ESTADISTICA ESPAÑOLA
Núm. 94, 1982, págs. 103 a 112

Aplicación del estimador de la razón a la determinación de los tamaños muestrales óptimos para la elaboración de números índices

por JOSE LUIS VIEDMA
Instituto Nacional de Estadística

RESUMEN

Como aplicación del estimador de la razón, se presenta un procedimiento para determinar los tamaños muestrales necesarios para la elaboración de un índice simple, en el muestreo aleatorio simple y en el muestreo aleatorio estratificado, de una población finita. En este último caso, se obtiene e interpreta la afijación óptima de la muestra.

Por último, se generaliza el método en ambos tipos de muestreo, al caso de la elaboración de números índices compuestos, obteniendo e interpretando las afijaciones óptimas resultantes.

Palabras clave: Estimador de la razón, tamaño muestral, afijación óptima, número índice simple, número índice compuesto.

I. INTRODUCCION

La mayoría de los textos y de los estudios relativos a la metodología de obtención de números índices centran su atención en la propia estructura del índice. Para estudiar la evolución en el tiempo de una variable (precio, producción, salario, etc.) o su distribución espacial referida a un mismo instante de observación, se suele elaborar un índice, bien sea un índice simple o un índice compuesto de índices simples agregados con un conjunto de ponderaciones determinado. En todos los casos se suele partir de una hipótesis inicial —generalmente no contrastada— de que el conjunto de unidades informantes acerca de los valores de la variable estudiada, es completamente representativo de la población a la que pertenece.

Desde un punto de vista objetivo, el procedimiento que se debería seguir para la obtención de un número índice sería:

1.º Decidir cuál es el tipo de índice que mejor se ajusta a las necesidades de la situación estudiada, y en el caso de requerirse un índice compuesto, establecer la fórmula agregativa de los índices simples que lo constituyen.

2.º Construir un marco o directorio, constituido por todas aquellas unidades que puedan suministrar la información a que se refiere el índice —o al menos sea un conjunto de conglomerados de dichas unidades—, que cubra la población total.

3.º Seleccionar una muestra de unidades de ese directorio que sea representativa del mismo, en el sentido de permitir la elaboración de un «índice estimado» en base a las observaciones muestrales, con un error de muestreo inferior a un valor prefijado en función de las necesidades y del coste del estudio.

El objeto del presente estudio es precisamente el de presentar una forma de determinar el tamaño muestral necesario para obtener un estimador de un índice simple en un muestreo aleatorio y en un muestreo aleatorio estratificado, en el cual se determina también la afijación óptima, como una aplicación directa del estimador de la razón. Por último, se generaliza el procedimiento para el caso en que se desee elaborar un índice compuesto de varios índices simples, en ambos tipos de muestreo.

II. DETERMINACION DEL TAMAÑO MUESTRAL OPTIMO PARA LA ELABORACION DE UN INDICE SIMPLE DE UNA POBLACION FINITA DE N UNIDADES

2.1. EN UN MUESTREO ALEATORIO SIMPLE SIN REEMPLAZAMIENTO

Sea X la variable a considerar y representemos por X_o y X_t a sus valores en los períodos cero (base) y *t*.

Se desea obtener el tamaño muestral necesario para estimar a partir de una *muestra común,* en ambos períodos el índice simple $R_t = X_t/X_o$, con un error de muestreo inferior a un valor prefijado e (por ejemplo, $e = 0{,}01$ para un error de muestreo del 1 por 100 en el valor de R_t).

Observando que R_t es la expresión de una razón de variables aleatorias, y que además dichas variables, generalmente, estarán muy positivamente correlacionadas, por ser la misma variable, aunque observada en instantes o lugares diferentes, se puede pensar en la conveniencia de utilizar como estimador de $R_t = X_t/X_o$ la razón muestral $\hat{R}_t = \bar{x}_t/\bar{x}_o$. Su varianza en el muestreo se obtiene a partir de la expresión general de la varianza aproximada de orden $1/n$ del estimador de la razón, que en este caso resulta:

$$V(\hat{R}_t) \doteq R_t^2(C_{\bar{x}_t} + C_{\bar{x}_o} - 2C_{\bar{x}_t\bar{x}_o})$$

