


SELECCIÓN DE LOS CASOS (II): Muestra estadística

 El objetivo de esta ficha de actividad es introducir el problema de la selección de los casos una vez que se ha definido (aunque sea) provisoriamente el tipo de diseño que organizará la investigación.

- ⇄ El tema del muestreo puede consultarse en varios textos. Los fundamentos estadísticos son los mismos; se diferencian en los campos de aplicación y por tanto en los ejemplos. Tal vez las mayores variaciones se producen en el momento de integrar el muestreo a la técnica de encuestas.
- ⇄ Una bibliografía introductoria en el tema se puede consultar en:


CEA D'ANCONA, Mariángeles (1996) *Metodología cuantitativa. Estrategias y técnicas de investigación social*. Editorial Síntesis. Madrid. Capítulo 5 “la selección de las unidades de observación: el diseño de la muestra”.

RODRIGUEZ, Jacinto (2000) “La muestra: teoría y aplicaciones”. En García Ferrando, Manuel et al *El análisis de la realidad social. Métodos y técnicas de investigación. 3era. Edición*. Alianza Editorial. Madrid.

- ⇄ Conviene en este nivel revisar las lecturas de los autores clásicos en la materia:

KISH, Leslie (1978) “La selección de la muestra” . En Leon Festinger & Daniel Katz *Metodología de investigación en las ciencias del comportamiento*. Editorial Paidós. Buenos Aires.

COCHRAN, William (2000) *Técnicas de muestreo*. Editorial CECSA. Decimoquinta reimpresión. México DF.

 Se denomina **muestreo estadístico** a aquel proceso que seleccionó un sub-conjunto de casos de un universo a partir de probabilidades conocidas de selección.

- ⇄ una muestra estadística permite al investigador establecer estimaciones de los valores de la población a partir de los datos de la muestra y luego calcular los intervalos de confianza respecto de estas estimaciones.
- ⇄ una muestra estadística será denominada **al azar** cuando el procedimiento implique que cada uno de los elementos de la muestra ha sido seleccionado con equiprobabilidades.


 **Glosario de términos sobre muestreo estadístico**

- ⇄ Universo es el conjunto de elementos definidos por extensión o comprensión para los cuales se desea construir información estadística a través de medidas de tenencia, de dispersión, de concentración, de forma.
- ⇄ Marco muestral: es el listado completo de los elementos que componen del universo. Debe estar

SELECCIÓN DE LOS CASOS (II): Muestra estadística

actualizado y comprobar que no existen exclusiones que puedan generar sesgos.

- Parámetros : se refieren a las medidas estadísticas de la población sobre las cuales se desea producir información por vía de una inferencia. Por convención estos estimadores se escriben utilizando letras griegas.
- Estimadores : es cualquier función calculable a partir de datos de la muestra, por ejemplo, la media o la varianza muestrales. Por convención estos estimadores se escriben utilizando letras minúsculas latinas.

 Existen varios tipos “puros” de muestreo: i) el aleatorio simple (MAS); ii) el muestreo sistemático; iii) el muestreo aleatorio estratificado (MAE) y iv) el muestreo por conglomerado.

- La estrategia de investigación en el momento de seleccionar los casos no requiere restringirse a un único tipo de muestreo. Por el contrario, en muchas ocasiones, suelen combinarse los distintos tipos de muestreo.
- Hay una forma de combinar sucesivamente distintos tipos de muestreo que se denomina muestreo polietápico (bi, tri, etc, *etápico*).

✱ *Un muestreo bi-etápico es aquel donde primero se estratifica a la población y luego se selecciona dentro de cada estrato un MAS.*

✱ *Un muestreo tri-etápico puede consistir en seleccionar aleatoriamente conglomerados, luego hacer una estratificación y luego sacar un MAS dentro de cada estrato.*

- También se realizan combinaciones de muestreo paralelas: por ejemplo, una muestra bi-etápica de conglomerados y MAS por un lado y una doble muestra aleatoria sucesiva de conglomerados finalizando por un MAS.

