes.

Para tal efecto se utiliza la información censal de los años 1940 y 1930; en este caso las variables son $y = 1950$, $x_1 = 1940$, $x_2 = 1930$.

Se toma una muestra de $n = 50$ ciudades aleatoriamente con los siguientes resultados :

$m_0 = 1896$

$m_1 = 1693$ $\qquad r_1 = 1.1199$

$m_2 = 1643$ $\qquad r_2 = 1.1539$

Por otra parte las medias poblacionales μ_1 y μ_2 correspondientes a 1940 y 1930 son :

$\mu_1 = 1482$

$\mu_2 = 1420$

La matriz de varianza-covarianza unitaria S es :

$$\begin{bmatrix} 4360512 & 3983501 & 3912601 \\ & 3731856 & 3713435 \\ & & 3727939 \end{bmatrix} *$$

El estimador de razón múltiple es

$$\hat{\mu}_0 = W_1 \times 1.1199 \times 1482 + W_2 \times 1.539 \times 1420$$

---

* Los espacios en blanco indican que es una matriz simétrica.

Para determinar w necesitamos

$$w = \frac{e\hat{A}^{-1}}{e\hat{A}^{-1}e'}$$

Donde la matriz A se obtiene con la siguiente transformación :

$$a_{ij} = C_o^2 - C_{oi} - C_{oj} + C_{ij}$$

Por lo tanto

$$A = \begin{bmatrix} 033 & .051 \\ & .082 \end{bmatrix}$$

<u>Y</u>

$$A^{-1} = \begin{bmatrix} 773.68 & -481.28 \\ & 311.51 \end{bmatrix}$$

De aquí que

$$w = (2.38, -1.38)$$

Por lo que finalmente

$$\hat{\mu}_o = 2.38 \times 1.1199 \times 1482 - 1.38 \times 1.539 \times 1.420 = \underline{1689}$$

La varianza de la estimación es :

$$V(\hat{\mu}_o) = \frac{\left(1 - \frac{50}{200}\right)}{50} \; \frac{.(1689^2)}{122.33} = (18.61)^2$$

Nota.- $e\hat{A}^{-1}e' = 122.33$

El estimador de diferencia múltiple es :

$$\tilde{\mu}_o = u_1[m_o + (\mu_1 - m_1)] + u_2[m_o + (\mu_2 - m_2)]$$

Lo único que necesitamos determinar es U

Para tal efecto se utiliza

$$u = \frac{e\,Z^{-1}}{e\,Z^{-1}e'}$$

La matriz Z se puede encontrar mediante la transformación :

$$Z_{ij} = S_o^2 - S_{oi} - S_{oj} + S_{ij}$$

Por lo tanto

$$Z = \begin{bmatrix} 125326 & 177825 \\ & 263249 \end{bmatrix} \qquad Z^{-1} = 10^{-4}\begin{bmatrix} 19 & -12 \\ & 9 \end{bmatrix}$$

De aquí que

$$U = (1.75, -0.75)$$

Y por lo tanto

$$\tilde{\mu}_o = 1.75\left[1896 + (1482 - 1693)\right] - 0.75$$
$$\left[1896 + (1420 - 1643)\right] = 1694$$

Su varianza está dada por :

$$V(\tilde{\mu}_o) = \frac{(1-50/200)}{50} \cdot \frac{1}{4 \times 10^{-5}} = (19.3)^2$$

Nota. $e\hat{\Sigma}e' = 4 \times 10^{-5}$

De aquí que el error estandar es 19.63

El valor real resultó ser de 1699 y se puede calcular el error relativo real en las dos estimaciones, el estimador de razón múltiple tuvo un error relativo total de

$$\frac{1699-1689}{1699} = .0059 = 0.59$$

Y para el estimador de diferencia múltiple el error fué de :

$$\frac{1699-1694}{1699} = .0029 = 0.29$$

Se puede observar que el error total es menor para el estimador de diferencia múltiple; sin embargo los coeficientes de variación son muy similares pues resultan ser de

$\frac{18.61}{1689} = 0.01102$ ó sea 1.102 % para

El estimador de razón

$\frac{19.3}{1694} = 0.01139$ ó sea 1.139 % para el de diferencia

El *CV* del de diferentes ligeramente superior en este caso al de razón.

