

TALLER DE MUESTREO

PROF. ELIZABETH GANDICA DE ROA

lizgandi@hotmail.com

INTRODUCCIÓN

Una parte fundamental para realizar un estudio estadístico de cualquier tipo es obtener unos resultados confiables y que puedan ser aplicables. Resulta casi imposible o impráctico llevar a cabo algunos estudios sobre toda una población, por lo que la solución es llevar a cabo el estudio basándose en un subconjunto de ésta denominada muestra.

Sin embargo, para que los estudios tengan la validez y confiabilidad buscada es necesario que tal subconjunto de datos, o muestra, posea algunas características específicas que permitan, al final, generalizar los resultados hacia la población en total. Esas características tienen que ver principalmente con el tamaño de la muestra y con la manera de obtenerla.

IMPORTANCIA DEL MUESTREO

A lo largo del curso se hacen uso de dos tipos de razonamiento: el **deductivo** y el **inductivo**. El primero está relacionado directamente con la teoría de probabilidad, que se aborda en la unidad 4, y que a partir de las características de la población se obtienen las posibles características de una muestra. El segundo tipo de razonamiento se relaciona con la denominada **inferencia estadística**: utilizar las características de un subconjunto de la población (la muestra) para hacer afirmaciones (inferir) sobre la población en general. Éste será el caso de esta unidad.

El muestreo, como ya se mencionó, implica algo de incertidumbre que debe ser aceptada para poder realizar el trabajo, pues aparte de que estudiar una población resulta ser un trabajo en ocasiones demasiado grande, existen otras razones extras:

- **Recursos limitados.** Es decir, no existen los recursos humanos, materiales o económicos para realizar el estudio sobre el total de la población. Es como cuando se compra un aparato, un automóvil usado (por ejemplo), que se prueba unos minutos (el encendido, una carrerita, etc.) para ver si funciona correctamente y luego se adquiere, pero no se espera a probarlo toda la vida (encendiéndolo y apagándolo o, simplemente, dejándolo encendida) antes de realizar la adquisición.
- **Escasez.** Es el caso en que se dispone de una sola muestra. Por ejemplo, para el estudio paleontológico de los dinosaurios (el T. Rex por ejemplo) sería muy bueno contar con, al menos, muchos restos fósiles y así realizar tales investigaciones; sin embargo, se cuenta sólo con una docena de esqueletos fosilizados (casi todos incompletos) de esas criaturas en todo el mundo.

- **Pruebas destructivas.** Es el caso en el que realizar el estudio sobre toda la población llevaría a la destrucción misma de la población. Por ejemplo, si se quisiese saber el conteo exacto de hemoglobina de una persona habría que extraerle **toda** la sangre.
- **El muestreo puede ser más exacto.** Esto es en el caso en el que el estudio sobre la población total puede causar errores por su tamaño o, en el caso de los censos, que sea necesario utilizar personal no lo suficientemente capacitado; mientras que, por otro lado, el estudio sobre una muestra podría ser realizada con menos personal pero más capacitado.

MUESTREO

Es un procedimiento por medio del cual se estudia una parte de la población llamada muestra, con el objetivo de inferir con respecto a toda la población.

Es importante relacionar el muestreo con lo que es el censo, el cual se define como la enumeración completa de todos los elementos de la población de interés.

VENTAJAS DEL MUESTREO:

- a) Costos reducidos.
- b) Mayor rapidez para obtener resultados.
- c) Mayor exactitud o mejor calidad de la información: debido a los siguientes factores
 - c.1 Volumen de trabajo reducido.
 - c.2 Puede existir mayor supervisión en el trabajo.
 - c.3 Se puede dar más entrenamiento al personal.
 - c.4 Menor probabilidad de cometer errores durante el procesamiento de la información.
- d) Factibilidad de hacer el estudio cuando la toma de datos implica técnicas destructivas, por ejemplo:
 - Pruebas de germinación.
 - Análisis de sangre.
 - Control de calidad.

VENTAJAS DEL CENSO:

Sin embargo, también se debe mencionar que el censo tiene algunas ventajas que son las siguientes:

- a) Existe una cobertura total.
- b) Tiene aceptación pública.
- c) No se requieren grandes conocimientos de estadística.