✱ *Tal es el caso de la ECH donde se siguen procedimientos de muestreo diferentes para Montevideo y para el Interior Urbano.*

✱ *Por lo general, estos muestreos complejos son evitados porque conllevan complejos procedimientos de inferencia para cada uno de los parámetros que han de utilizar.*

SELECCIÓN DE LOS CASOS (II): Muestra estadística

Cuadro I
Esquema de los tipos de muestreo y sus principales características


Muestreo Aleatorio Simple	Muestreo Aleatorio sistemático	Muestreo Aleatorio Estratificado	Muestreo Aleatorio por Conglomerados
Una vez definido el marco muestral, cada elemento tiene la misma probabilidad de ser seleccionado para integrar la muestra	Una vez definido el marco muestral, se numeran los elementos y se selecciona cada componente de la muestra cada J veces.	Implica una división del universo en sub-universos, llamados estratos, siguiendo una o varias variables teóricamente relevantes y posteriormente tomando un MAS dentro de cada estrato. El estrato es creado por el investigador, aunque puede pre-existir como fenómeno social. En tal caso, conviene tratar esa circunstancia como MAC. El tamaño de los estratos en la muestra no tiene necesariamente que ser igual que en la población.	La unidad primaria de muestreo (UPM) es una agrupación de los elementos de interés (USM). Pueden ser conglomerados de personas, hogares, organizaciones, de actividades, de documentos, etc, dependiendo cuál sea la unidad de análisis. A su vez, los conglomerados en sí pueden ser organizaciones, barrios, pueblos, ciudades, lugares (por ejemplo, un aeropuerto)
Todas las inferencias estadísticas se realizan por lo general suponiendo este muestreo		La inferencia se realiza en dos pasos: primero dentro de cada estrato se calculan los estimadores y luego se procede a la estimación total. Los procedimientos para hacer las inferencias totales dependerán de como se distribuyen los estratos en la población y en la muestra. Es frecuente la corrección muestral utilizando ponderadores	Las inferencias son complejas, dado que no existe independencia en la elección de los elementos últimos (USM). Los errores estándares son mayores que en el caso del MAS. Hay que corregir por conglomerados los métodos multivariados (por ejemplo, en regresión).
Puede utilizarse sólo o en combinación con el MAEL y el MAC	Es muy utilizado en muestreos polietápicos de hogares donde a nivel de la manzana se marca un lugar de comienzo y se le dice "entre cada 4 casas".	Supone que la estratificación permite descomponer la varianza en la variable dependiente de tal forma que la intervarianza (entre estratos) representa una alta proporción de la varianza total	Se utiliza generalmente por razones económicas, dado que se encuentran las unidades reunidas en un espacio físico delimitado y pequeño.

 En el tipo más simple de todos, la fórmula de cálculo del tamaño de la muestra es:

$$n = \frac{Z^2 \sigma^2}{d^2}$$

SELECCIÓN DE LOS CASOS (II): Muestra estadística

- ↔ Todos los tipos más complejos de muestreo se derivan de este.
- ↔ Esta fórmula se aplica directamente así cuando:
 - ✓ el universo puede definirse como una población infinita, o en términos prácticos, cuando el número de elementos supera los cien mil ($N > 100.000$)
 - ✓ o se utiliza el muestreo con reposición .
- ↔ En la fórmula, cada uno de los elementos significa:
 - ✓ d^2 es el cuadrado de la **diferencia máxima aceptable** en la estimación de los parámetros utilizando los estadísticos muestrales.
 - * Este punto depende del marco teórico, del presupuesto que se tenga y sobre todo de los objetivos de la investigación.
 - * Resulta de la resta entre cualquier par de hipotéticos pares de valores que se tengan. Si se trata de porcentajes, por ejemplo, se expresa en términos de proporciones.
 - * Una $d= 0.05$ significa que la máxima diferencia que se aceptará entre el estimador “ p ” y el parámetro “ P ” equivale a cinco puntos porcentuales.
 - * Tiene una relación cuadrática negativa con el tamaño de la muestra.
 - ↔ Z^2 es el cuadrado del valor típico relacionado con la confianza establecida por el investigador.
 - * Didácticamente, una confianza del 95% por ejemplo, indica que el investigador espera que de repetir 100 veces la extracción de una muestra y el cálculo del estadístico, esperará que en 95 de las muestras el valor estimado para el parámetro de interés se encuentre en un intervalo “tal y tal”.
 - * La confianza Resulta de la tabla de la distribución de Z.
 - * El intervalo de **confianza** es el único puente estadístico entre el estimador y el parámetro poblacional. Se basan en las estimaciones del error estándar. Pero sólo es posible calcular errores estándares en las muestras estadísticas.
 - ↔ σ^2 es la varianza en la población de la variable de interés. Este es el único dato empírico que se requiere para determinar el tamaño de la muestra.
 - ↔ Conceptualmente, el tamaño de una muestra está directamente relacionado al cuadrado de la varianza, al cuadrado de la confianza y tiene una relación inversa al cuadrado de la diferencia máxima aceptable.