siendo $C^2_{\bar{x}t}$ y $C^2_{\bar{x}o}$ los cuadrados de los coeficientes de variación (o viarianzas relativas) de la media muestral de la variable X en los períodos o lugares t y o, respectivamente, y $C_{\bar{x}t\bar{x}o}$ el coeficiente de covariación (o covarianza relativa) de las medias muestrales x_t, x_o, dados por:

$$C^2_{\bar{x}t} = \frac{V(\bar{x}_t)}{E(\bar{x}_t)^2} = \frac{\frac{1-f}{n}S^2_t}{\bar{X}^2_t} \qquad \text{con } S^2_t = \frac{\sum_{i=1}^{N}(X_{it} - \bar{X}_t)^2}{N-1}$$

$$C^2_{\bar{x}_o} = \frac{V(\bar{x}_o)}{E(\bar{x}_o)^2} = \frac{\frac{1-f}{n}S^2_o}{\bar{X}^2_o} \qquad \text{con } S^2_o = \frac{\sum_{i=1}^{N}(X_{io} - \bar{X}_o)^2}{N-1}$$

$$C_{\bar{x}_t\bar{x}_o} = \frac{\text{cov}(\bar{x}_t\bar{x}_o)}{E(\bar{x}_t)E(\bar{x}_o)} = \frac{\frac{1-f}{n}S_{to}}{\bar{X}_t\bar{X}_o} \qquad \text{con } S_{to} = \frac{\sum_{i=1}^{N}(X_{it} - \bar{X}_t)(X_{io} - \bar{X}_o)}{N-1}$$

siendo $f = \frac{n}{N}$ la fracción de muestreo.

Sustituyendo en la expresión aproximada de $V(\hat{R}_t)$ se obtiene:

$$V(\hat{R}_t) = \frac{X_t^2}{X_o^2}\,\frac{(1-f)}{n}\left(\frac{S_t^2}{X_t^2} + \frac{S_o^2}{X_o^2} - 2\,\frac{S_{to}}{X_tX_o}\right) =$$

$$= \frac{(1-f)}{\bar{X}_o^2 n}(S_t^2 + R_t^2S_o^2 - 2R_tS_{to})$$

de donde despejando n

$$n = \frac{NR_t^2\left(\frac{S_t^2}{\bar{X}_t^2} + \frac{S_o^2}{\bar{X}_o^2} - 2\,\frac{S_{to}}{\bar{X}_t\bar{X}_o}\right)}{NV(R_t) + R_t^2\left(\frac{S_t^2}{\bar{X}_t^2} + \frac{S_o^2}{\bar{X}_o^2} - 2\,\frac{S_{to}}{\bar{X}_t\bar{X}_o}\right)} =$$

$$= \frac{N(S_t^2 + R_t^2 S_o^2 + 2R_t S_{to})}{N\bar{X}_o^2 V(R_t) + (S_t^2 + R_t^2 S_o^2 - 2R_t S_{to})}$$

expresión que permite obtener el tamaño muestral necesario para estimar un índice simple $R_t = X_t/X_o$ mediante $\hat{R}_t = \bar{x}_t/\bar{x}_o$ con un error de muestreo $\sqrt{V(\hat{R}_t)}$ prefijado.

En la práctica los valores poblacionales de R_t, X_t, X_o, S_t^2, S_o^2 y S_{to} serán desconocidos. En este caso se estimarán a partir de una muestra ensayo.

II.2. En un muestreo aleatorio estratificado sin reemplazamiento. Afijación óptima

Supongamos una población finita de N unidades estratificada en L estratos h = 1, 2, ..., L, de tamaños N_h conocidos. Utilizaremos como estimador de $R_t = X_t/X_o$, el estimador combinado de la razón en el muestreo estratificado

$$\hat{R}_{ct} = \frac{\bar{x}_{st}}{\bar{x}_{so}}$$

siendo

$$\bar{x}_{st} = \sum_{h=1}^{L} \frac{N_h \bar{x}_{ht}}{N}$$

$$\bar{x}_{so} = \sum_{h=1}^{L} \frac{N_h \bar{x}_{ho}}{N}$$

por presentar, en general, una variabilidad mayor que el estimador separado, y por simplificar los cálculos al no ser necesario considerar estimadores de la razón distintos en cada uno de los estratos.

