ALGUNAS CONSIDERACIONES EN EL MUESTREO

Fernando Rivero Suguiura

Tamaño de Muestra Optimo

Un tamaño de muestra optimo para la ejecución de un encuesta, que cuenta con información de la varianza de las variables que se desean estudiar, así como el interés de estimar una proporción de tipo binomial bajo un esquema de selección aleatorio simple (MAS), fue dado por Cochran (1953), bajo el supuesto de una distribución normal en el parámetro, propuso un expresión para calcular el tamaño de muestra dado por:

$$n = \frac{z^2pq}{\varepsilon^2}$$

que se utiliza con mucha frecuencia y ligereza en investigaciones por muestreo y en los trabajos de tesis de estudiantes de diferentes carreras universitarias.

Cuando no se cuenta con información de la varianza del parámetro de interés -lo cual sucede frecuentemente- la práctica aconseja asumir un valor de el cual maximiza la variabilidad del estimador y genera un tamaño de muestra que en apariencia garantiza la precisión deseada.

Esta forma de proceder llevaría a suponer que la encuesta sólo desea obtener estimaciones de una variable sin considerar que prácticamente todas las encuestas son de propósitos múltiples. Lo cual hace pensar que la variable de interés tenga más bien una distribución multinomial.

El procedimiento sugerido por Cochran puede subestimar el tamaño de muestra, pues no permite fijar en forma simultánea un coeficiente de confianza para todas las categorías en que se distribuye la variable de estudio lo cual inhibe al investigador sobre la posibilidad de controlar la precisión deseada de las estimaciones.

Para resolver este problema, se han propuesto algunos algoritmos de optimización no lineal que permiten calcular el tamaño de muestra óptimo (Kokan 1963 y Khan 1967).

Efecto de Diseño

Para resolver esta situación que se presenta en el cálculo del tamaño de muestra en encuestas complejas, donde no se cuenta con información sobre la varianza de las variables de interés, Kish (1979) p r o p o n e la definición de un factor de ajuste que a partir de una muestra aleatoria simple, permita aproximarse al número de selecciones necesarias para que un diseño de conglomerados proporcione la misma varianza. El factor se conoce como el efecto de diseño (deft) y está dado por:

$$deft = \frac{Var_C(\bar{Y})}{(1-f)S^2/n}$$

donde es la varianza de la variable de interés en el diseño de conglomerados y representa la varianza de un muestreo aleatorio simple, de tal forma que dicha varianza se puede escibir como:

$$(1-f)S^2/n$$

Según Kish, este factor permite evitar la complejidad en el diseño de una muestra compleja, sobre todo los de conglomerados o estatificación.

$$Var_C(\overline{Y}) = (1-f)S^2/n * deft$$

Es cotidiano que en un diseño de muestra estratificado, el efecto de diseño es menor a uno ya que los procedimientos tienden a reducir la varianza del estimador debido a la homogeneidad al interior de los grupos formados. Lo contrario sucede en el caso de un diseño de muestreo por conglomerados, donde generalmente dicho efecto de diseño es mayor a uno.

En caso de que los conglomerados sean en promedio de igual tamaño, el efecto de diseño se puede expresar como:

$deft = [1+\rho(M-1)]$ o $deft = [1+\rho(\overline{M}-1)]$

donde y representan el tamaño del conglomerado y el tamaño medio, respectivamente y es el denominado coeficiente de correlación intraconglomerado y se puede calcular por medio de:

$$\rho = \frac{\sum\sum(Y_{ij}-\overline{Y})(Y_{ik}-\overline{Y})}{m(\overline{M}-1)\overline{M}\sigma^2}$$

donde representa el número de conglomerados en la población, mientras que:

$$\sigma^2 = (M\overline{M}-1)S^2/M\overline{M}$$

Luego, la varianza del estimador se puede escirbir en términos del efecto de diseño como se muestra a continuación:

$$Var_C(\overline{Y}) = (1-f)S^2[1+(\overline{M}-1)\rho]nM$$

lo cual permite comparar la varianza de un diseño aleatorio simple sin reemplazo con la que se obtiene bajo un diseño de muestreo por conglomerados, lo que se interpreta como una medida de la eficiencia relativa del diseño.

Finalmente, el comentario afirma lo importante que significa la expresión efecto de diseño, pues permite relacionar un diseño complejo de muestreo con un muestreo aleatorio simple y expresar los tamaños de muestra y varianzas de los estimadores de tal forma que no requiera una construcción compleja, sino en términos de un muestreo aleatorio simple.