TAMAÑO DE LAS MUESTRA

Para calcular el tamaño de una muestra hay que tomar en cuenta tres factores:

1. El porcentaje de confianza con el cual se quiere generalizar los datos desde la muestra hacia la población total.
2. El porcentaje de error que se pretende aceptar al momento de hacer la generalización.
3. El nivel de variabilidad que se calcula para comprobar la hipótesis.

La **confianza** o el **porcentaje de confianza** es el porcentaje de seguridad que existe para generalizar los resultados obtenidos. Esto quiere decir que un porcentaje del 100% equivale a decir que no existe ninguna duda para generalizar tales resultados, pero también implica estudiar a la totalidad de los casos de la población.

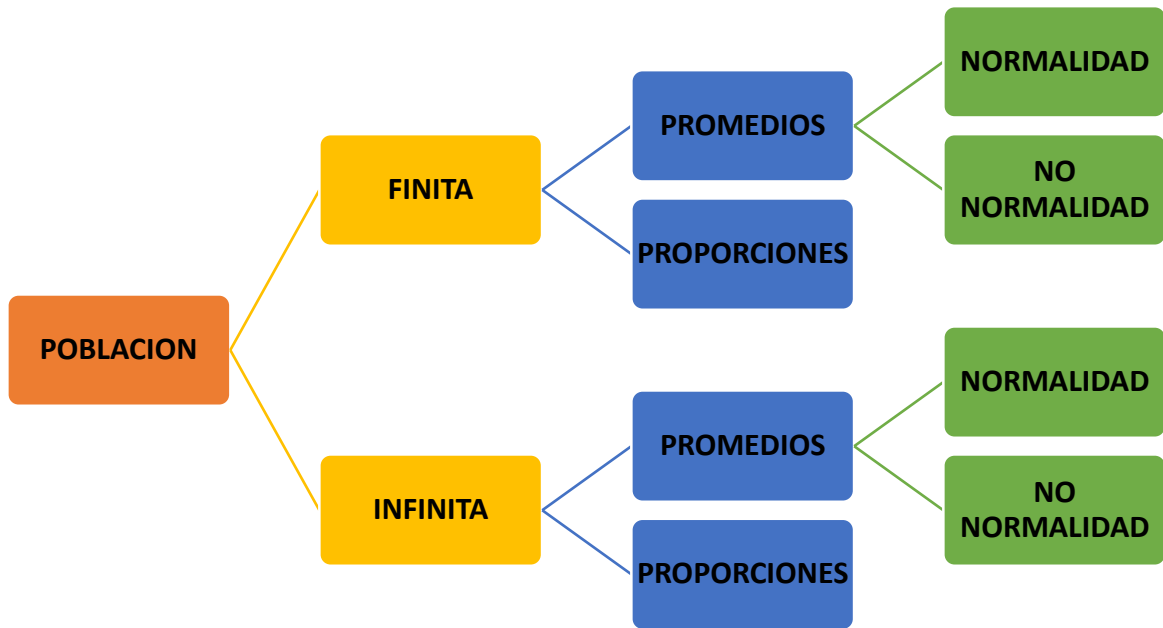
Para evitar un costo muy alto para el estudio o debido a que en ocasiones llega a ser prácticamente imposible el estudio de todos los casos, entonces se busca un porcentaje de confianza menor. Comúnmente en las investigaciones sociales se busca un 95%.

El **error** o **porcentaje de error** equivale a elegir una probabilidad de aceptar una hipótesis que sea falsa como si fuera verdadera, o la inversa: rechazar a hipótesis verdadera por considerarla falsa. Al igual que en el caso de la confianza, si se quiere eliminar el riesgo del error y considerarlo como 0%, entonces la muestra es del mismo tamaño que la población, por lo que conviene correr un cierto riesgo de equivocarse.

Comúnmente se aceptan entre el 4% y el 6% como error, tomando en cuenta de que **no** son complementarios la confianza y el error.

La **variabilidad** es la probabilidad (o porcentaje) con el que se aceptó y se rechazó la hipótesis que se quiere investigar en alguna investigación anterior o en un ensayo previo a la investigación actual.