La varianza aproximada de orden $1/n$ de $\hat{R}_{ct}$ viene dada por

$$V(\hat{R}_{ct}) \doteq R^2(C^2_{\bar{x}st} + C^2_{\bar{x}so} - 2C_{\bar{x}st\bar{x}so})$$

siendo

$$C^2_{\bar{x}_{st}} = \frac{V(\bar{x}_{st})}{E(\bar{x}_{st})^2} = \frac{\sum_{h=1}^{L} \frac{N_h^2}{N^2} \frac{1 - f_h}{n_h} S^2_{ht}}{\bar{X}_t^2} \quad \text{con } S^2_{ht} = \frac{\sum_{i=1}^{Nh} (X_{iht} - \bar{X}_{ht})^2}{N_h - 1}$$

$$X^2_{\bar{x}_{so}} = \frac{V(\bar{x}_{so})}{E(\bar{x}_{so})^2} = \frac{\sum_h \frac{N_h^2}{N^2} \frac{1 - f_h^2}{n_h} S_{ho}}{X_o^2} \quad \text{con } S^2_{ho} = \frac{\sum_i (X_{iho} - \bar{X}_{ho})^2}{N_h - 1}$$

$$C_{\bar{x}_{st}\bar{x}_{so}} = \frac{COV(\bar{x}_{st}\bar{x}_{so})}{E(x_{st})\,E_{(xso)}} = \frac{\sum_h \frac{N_h^2}{N^2} \frac{1 - f_h}{n_h} S_{hto}}{\bar{X}_t \bar{X}_o}$$

con

$$S_{hto} = \frac{\sum_i (X_{iht} - \bar{X}_{ht})(X_{iho} - \bar{X}_{ho})}{N_h - 1}$$

Sustituyendo en $V(\hat{R}_{ct})$ y simplificando, se obtiene

$$V(\hat{R}_{ct}) = \frac{1}{\bar{X}_o N^2}\left(\sum_h \frac{N_h^2 S_{ht}^2}{n_h} + R^2 \sum_h \frac{N_h^2 S_{ho}^2}{n_h} - 2R \sum_h \frac{N_h^2}{n_h} S_{hto}\right) -$$

$$- \frac{1}{\bar{X}_o N^2}\left(\sum_h N_h S_{ht}^2 + R^2 \sum_h N_h S_{ho}^2 - 2R \sum_h N_h S_{hto}\right)$$

La afijación óptima n_h se halla minimizando $V(\hat{R}_{ct})$ condicionado a que $\sum_h n_h = n$.

Construyendo la función

$$\phi(n_h) = V(\hat{R}_{ct}) - \lambda\left(\sum_h n_h - n\right)$$

derivando respecto a n_h, e igualando a cero

$$\frac{1}{\bar{X}_o^2 N^2}\left(- \frac{N_h^2 S_{ht}^2}{n_h^2} - R^2 \frac{N_h^2 S_{ho}^2}{n_h^2} + 2R \frac{N_h^2 S_{hto}}{n_h^2}\right) - \lambda = 0$$

de donde

$$n_h = -\frac{1}{N\bar{X}_o\sqrt{\lambda}} N_h (S_{ht}^2 + R^2 S_{ho}^2 - 2RS_{hto})^{1/2}$$

e imponiendo que $\sum_h n_h = n$, se deduce finalmente

$$n_h = n \frac{N_h (S_{ht}^2 + R^2 S_{ho}^2 - 2RS_{\overline{hto}})^{1/2}}{\sum_h N_h (S_{ht}^2 + R^2 S_{ho} - 2RS_{hto})^{1/2}}$$

Sustituyendo en las expresiones de las varianzas relativas, y despejando n se tiene

$$n = \frac{\left[\sum_h N_h (S^2_{ht} + R^2 S^2_{ho} - 2RS_{hto})^{1/2} \right]^2}{N^2 \bar{X}^2_o V(R_{ct}) + \sum_h N_h (S^2_{ht} + R^2 S^2_{ho} - 2RS_{hto})}$$

que es la expresión que permite obtener n en función de R, $\bar{X}_t$, $\bar{X}_o$, S^2_{ht}, S^2_{ho} y S_{hto}, para unos tamaños poblacionales N_h conocidos y una varianza $V(\hat{R}_{ct})$ prefijada.