Ejemplo 3.-

Se desea estimar el número promedio de cabezas de ganado de un cierto municipio, el cual comprende 250 fincas. Los datos a utilizar son datos muestrales del ganado en el período actual y datos censales de 4 años anteriores.

Para tal efecto se tomó una muestra aleatoria estratificada. Se hizo esta estratificación debido a la gran variación que en cabezas de ganado habia de finca a finca, en otras palabras se agruparon fincas con características similares, con el objeto de aumentar la precisión de las estimaciones.

La variable de estratificación usada fué el ganado existente en el período inmediato anterior.

Para la afijación del tamaño de muestra ($n = 70$) entre los diversos estratos se utilizó afijación de Neyman entre los 7 estratos escogidos por 1/2 de valores con los siguientes resultados :

h	N_h	S_h	$N_h\ S_h$
1	8	514.21	4113.68
2	13	238.11	3095.43
3	20	171.88	3437.60
4	20	91.88	1836.00
5	31	113.17	3508.27
6	47	38.93	1829.71
7	111	80.00	8880.00
	250		26700.69

Al aplicar la fórmula

$$n_h = n \cdot \frac{N_h}{\sum N_h S_h}$$

Se llega al resultado siguiente :

$n_1 = 8$

$n_2 = 11$

$n_3 = 9$

$n_4 = 5$

$n_5 = 9$

$n_6 = 5$

$n_7 = 23$

Las medias de los censos anteriores son :

$$\mu_1 = 542$$

$$\mu_2 = 555$$

$$\mu_3 = 552$$

$$\mu_4 = 571$$

Los resultados obtenidos de la muestra fueron :

Medias ponderadas	Razones
$\hat{m}_0 = 569.37$	
$\hat{m}_1 = 562.91$	$r_{e_1} = 1.0114$
$\hat{m}_2 = 576.24$	$r_{e_2} = 0.988$
$\hat{m}_3 = 567.28$	$r_{e_3} = 1.0036$
$\hat{m}_4 = 589.33$	$r_{e_4} = 0.9661$

La matriz de covarianza unitaria S fué :

$$\begin{bmatrix} 908201 & 873197 & 895813 & 893383 & 951247 \\ & 845872 & 866169 & 862720 & 918156 \\ & & 891871 & 887355 & 946763 \\ & & & 887030 & 944319 \\ & & & & 1013740 \end{bmatrix}.$$

La matriz de covarianzas (C) relativas es :

$$\begin{bmatrix} 0.909860 & 0.889007 & 0.891321 & 0.894366 & 0.907695 \\ & 0.875178 & 0.875837 & 0.877701 & 0.890352 \\ & & 0.881334 & 0.881380 & 0.897243 \\ & & & 0.887360 & 0.900428 \\ & & & & 0.931351 \end{bmatrix}$$

El estimador de razón múltiple es :

$$\hat{\mu}_o^c = l_1 r_{e1} \mu_1 + l_2 r_{e2} \mu_2 + l_3 r_{e3} \mu_3 + l_4 r_{e4} \mu_4$$

Donde :

$$l = \frac{e \hat{A}^{-1}}{e \hat{A}^{-1} e'}$$

La matriz A es como sigue :

$$\begin{bmatrix} .005060 & .003874 & .00302 & .00253 \\ & .001620 & .00464 & .00583 \\ & & .00611 & .00593 \\ & & & .01140 \end{bmatrix}$$

Su inverso es :

$$A^{-1} = \begin{bmatrix} 426.29 & -305.05 & -77.68 & 161.75 \\ & 645.89 & -170.88 & -173.70 \\ & & 464.08 & -136.76 \\ & & & 225.17 \end{bmatrix}$$

De donde :

$$\begin{aligned} l_1 &= 0.575436 \\ l_2 &= -0.010482 \\ l_3 &= 0.220746 \\ l_4 &= 0.21430 \end{aligned} \qquad e \hat{A}^{-1} e' = 356.79$$