 Es frecuente encontrar la creencia que la muestra debería tener cierta relación proporcional con la población. **Esto es un error** que puede fácilmente despejarse haciendo la observación de que en la fórmula para calcular el tamaño de la muestra, para

SELECCIÓN DE LOS CASOS (II): Muestra estadística

nada entra el tamaño de la población.

- El único dato empírico sobre la población que se requiere es la heterogeneidad o dispersión en la o las variables de interés, medida a través de la varianza.
- El tamaño de la población, cuando esta es pequeña, impone un recálculo de la muestra, ajustando por el factor de corrección para muestras finitas. En la siguiente fórmula, se introdujo el tamaño n_0 de la muestra calculado según el MAS y se ajusta, de tal forma que al final las muestras se reducen en forma significativa, de acuerdo al tamaño de la población.

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

✱ Conceptualmente, cuando mayor es la fracción de muestreo, menor será el tamaño de la muestra final corregida respecto a la muestra inicial.

- Fracción de muestreo: es el cociente entre el tamaño de la muestra y el tamaño de la población.

$$FM = n / N$$

✱ Es común que en el tipo de muestreo estratificado, se utilicen distintas fracciones de muestreo para cada estrato (denominado como "O" por observados). Por la sencilla razón de que la distribución de los elementos en el Universo para cada estrato no es exactamente igual. De aquí tendremos diferentes fracciones de muestreo:

$$FM_U = N_U / N$$

✱ También puede suceder que la distribución de los elementos en el universo y en la muestra sean deliberadamente distintos. Esto sucede por ejemplo, cuando por alguna razón sustantiva se ha deseado sobre-representar a un estrato incrementando la cantidad de "cupos" que le corresponderán en la muestra. Por ejemplo, en las evaluaciones de aprendizaje de la educación primaria de Uruguay, generalmente se sobre-representa el estrato rural, el cual de otra forma tendría muy pocos niños, insuficientes como para realizar estimaciones precisas. Esto nos conduce a tener tantos cocientes observados como estratos.

$$w_O = n_O / n$$

✱ En tal caso, esta discrepancia se debe utilizar para construir el **ponderador** de cada estrato

SELECCIÓN DE LOS CASOS (II): Muestra estadística

y se deberá utilizar para hacer las estimaciones de los parámetros.

* Para el cálculo del ponderador se utiliza la fracción de muestreo en el universo (w_U) y la fracción de muestreo observada en la muestra donde se ha alterado el (w_o).

$$w_h = \frac{\frac{n_U}{N}}{\frac{n_o}{N}} = \frac{n_U}{N} * \frac{N}{n_o} = \frac{n_U}{n_o}$$

* Cuando las fracciones de muestreo o ponderadores de los estratos son iguales al peso que el estrato tiene en el universo, se dice que la muestra es auto-ponderada ya que la afijación de los elementos en la muestra se ha hecho en forma proporcional .

SELECCIÓN DE LOS CASOS (II): Muestra estadística

Cuadro I
Cálculo del tamaño de la muestra bajo la asignación aleatoria simple: población infinita y población finita

Confianza en la estimación Z		Diferencia máxima aceptable D		Tamaño de Población	Fracción de muestreo	
expresada en valores de t de student o de z		(expresado como diferencia entre porcentajes)		N	n/N	Tamaño Final corregido
2	4	0.05	0.0025	500	1000	Población finita
P	I-P	Varianza	Numerador	Muestra	Fracción	Muestra
0,10	0,90	0,090	0,36	144	0,144	125,87
0,15	0,85	0,128	0,51	204	0,204	169,44
0,20	0,80	0,160	0,64	256	0,256	203,82
0,25	0,75	0,188	0,75	300	0,300	230,77
0,30	0,70	0,210	0,84	336	0,336	251,50
0,35	0,65	0,228	0,91	364	0,364	266,86
0,40	0,60	0,240	0,96	384	0,384	277,46
0,45	0,55	0,248	0,99	396	0,396	283,67
0,50	0,50	0,250	1	400	0,400	285,71
0,55	0,45	0,248	0,99	396	0,396	283,67
0,60	0,40	0,240	0,96	384	0,384	277,46
0,65	0,35	0,228	0,91	364	0,364	266,86
0,70	0,30	0,210	0,84	336	0,336	251,50
0,75	0,25	0,188	0,75	300	0,300	230,77
0,80	0,20	0,160	0,64	256	0,256	203,82
0,85	0,15	0,128	0,51	204	0,204	169,44
0,90	0,10	0,090	0,36	144	0,144	125,87