FORMULAS PARA CALCULAR EL TAMAÑO DE LA MUESTRA



FORMULAS DE MUESTREO PARA POBLACIONES FINITAS E INFINITAS

Población finita e infinita. Estimación de promedios. Normalidad

$$n = \frac{NZ^2\sigma^2}{(N-1)E^2 + Z^2\sigma^2}$$

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

Población finita e infinita. Estimación de promedios. No Normalidad

$$n = \frac{\sigma^2}{E^2 \alpha}$$

Población finita e infinita. Estimación de proporciones.

$$n = \frac{NZ^2pq}{(N-1)E^2 + Z^2pq}$$

$$n = \frac{Z^2pq}{E^2}$$

TIPOS DE MUESTREO

Muestreo Aleatorio Simple: Es la forma más común de obtener una muestra en la selección al azar, es decir, cada uno de los individuos de una población tiene la misma posibilidad de ser elegido. Si no se cumple este requisito, se dice que la muestra es viciada. Para tener la seguridad de que la muestra aleatoria no es viciada, debe emplearse para su constitución una tabla de números aleatorios. Este procedimiento, atractivo por su simpleza, tiene poca o nula utilidad práctica cuando la población que estamos manejando es muy grande. Ejemplo:

Supongamos que nos interesa elegir una muestra aleatoria de 5 estudiantes en un grupo de estadística de 20 alumnos. 20C5 da el número total de formas de elegir una muestra no ordenada y este resultado es 15,504 maneras diferentes de tomar la muestra. Si listamos las 15,504 en trozos separados de papel, una tarea tremenda, luego los colocamos en un recipiente y después los revolvemos, entonces podremos tener una muestra aleatoria de 5 si seleccionamos un trozo de papel con cinco nombres. Un procedimiento más simple para elegir una muestra aleatoria sería escribir cada uno de los 20 nombres en pedazos separados de papel, colocarlos en un recipiente, revolverlos y después extraer cinco papeles al mismo tiempo.

Otro método para obtener una muestra aleatoria de 5 estudiantes en un grupo de 20 utiliza una tabla de números aleatorios. Se puede construir la tabla usando una calculadora o una computadora. También se puede prescindir de estas y hacer la tabla escribiendo diez dígitos del 0 al 9 en tiras de papel, las colocamos en un recipiente y los revolvemos, de ahí, la primera tira seleccionada determina el primer número de la tabla, se regresa al recipiente y después de revolver otra vez se selecciona la segunda tira que determina el segundo número de la tabla; el proceso continúa hasta obtener una tabla de dígitos aleatorios con tantos números como se desee.

Hay muchas situaciones en las cuales el muestreo aleatorio simple es poco práctico, imposible o no deseado; aunque sería deseable usar muestras aleatorias simples para las encuestas nacionales de opinión sobre productos o sobre elecciones presidenciales, sería muy costoso o tardado.

Es recomendable:

- Cuando la población no es numerosa y las unidades se concentran en un área pequeña
- La característica no debe tener gran variabilidad
- La población debe facilitar su enumeración para que permita la aplicación de este método

Inconvenientes:

- Se requiere un listado de unidades de la población, la cual no es fácil, en muchos casos
- Si el área es amplia, es probable que haya necesidad de un traslado a lugares lejanos, para investigar unas pocas unidades, dificultando su aplicación (costo y tiempo)
- No existe garantía de que todas las unidades queden representadas en la muestra, cuando la característica tiene gran variabilidad.
- Si el coeficiente de variación es mayor a 30%, el tamaño de la muestra se va haciendo más grande.

MUESTREO ALEATORIO SIMPLE- PROMEDIOS



$$n = \frac{NZ^2\sigma^2}{(N-1)E^2 + Z^2\sigma^2}$$

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

$$n = \frac{\sigma^2}{E^2 \alpha}$$

Muestreo Aleatorio Sistemático: Es una técnica de muestreo que requiere de una selección aleatoria inicial de observaciones seguida de otra selección de observaciones obtenida usando algún sistema o regla. Ejemplo:

Para obtener una muestra de suscriptores telefónicos en una ciudad grande, puede obtenerse primero una muestra aleatoria de los números de las páginas del directorio telefónico; al elegir el vigésimo nombre de cada página obtendríamos un muestreo sistemático, también podemos escoger un nombre de la primera página del directorio y después seleccionar cada nombre del lugar número cien a partir del ya seleccionado. Por ejemplo, podríamos seleccionar un número al azar entre los primeros 100; supongamos que el elegido es el 40, entonces seleccionamos los nombres del directorio que corresponden a los números 40, 140, 240, 340 y así sucesivamente.