Se puede observar cómo esta expresión generaliza la fórmula obtenida para un muestreo aleatorio simple, que queda como caso particular si se considera solamente un estrato.

Interpretación de la afijación óptima obtenida

La afijación óptima obtenida en el muestreo aleatorio estratificado para el estimador de la razon combinado favorecerá los tamaños muestrales de aquellos estratos con mayor número de unidades, y con mayor varianza en t y en o, en tanto que se reducirá el tamaño n_h en aquellos estratos en los que la covarianza entre t y o sea mayor, supuesta positiva.

III. DETERMINACION DEL TAMAÑO MUESTRAL OPTIMO PARA LA ELABORACION DE UN INDICE COMPUESTO DE UNA POBLACION FINITA DE N UNIDADES

Supongamos que se desea obtener el tamaño muestral necesario para estimar un índice compuesto

$$R^*_t = \sum_{i=1}^{I} W_i R^i_t \quad \text{con} \quad W_i \geq 0 \quad \text{y} \quad \sum_{i=1}^{I} W_i = 1$$

de un conjunto de índices simples R^i_t, $i = 1, 2, \ldots, I$, referidos a I modalidades de la característica X investigada, con una precisión prefijada $V(R^*_t)$. Este es el caso que se presentaría, por ejemplo, al tratar de estimar un índice de producción industrial, en cuyo caso la variable X representaría las cantidades producidas y las modalidades i serían los productos que comprendiera el índice complejo.

En primer lugar se supondrá que el tamaño muestral total n, necesario para obtener el índice compuesto R^*_t, es conocido, lo que permitirá obtener la afijación óptima de mínima varianza para el conjunto de las I modalidades consideradas. Posteriormente, partiendo de la afijación resultante se determinará el tamaño n necesario para estimar R_t^* con la precisión prefijada $V(R_t^*)$.

La expresión de partida será:

$$V(R_t) = \sum_i W_i^2 V(R_t^i)$$

lo que supone aceptar por simplicidad que los índices elementales R_t^i para el conjunto de las I modalidades consideradas son mutuamente independientes, de manera que la evolución de la variable X entre o y t, en una modalidad i, no depende del comportamiento de dicha variable en las restantes modalidades.

III.1. EN UN MUESTREO ALEATORIO SIMPLE SIN REEMPLAZAMIENTO. AFIJACIÓN ÓPTIMA

Se trata de minimizar $V(R_t)$ dado por

$$V(R_t) \doteq \sum_i \frac{W_i^2\left(1 - \frac{n_i}{N_i}\right)}{\overline{X}_{io}^2 n_i}(S_{it}^2 + (R_t^i)^2 S_{io}^2 - 2R_t^i S_{ito})$$

sujeto a $\sum_i n_i = n$

Derivando respecto a n_i e igualando a cero, resulta la afijación óptima siguiente:

$$n_i = n \frac{\frac{W_i}{\overline{X}_{io}}(S_{it}^2 + (R_t^i)^2 S_{io}^2 - R_t^i S_{ito})^{1/2}}{\sum_i \frac{W_i}{\overline{X}_{io}}(S_{it}^2 + (R_t^i) S_{io} - R_t^i S_{ito})^{1/2}}$$

Sustituyendo el valor obtenido de n_i en la expresión $V(R_t)$ y despejando n se obtiene:

$$n = \frac{\left[\sum_i \frac{W_i}{\overline{X}_{io}}(S_{it}^2 + (R_t^i)^2 S_{io}^2 - 2R_t^i S_{ito})^{1/2}\right]^2}{V(R_t^*) + \sum_i \frac{W_i^2}{\overline{X}_{io}^2 N_i}(S_{it}^2 + (R_t^i)^2 S_{io}^2 - 2R_t S)}$$

Esta expresion se particulariza a la obtenida en el apartado II.1, para el caso de considerar índices simples, haciendo $W_i = 0$, para todas las modalidades excepto para una —la que se refiera al índice simple—, en que $W_i = 1$.

