Visión algebraica unificada
para el cálculo del tamaño de la muestra

Vicente Manzano Arrondo. 1997

Resumen

La decisión sobre el tamaño de una muestra está en función de multitud de aspectos, principalmente del modelo de selección de la muestra y de la función poblacional a estimar. La combinación de los factores implicados genera una gran cantidad de expresiones de cálculo específicas. Esta situación lleva consigo dos consecuencias indeseables. Por un lado, dificulta la automatización del proceso en la elaboración de algoritmos de cálculo. Por otro, desorienta a los lectores e investigadores que buscan una expresión idonea. En el presente trabajo se sugiere la inclusión de variables globales en las fórmulas para el cálculo del tamaño de la muestra. Con ello, se generan dos consecuencias positivas. Por un lado, se reduce sensiblemente el número de expresiones de cálculo requeridas. Por otro, se facilita la obtención de una visión global, general o unificada en el problema del cálculo de un tamaño para la muestra.

Palabras clave: tamaño de muestra.

Cuestiones previas

Entre los procedimientos de extracción de muestras aleatorias, puede considerarse que existen cuatro modelos estándares: aleatorio simple, estratificado, de conglomerados monoetápico y de conglomerados con submuestreo. A su vez, las estimaciones más comunes se realizan para proporciones y medias, bien sea en problemas de una sola variable o estableciendo diferencias (en grupos relacionados o independientes. Además, las expresiones se duplican cuando se considera la existencia o no de una hipótesis alternativa en las pruebas de hipótesis. Por último, las situaciones de aplicación requieren distinguir entre poblaciones finitas e infinitas. De esta forma, considerando únicamente los contextos más usuales, deberían generarse un centenar de expresiones de cálculo para las estimaciones de las varianzas o para la decisión sobre el tamaño de la muestra.

Esta variedad de situaciones complica enormemente el establecimiento de algoritmos de cálculo automático y, sobretodo, la comunicación de expresiones en manuales y la visión de patrones comunes.

La inclusión de las variables globales A, B y C

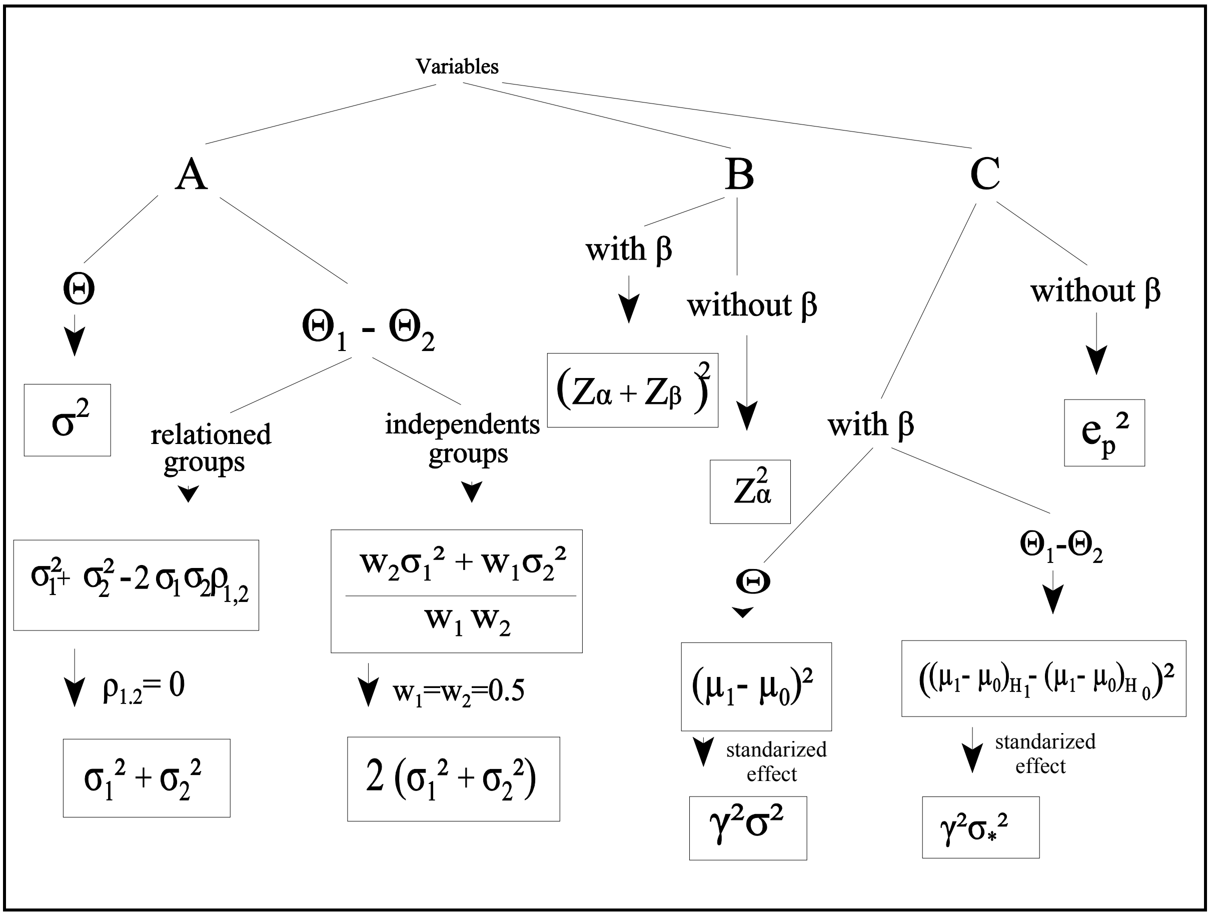
Una posible solución es la identificación de entidades comunes a grupos amplios de expresiones de cálculo y su inclusión simbólica en las fórmulas. Con tal objetivo, sugiero el establecimiento de las siguientes variables globales:

A: medida de variación en la población.