EJEMPLO DE MUESTREO SISTEMATICO

$$k = \frac{N}{n} = \frac{2000}{100} = 20$$



Hasta completar los 100

Muestreo Aleatorio Estratificado: Una muestra es estratificada cuando los elementos de la muestra son proporcionales a su presencia en la población. La presencia de un elemento en un estrato excluye su presencia en otro. Para este tipo de muestreo, se divide a la población en varios grupos o estratos con el fin de dar representatividad a los distintos factores que integran el universo de estudio. Para la selección de los elementos o unidades representantes, se utiliza el método de muestreo aleatorio.

En síntesis, requiere de separar a la población según grupos llamados estratos, y de elegir después una muestra aleatoria simple en cada estrato. La información de las muestras aleatorias simples de cada estrato constituiría entonces una muestra global. Ejemplo:

Supongamos que nos interesa obtener una muestra de las opiniones de los profesores de una gran universidad. Puede ser difícil obtener una muestra con todos los profesores, así que supongamos que elegimos una muestra aleatoria de cada colegio, o departamento académico; los estratos vendrían a ser los colegios, o departamentos académicos.

La escogencia final en cada estrato se puede realizar con:

- Asignación igual: se da cuando los elementos quedan repartidos por igual en cada estrato
- Asignación proporcional: se da cuando extraigo el mismo porcentaje de cada estrato

- Asignación óptima: se da cuando el tamaño de la muestra depende de la variabilidad del estrato y del costo mínimo para una precisión dada.

MUESTREO ALEATORIO ESTRATIFICADO PROMEDIOS POBLACION FINITA-NORMALIDAD

$$n = \frac{NZ^2\sigma^2}{(N-1)E^2 + Z^2\sigma^2}$$

$$\sigma^2 = \sum(w_i * \sigma_i^2)$$

$$\bar{x} = \sum(w_i * \bar{x}_i)$$

σ_i^2 = varianza en cada estrato

W_i = peso de cada estrato

\bar{x}_i = media de cada estrato

MUESTREO ESTRATIFICADO PROPORCIONES POBLACION FINITA

$$n = \frac{NZ^2pq}{(N-1)E^2 + Z^2pq}$$

$$p * q = \sum(w_i * p_i * q_i)$$

Muestreo Aleatorio por Área o Conglomerado: Requiere de elegir una muestra aleatoria simple de unidades heterogéneas entre sí de la población llamadas conglomerados. Cada elemento de la población pertenece exactamente a un conglomerado, y los elementos dentro de cada conglomerado son usualmente heterogéneos o disímiles. Ejemplo:

Supongamos que una compañía de servicio de televisión por cable está pensando en abrir una sucursal en una ciudad grande; la compañía planea realizar un estudio para determinar el porcentaje de familias que utilizarían sus servicios, como no es práctico preguntar en cada casa, la empresa decide seleccionar una parte de la ciudad al azar, la cual forma un conglomerado.

En el muestreo por conglomerados, éstos se forman para representar, tan fielmente como sea posible, a toda la población; entonces se usa una muestra aleatoria simple de conglomerados para estudiarla. Los estudios de instituciones sociales como iglesias, hospitales, escuelas y prisiones se realizan, generalmente, con base en el muestreo por conglomerados.

MUESTREOS NO PROBABILÍSTICOS

Muestreo Discrecional: A criterio del investigador los elementos son elegidos sobre lo que él cree que pueden aportar al estudio. Ejemplo.: muestreo por juicios; cajeros de un banco o un supermercado; etc.

Muestreo Doble: Bajo este tipo de muestreo, cuando el resultado del estudio de la primera muestra no es decisivo, una segunda muestra es extraída de la misma población. Las dos muestras son combinadas para analizar los resultados. Este método permite a una persona principiar con una muestra relativamente pequeña para ahorrar costos y tiempo. Si la primera muestra arroja un resultado definitivo, la segunda muestra puede no necesitarse. Por ejemplo, al probar la calidad de un lote de productos manufacturados, si la primera muestra arroja una calidad muy alta, el lote es aceptado; si arroja una calidad muy pobre, el lote es rechazado. Solamente si la primera muestra arroja una calidad intermedia, será requerida la segunda muestra.