Interpretación de la afijación óptima obtenida

En general, la afijación obtenida potenciará los tamaños muestrales de aquellas modalidades con mayor ponderación en el índice complejo, con mayor variabilidad en *o* y en *t*, y con menor covarianza (positiva) entre *o* y *t*.

III.2. En un muestreo aleatorio estratificado sin reemplazamiento. Afijación óptima

Se trata de minimizar la expresión:

$$V(R_t) \doteq \sum_i \frac{W_i^2}{\bar{X}_{io}^2 N_i^2}\left[\left(\sum_{hi} \frac{N_{hi}^2 S_{hit}^2}{n_{hi}} + (R_t^i)^2 \sum_{hi} \frac{N_{hi}^2 S_{hit}^2}{n_{hi}} - \right.\right.$$

$$\left.\left. - 2R_t^i \sum_{h_i} \frac{N_{hi}^2 S_{hito}^2}{n_{hi}}\right) - \left(\sum_{hi} N_{hi} S_{hit}^2 + (R_t^i) \sum_{hi} N_{hi} S_{hio}^2 - 2R_i^t \sum_{ni} N_{hi} S_{hito}\right)\right]$$

donde:

$$n_{hi} = n_i \cdot \frac{N_{hi}[S_{hit}^2 + (R_t^i)^2 S_{hio}^2 - 2R_t^i S_{hito}]^{1/2}}{\sum_{hi} N_{hi}(S_{hit}^2 + (R_t^i)^2 S_{hio}^2 - 2R_t^i S_{hito})^{1/2}}$$

Derivando respecto a n_i e igualando a cero, la afijación óptima que resulta es:

$$n_i = n \frac{\frac{W_i}{\bar{X}_{io} N_i}\left[\sum_{hi} N_{hi}(S_{hit}^2 + (R_t^i)^2 S_{hio}^2 - 2R_t S_{hito})^{1/2}\right]}{\sum_i \frac{W_i}{\bar{X}_{io} N_i}\left[\sum_{hi} N_{hi}(S_{hit}^2 + (R_t^i)^2 S_{hio}^2 - 2R_t^i S_{hito})^{1/2}\right]}$$

Sustituyendo este resultado de la expresión de n_i en la fórmula de $V(R_t)$ y despejando n, resulta finalmente

$$n = \frac{\left\{\sum_i \frac{W_i}{\bar{X}_{io} N_i}\left[\sum_{hi} N_{hi}(S_{hit}^2 + (R_t^i)^2 S_{hio}^2 - 2R_t^i S_{hito})^{1/2}\right]\right\}^2}{V(R^*) + \sum_i \frac{W_i^{\;2}}{\bar{X}_{io}^2 N_i^2}\left[\sum_{hi} N_{hi}(S_{nit}^2 + (R_t^i)^2 S_{hio}^2 - 2R_t^i S_{hito})\right]}$$

Como en el caso anterior, esta expresión coincide con la obtenida en el apartado II.2 en el caso de considerar índices simples.

Interpretación de la afijación óptima obtenida

En general, la afijación obtenida potenciará los tamaños muestrales de aquellas modalidades con mayor ponderación en el índice complejo, y dentro de cada modalidad favorecerá los tamaños de los estratos con mayor proporción de unidades N_{hi}/N_i, y con mayores variables en los instantes *o* y *t* y menor covarianza entre *o* y *t*.

SUMMARY

APPLICATION OF THE RATIO ESTIMATOR TO THE DETERMINATION OF THE OPTIMUM SAMPLE SIZES FOR THE ELABORATION OF INDEX NUMBERS

As an application of the ratio estimator, the author proposes a method to fix the sample sizes necessary to elaborate a simple index, in the simple random sampling as well as in the stratified random sampling, of a finite population. In the latter case, the optimum sampling allocation is achieved and interpreted.

Afterwards in both sampling type, the method is extended to the case of elaborating composite index numbers whereas the resulting optimum allocations are achieved and interpreted.

Key words: ratio estimator, sample size, optimum allocation, simple index number, composite index number.

AMS, 1970. Subject classification 62D05.