B: medida de probabilidad o riesgos asociados a la estimación y/o decisión.

C: medida de error en la estimación o en la decisión.

El cuadro 1 muestra los valores que definen a estas tres variables, en función del conjunto de factores mencionados en el punto anterior.



Cuadro 1: variables globales para la definición del cálculo del tamaño c

donde:

Θ	proporción o media aritmética en un problema univariable
$\Theta_1 - \Theta_2$	diferencia entre proporciones o entre medias
σ_i	desviación estandar del grupo i
σ_*	$\sqrt{w_1 \sigma_1^2 \% w_2 \sigma_2^2}$
$\rho_{1,2}$	correlación lineal entre los grupos 1 y 2
w_i	peso del grupo i
Z_α	medida estandarizada para el error tipo I
Z_β	medida estandarizada para el error tipo II
e_p	error de precisión (radio del intervalo de estimación)

μ_i media aritmética del grupo i

γ tamaño estandarizado de efecto

De esta forma, considerando explícito únicamente el modelo de selección de muestras y la finitud de las poblaciones, se generan las siguientes expresiones de cálculo para el tamaño de la muestra:

ALEATORIO SIMPLE

$$\frac{N}{\frac{C(N-1)}{A\cdot B}+1} \qquad \text{Para } N\to\infty \quad \frac{A\cdot B}{C}$$

ESTRATIFICADO

$$\frac{\left(\sum N_i\sqrt{A_i}\right)^2}{\frac{C\cdot N^2}{B}+\sum N_i A_i} \qquad \text{Para } N\to\infty \quad \frac{B\left(\sum w_i\sqrt{A_i}\right)^2}{C}$$

CONGLOMERADOS MONOETÁPICO

$$\frac{N_C}{\frac{C(N_C-1)}{A\cdot B}+\frac{1}{G}} \qquad \text{Para } N_C\to\infty \quad \frac{A\cdot B}{C}[\delta(G-1)+1]$$

CONGLOMERADOS CON SUBMUESTREO

$$\frac{G\sigma_d^2 n_C}{\sigma_d^2+(G-1)\left[n_C\frac{C}{B}-\sigma_e^2\frac{N_C-n_C}{N_C-1}\right]}$$

Para N ÷ 4 $$\frac{G\sigma_d^2 n_C}{\sigma_d^2 \% (G \& 1)\left[n_C \frac{C}{B} \& \sigma_e^2\right]}$$

Para G ÷ 4 $$\frac{\sigma_d^2}{n_C \frac{C}{B} \& \sigma_e^2 \frac{N_C \& n_C}{N_C \& 1}}$$

Para G,N_C ÷ 4 $$\frac{\sigma_d^2}{n_C \frac{C}{B} \& \sigma_e^2}$$

donde:

N	Número de unidades elementales en la población
N_C	Número de conglomerados en la población
N_i	Número de unidades elementales en el estrato **i**
n	Número de unidades elementales en la muestra
n_C	Número de conglomerados en la muestra
w_i	Peso del estrato **i** en la población
A_i	Valor de la variable A en el estrato **i**
G	N / N_C
g	n / n_C
δ	correlación u homogeneidad entre conglomerados
σ_e^2	$A\frac{\delta(G \& 1) \% 1}{G}$ (si G ÷ 4, $\sigma^2 \cdot \delta$)
σ_d^2	A - σ_e^2

Nota del autor

Las expresiones de cálculo para el tamaño de la muestra han sido deducidas de las correspondientes para el error típico que se encuentran en los textos al uso sobre teoría del muestreo, como por ejemplo Sukhatme (1951), Hansen y otros (1953), Kish (1965), Raj (1968), Hájek (1981) o Yates (1981) . Las expresiones de partida y las deducciones se encuentran íntegras en la tesis doctoral del autor, Manzano (1996). Las expresiones sobre el tamaño del efecto para los contextos de inferencia aquí mencionados, se encuentran en Cohen (1988).

Referencias

Cohen, J. (1988). *Statistical power for the behavioral sciences.*. Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers.

Hàjek, J. (1981). *Sampling from a finite population*. Nueva York: Marcel Dekker.

Hansen, M. H.; Hurwitz, W. N. y Madow, W. G. (1953). *Sample survey. Methods and theory. Vol 1: Methods and applications. Serie Wiley Classics Library.*. Nueva York: John Wiley & Sons.

Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons.

Manzano, V. (1996). *Tamaño óptimo de muestra en investigación mediante encuestas. Fundamentos e implementación de un sistema de ayuda a la decisión*. Tesis doctoral no publicada. Facultad de Psicología. Universidad de Sevilla.

Raj, D. (1968). *Sampling theory*. Nueva York: McGraw-Hill.

Sukhatme, P.V. (1953). *Sampling theory with applications*. Iowa: Iowa State College Press.

Yates, F. (1981). *Sampling methods for censures and surveys*. High Wycombe, England: Charles Griffin.