Muestreo Múltiple: El procedimiento bajo este método es similar al expuesto en el muestreo doble, excepto que el número de muestras sucesivas requerido para llegar a una decisión es más de dos muestras.

Muestreo Opinático o Intencional: Este tipo de muestreo se caracteriza por un esfuerzo deliberado de obtener muestras "representativas" mediante la inclusión en la muestra de grupos supuestamente típicos. Es muy frecuente su utilización en sondeos preelectorales de zonas que en anteriores votaciones han marcado tendencias de voto.

Muestreo Casual o Incidenta: Se trata de un proceso en el que el investigador selecciona directa e intencionadamente los individuos de la población. El caso más frecuente de este procedimiento es el utilizar como muestra los individuos a los que se tiene fácil acceso (los profesores de universidad emplean con mucha frecuencia a sus propios alumnos). Un caso particular es el de los voluntarios.

RAZONES PARA TOMAR MUESTRAS

1- No estamos interesados realmente en todos los elementos: sino sólo en algunos ejemplares o casos de la población. Este tipo habitual de investigación no es de hecho un estudio de muestreo, sino un estudio de caso ampliado.

2- Estamos interesados por igual en todos los elementos de la población y querríamos estudiarlos todos. Pero por razones prácticas, tendremos que escoger solo una muestra. Tal vez tenemos una población de millones de objetos y es imposible abarcar incluso una mayoría de entre ellos. También en aquellos casos (con poblaciones, digamos, de 10.000) en que podríamos escoger estudiar cada objeto, el estudio de muestreo puede ser una elección prudente, porque ahorra tiempo y podemos usar el tiempo ahorrado para estudiar los elementos más cuidadosamente. Todos estos son buenos casos para un estudio de muestreo.

En la investigación de muestreo estamos siempre interesados no en la muestra sino en la población; más exactamente, en los atributos de los elementos de la población. Cuando estamos estudiando los elementos del ejemplo querríamos escoger elementos que tengan los mismos atributos que la media de la población. Si ese es el caso, nuestra muestra es representativa.

Este es el caso ideal, pero en la práctica no tenemos medio de saber si los elementos son representativos en realidad; el cálculo de probabilidades nos dice que en la mayor parte de los casos habrá algunas diferencias entre la muestra y la población. La diferencia se llama sesgo, y en alguna medida casi siempre está presente en la muestra, simplemente por el carácter accidental del muestreo.

Sin embargo, si tenemos razones para sospechar de la presencia de un sesgo sistemático en la muestra, debiéramos siempre intentar encontrar cuál es y ver si puede ser eliminado.

ERROR DE MUESTREO Y ERROR NO MUESTRAL

Diferencia entre los siguientes términos: Error de Muestreo y Error no de Muestreo.

La diferencia radica en los tipos de errores que son medidos o detectados en los resultados que arrojan las encuestas. Mientras que el Error de Muestreo señala desde las preguntas mal

redactadas por los entrevistadores en las encuestas, indisposición por parte de los entrevistados y cálculos errados; el Error no de Muestreo localiza la información falsa suministrada por los entrevistados.

En síntesis un Error de Muestreo usualmente ocurre cuando no se lleva a cabo la encuesta completa de la población y, los Errores no de Muestreo pueden ocurrir en una encuesta completa de la población.

PRECISION Y EXACTITUD

La Precisión: Se refiere a la identidad o por lo menos a la similitud entre dos o más mediciones de la misma cantidad. En cierto grado, la precisión está relacionada con la estabilidad de la técnica del experimentador, que puede necesitar más mejoras de las que él cree. Sin embargo, dependiendo de la naturaleza de la medición particular que se está considerando, puede aparecer una falta de precisión debido a un control defectuoso de temperatura, a una pieza de vidrio astillada, corroídas o flojas en los instrumentos utilizados.

La Exactitud: Se refiere a la cercanía a su valor verdadero de las mediciones obtenidas. Para un procedimiento dado, la exactitud se estima llevando a cabo mediciones físicas o químicas de un patrón conocido. Por ejemplo, si una investigación depende de una titulación con una solución normalizada de álcali, es conveniente comprobar la confiabilidad de este reactivo de tiempo en tiempo, titulando una cantidad conocida o determinada gravimétricamente de una sal de un ácido.