De aquí :

$$\hat{\mu}_o^c = 0.575436 \times 1.0114 \times 542 - .010482 \times .988 \times 55$$
$$.21430 \times .96661 \times 571$$

$$\hat{\mu}_o^c = 550$$

La varianza de la estimación es

$$V(\hat{\mu}_o) = \frac{(1 - 70/250)}{70} \; \frac{(550)^2}{356.79} = (3.6)^2$$

Para el estimador de diferencia múltiple.-

Matriz Z

$$\begin{bmatrix} 7679 & 5360 & 4341 & 1913 \\ & 8446 & 6360 & 7904 \\ & & 8465 & 7890 \\ & & & 19447 \end{bmatrix}$$

$$Z^{-1}$$

$$10^{-6} \times \begin{bmatrix} 3.042 & -2.226 & -0.727 & 0.901 \\ & 4.619 & -1.259 & -1.147 \\ & & 3.146 & -0.693 \\ & & & 1.173 \end{bmatrix}$$

Como :

$$u = \frac{e\hat{Z}^{-1}}{e\hat{Z}^{-1}e'}$$

$$\begin{aligned} u_1 &= 0.58999 \\ u_2 &= -0.00775 \\ u_3 &= 0.27831 \\ u_4 &= 0.13945 \end{aligned}$$

$$\tilde{\mu}_0 = 0.592833\,[569.37 + (542-562.91)] - .008402\,[569.37\quad(555 - 576.24)] + 0.279584\,[569.37 + (552 - 567.28)] + 0.13958\,[569.37 + (571-589.33)]$$

$$= \underline{552}$$

Su varianza esta dada por

$$V(\tilde{\mu}_0) = \frac{(1 - 70/251)}{70} \cdot \frac{1}{.000168} = (9.4)^2$$

Mientras que el de diferencia es

$$\frac{9.4}{552} = .017 = 1.7$$

El coeficiente de variación del estimador de razón en este caso resuelto menor que el de diferencia, esto se debe como mencionamos anteriormente a que la serie es más bien decreciente, lo cual se podrá observar de las medias poblacionales.

Es interesante notar cuales fueron los coeficientes de correlación del problema anterior.

<u>Matriz de Correlación</u>

$$\begin{bmatrix} 1 & .996251 & .995348 & .995354 & .991380 \\ & 1 & .997251 & .995974 & .991521 \\ & & 1 & .997648 & .995699 \\ & & & 1 & .995835 \\ & & & & 1 \end{bmatrix}$$

C O N C L U S I O N E S

1.- Los métodos de razón y de diferencia múltiple pueden proporcionar estimadores que aprovechan la información censal anterior y logran de esa maner a aumentar la precisión de las estimaciones considerablemente.

2.- Cuando las razones son cercanas a la unidad, el estimador de diferencia múltiple y el de razón múltiple son métodos equivalentes.

3.- Bajo ciertas condiciones el estimador de diferencia tiene menor varianza que el de razón.

4.- Los estimadores de razón y diferencia múltiples tal como han sido enfocados en esta tesis, pueden ser aplicados como ya mencionamos; en industria, en empresas privadas, en el sector público y en general en cualquier situación donde se desee aprovechar la información anterior para mejorar la precisión de las estimaciones muestrales. Una ventaja importante de los métodos anteriores, es que no solamente se pueden aplicar en series de tiempo, sino también tienen muchas otras aplicaciones cuando las variables auxiliares no son necesariamente datos censales.

5.- Un tema que puede ser importante conocer, es el método de regresión múltiple, el cual podría ser tema de otra tesis. Igualmente la aplicación de los estimadores de razón, diferencia y regresión múltiple en muestreo doble.

BIBLIOGRAFIA

COCHRAN, W.G. — "Sampling Techniques"
Wiley and Sons (1963)

HANSEN, M.H., HURWITZ, W.N. Y MADOW, W. G. — Sample Survey Methods and Theory
Vols. I y II
Wiley and Sons (1953)

KISH, LESLIE. — "Survey Sampling"
Wiley and Sons (1965)

NACIONES UNIDAS — "Breve Manual de Muestreo" Vol. I
Instituto Interamericano de Estadística
Washington, D. C. (1962)

OLKIN, INGRAM — "Multivariate Ratio Estimation for finite Population"

SUKHATME, P.V. — "Teoría de Encuestas por Muestreo con Aplicaciones"
Fondo de Cultura Económica,
México, (1956)