

Maestría en Marketing
Métodos y técnicas de análisis cuantitativo y cualitativo

Métodos de Muestreo

- Referencia: PR Cap 11, KT 12-14.

Otoño 2004

Problemas a tratar

- Que alternativas hay para confeccionar una muestra
- A cuántas personas es necesario encuestar
- Cómo tengo que elegir las
- Qué problemas pueden ocurrir si elijo mal la muestra

Algunas definiciones útiles

- Elemento o unidad de análisis
- Población
 - Elemento
 - Unidades de muestreo (etapas)
 - Alcance
 - Tiempo
- Marco muestral

Proceso de muestreo

- Definición de a población (elementos, unidades de muestreo, alcance y tiempo)
- Identificar el marco muestral
- Elegir el procedimiento de muestreo y definir el tamaño de la muestra
- Realizar la muestra

Sesgo de muestreo

- Una muestra se dice que es *sesgada* cuando no es representativa de la población
- Hay dos tipos de sesgos:
 - De selección
 - De no respuesta
- Estos sesgos no se eliminan con una muestra mayor, sino cambiando el procedimiento de muestreo

Notación

- Tamaño de la población : N
- Tamaño de la muestra : n
- Media muestral: $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$
- Desvío estandar muestral: $s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}}$
- Desvío estandar de la media: $s_{\bar{X}} = \frac{s}{\sqrt{n}}$

Intervalos de confianza

- Definimos al intervalo de confianza de nivel $(1-\alpha)$ como

$$IC_{1-\alpha} = \bar{X} \pm Z_{\alpha/2} \frac{s}{\sqrt{n}} = \bar{X} \pm Z_{\alpha/2} s_{\bar{X}}$$

- Para el nivel 95%, el intervalo es

$$IC_{0.95} = \bar{X} \pm 1.96 s_{\bar{X}}$$

- También definimos la precisión como

$$P = Z_{\alpha/2} \frac{s}{\sqrt{n}} = Z_{\alpha/2} s_{\bar{X}}$$

Muestreo Aleatorio Simple

- Una muestra es *aleatoria simple* si cumple que:
 - Cada elemento de la población tiene la misma probabilidad de ser elegido.
 - Los elementos se seleccionan de a uno en uno y con reposición.
- Para seleccionar la muestra por lo general se utilizan los números aleatorios

Ejemplo: ENGH

	Consumo de Tabaco	
N	4907	
n	100	
Media Poblacional	39,38	
Media Muestral	37,59	
Desv Muestral	29,31	
Desv Media	2,93	
Intervalo de Confianza	43,3301	31,8419

Efectos del tamaño de la muestra

- Aumentar el tamaño de la muestra aumenta la precisión en términos del intervalo de confianza

n	Intervalo de Confianza		Precisión
100	43,3301	31,8419	5,7441
200	41,6477	33,5244	4,0617
500	40,1549	35,0172	2,5688
1000	39,4025	35,7696	1,8164

Tamaño óptimo de la muestra

- El tamaño óptimo de la muestra será función de la precisión que queramos para nuestra estimación, dado el nivel de confianza deseado.

- Por ejemplo, para un nivel de confianza de 95%

$$P = 1.96 \frac{s}{\sqrt{n}} \Rightarrow n = \left(1.96 \frac{s}{P} \right)^2$$

- Si expresamos la precisión en términos porcentuales, llamando b a la precisión relativa, entonces

$$b\bar{X} = 1.96 \frac{s}{\sqrt{n}} \Rightarrow n = \left(1.96 \frac{s}{b\bar{X}} \right)^2$$

En en el ejemplo

Precisión Absoluta	n óptimo
5	132
2	825
1	3299
0,5	13198

Precisión Relativa	n óptimo
0,2	58
0,1	234
0,05	934
0,01	23356

Muestreo sistemático

- Cuando los elementos están ordenados (en listas, calles, etc), una alternativa más sencilla que el muestreo aleatorio simple es el sistemático.
- Supongamos que la población tiene un tamaño de 1000 y queremos una muestra de 50 observaciones. Entonces, en toda la población habrá 20 unidades por cada observación seleccionada.
- Lo que hacemos es tomar a los primeros 20 y elegir por muestreo aleatorio simple a uno de ellos, y ese será el primer elemento seleccionado. Luego, elegimos el segundo sumando a ese número 20, y así sucesivamente.

Muestreo sistemático

- Si los elementos en la lista están ordenados aleatoriamente, este método es igual al aleatorio simple.
- Si existe algún ciclo en la lista, la muestra puede sufrir de sesgo de selección.

Muestreo estratificado

- Para seleccionar una *muestra estratificada*:
 - La población se divide en estratos mutuamente excluyentes
 - Se toma una muestra aleatoria simple de individuos dentro del estrato
- Para saber cuantos individuos elegir dentro de cada estrato una alternativa es hacerlo proporcional a la cantidad de individuos de ese estrato en la población.
- La ventaja de usar ese método es que podemos reducir la varianza del estimador

Notación

- Tamaño de la población : N
- Tamaño de la muestra : n
- Tamaño de la población en el estrato 1 y del 2: N_1 N_2
- Tamaño de la muestra en el estrato 1 y del 2: n_1 n_2
- Media muestral del estrato 1 y del 2: \bar{X}_1 \bar{X}_2
- Desvio estandar muestral del estrato 1 y del 2: s_1 s_2
- Desvio estandar de la media del estrato 1 y del 2:
$$s_{\bar{X}_1} = \frac{s_1}{\sqrt{n_1}} \quad s_{\bar{X}_2} = \frac{s_2}{\sqrt{n_2}}$$
- Ponderadores del estrato 1 y del 2: $W_1 = N_1 / N$ $W_2 = N_2 / N$

Algunos cálculos

- Media muestral: $\bar{X} = W_1 \bar{X}_1 + W_2 \bar{X}_2$

- Desvío estandar del promedio de la muestra:

$$s_{\bar{X}} = \sqrt{W_1^2 s_{X_1}^2 + W_2^2 s_{X_2}^2}$$

- Luego, dependiendo como sean la varianzas de cada estrato, la varianza del promedio de la muestra puede ser menor que la del promedio aleatorio simple. Luego la precisión puede ser mayor con este método

En nuestro ejemplo

- El estrato 1 lo conformarán quienes tengan un nivel educativo menor o igual a secundario completo, y el estrato 2 los que tengan un nivel mayor

	Estrato 1	Estrato 2
<i>n</i>	84	14
<i>W</i>	0,84	0,14
Desv muestral	27,06	34,80
Desv media	2,95	9,30
Desv media total	2,80	

Generalizando

- En general, el aumento de la precisión esta asociado con que haya una variabilidad menor dentro de los estratos que la que existe en la muestra total.
- Lo que deseamos es que las variables que utilizamos para estratificar estén correlacionadas con la variable que queremos investigar.
- Notemos que esta característica de; nuestro estratificado puede verse desde dos perspectivas: este método aumenta la precisión ara una muestra dada, o también puede usarse para obtener una misma precisión pero con una muestra menor.

Muestreo por áreas

- Lo métodos vistos hasta ahora suponen que conocemos la lista exacta de todos los elementos en la población, lo cual por lo general no es así.
- La alternativa del muestreo por áreas lo que considera es realizar una muestra aleatoria de las distintas áreas geográficas en la cual esta situada la población de interés
- Este método procede siempre en distintas etapas.
- La ventaja de este método es que es la reducción de costos de implementación. La desventaja es que esta sujeto a mayo error muestral