

Determinación del tamaño de muestra

Muestreo II

Licenciatura en Estadística

2023

La determinación del tamaño de muestra n para una encuesta debe ser calculado de forma de tener en cuenta los fines analíticos de la encuesta:

- ▶ $moe = cte \times SE(\hat{\theta})$ para distintos parámetros a nivel de U
- ▶ moe a nivel de distintos dominios (aperturas) U_d

En general, los diseños muestrales son complejos (SE difíciles de calcular), existen marcos imperfectos, problemas de no respuesta y presupuestos acotados. El reto es encontrar un n aproximado que satisfaga lo anterior

En un primer paso se asume un muestreo aleatorio simple sin reposición y que el parámetro de interés es la media \bar{Y} .

El tamaño de muestra n para estimar \bar{Y} es:

$$n_0 = (z^* \times \text{var}[y] \times \text{moe}^{-1})^2$$

donde $\text{var}[y]$ puede provenir de una muestra de iluminación (piloto), encuestas anteriores o de una variable x correlacionada con la variable de interés y .

Si el tamaño de población N es pequeño y/o la tasa de muestreo f no es despreciable, n_0 es ajustado:

$$n_1 = \frac{n_0}{1 + n_0/N}$$

El efecto de diseño (deff) puede ser utilizado para ajustar el tamaño de muestra si el diseño muestral utilizado es más "complejo" que el m.a.s.

$$n_2 = n_1 \times \text{deff}$$

donde el deff es generalmente calculado utilizando muestras anteriores (para el parámetro clave) o realizando suposiciones (e.g. en muestreo por conglomerados en varias etapas, el deff, en la práctica, es fijado entre 1.5 y 2)

En el ultimo paso, se realiza un ajuste por perdida de tamaño de muestra por problemas de elegibilidad y no respuesta

$$n_{\text{final}} = \frac{n_2}{TE \times TR}$$

donde

- ▶ la tasa de elegibilidad (TE) puede provenir del marco o relevamientos anteriores
- ▶ la tasa de respuesta (TR) puede ser en base a suposiciones o encuestas anteriores. Indendentemiente de lo anterior, en este paso se asume que la NR es MCAR o MAR.

Una técnica utilizada en la práctica es seleccionar aleatoriamente una muestra de tamaño n asumiendo en el “peor de los casos”, subdividir aleatoriamente la muestra completa s en submuestras de recolección de los datos (réplicas) y utilizar solo el número de réplicas necesarios para alcanzar los objetivos analíticos de la encuesta.

- ▶ Debido a que las réplicas se construyen al azar, decidir retener cualquier réplica no niega la aleatoriedad de la muestra
- ▶ las ponderaciones de las unidades utilizadas se ajustan para reflejar solo aquellas réplicas que fueron liberadas.
- ▶ las réplicas utilizadas se consideran una muestra aleatoria simple de la muestra original a los efectos de calcular el ajuste de los ponderadores.
- ▶ una vez que una réplica es liberada, la misma debe ser realizada en su totalidad.
- ▶ generalmente, las réplicas son construidas a nivel de las UPM dentro de cada uno de los estratos de diseño