

Distribuciones muestrales.
Distribución muestral de
Medias

TEORIA DEL MUESTREO

Uno de los propósitos de la estadística inferencial es estimar las características poblacionales desconocidas, examinando la información obtenida de una muestra, de una población. El punto de interés es la muestra, la cual debe ser representativa de la población objeto de estudio.

Se seguirán ciertos procedimientos de selección para asegurar de que las muestras reflejen observaciones a la población de la que proceden, ya que solo se pueden hacer observaciones probabilísticas sobre una población cuando se usan muestras representativas de la misma.

Una **población** está formada por la totalidad de las observaciones en las cuales se tiene cierto observa.

Una **muestra** es un subconjunto de observaciones seleccionadas de una población.

A continuación se verá algunos usos del muestreo en diversos campos:

1. *Política*. Las muestras de las opiniones de los votantes se usan para que los candidatos midan la opinión pública y el apoyo en las elecciones.
2. *Educación*. Las muestras de las calificaciones de los exámenes de estudiantes se usan para determinar la eficiencia de una técnica o programa de enseñanza.
3. *Industria*. Muestras de los productos de una línea de ensamble sirve para controlar la calidad.
4. *Medicina*. Muestras de medidas de azúcar en la sangre de pacientes diabéticos prueban la eficacia de una técnica o de un fármaco nuevo.
5. *Agricultura*. Las muestras del maíz cosechado en una parcela proyectan en la producción los efectos de un fertilizante nuevo.
6. *Gobierno*. Una muestra de opiniones de los votantes se usaría para determinar los criterios del público sobre cuestiones relacionadas con el bienestar y la seguridad nacional.

Errores en el Muestreo

Cuando se utilizan valores muestrales (**parámetros**), o **estadísticos** para estimar valores poblacionales, pueden ocurrir dos tipos generales de errores:

El error muestral y
El error no muestral.

El error muestral se refiere a la variación natural existente entre muestras tomadas de la misma población. Aún si se ha tenido gran cuidado para asegurar que dos muestras del mismo tamaño sean representativas de una cierta población, no esperaríamos que las dos sean idénticas en todos sus detalles. El error muestral es un concepto importante que ayudará a entender mejor la naturaleza de la estadística inferencial.

Los errores que surgen al tomar las muestras y que no pueden clasificarse como errores muestrales y se denominan errores **no muestrales**.

El sesgo de las muestras es un tipo de error no muestral. El sesgo muestral se refiere a una tendencia sistemática inherente a un método de muestreo que da estimaciones de un parámetro que son, en promedio, menores (sesgo negativo), o mayores (sesgo positivo) que el parámetro real. Ejemplo: la longitud del dedo índice de personas de la misma edad y sexo.

El sesgo muestral puede suprimirse, o minimizarse, usando la **aleatorización**.

La aleatorización se refiere a cualquier proceso de selección de una muestra de la población en el que la selección es imparcial o no está sesgada; una muestra elegida con procedimientos aleatorios se llama **muestra aleatoria**.

Los tipos más comunes de técnicas de muestreo aleatorios son:

- el muestreo aleatorio simple,
- el muestreo estratificado,
- el muestreo por conglomerados y
- el muestreo sistemático.

Si una muestra aleatoria se elige de tal forma que todos los elementos de la población tengan la misma probabilidad de ser seleccionados, la llamamos muestra aleatoria simple.

Ejemplo 1

Suponga que nos interesa elegir una muestra aleatoria de 5 estudiantes en un grupo de estadística de 20 alumnos. Las combinaciones se escriben ${}_{20}C_5$ lo que da el número total de formas de elegir una muestra no ordenada y este resultado es igual a 15,504 maneras diferentes de tomar la muestra. Un procedimiento 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 o con métodos de selección al azar.

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.

El **muestreo estratificado** requiere de separar a la población según grupos que no se traslapen 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 2

Suponga 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.

El **muestreo por conglomerados** 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 3

Suponga 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.

El **muestreo 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 4

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.

En este caso, 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.

Distribuciones Muestrales

Las muestras aleatorias obtenidas de una población son, por naturaleza propia, impredecibles. No se esperaría que dos muestras aleatorias del mismo tamaño y tomadas de la misma población tenga la misma media muestral o que sean completamente parecidas; puede esperarse que cualquier estadístico, como la media muestral, calculado a partir de las medias en una muestra aleatoria, cambie su valor de una muestra a otra, por ello, se quiere estudiar la distribución de todos los valores posibles de un estadístico.

Tales distribuciones serán muy importantes en el estudio de la estadística inferencial, porque las inferencias sobre las poblaciones se harán usando estadísticas muestrales.

Con el análisis de las distribuciones asociadas con los estadísticos muestrales, podremos juzgar la confiabilidad de un estadístico muestral como un instrumento para hacer inferencias sobre un parámetro poblacional desconocido.

Como los valores de un estadístico, tal como la media, varían de una muestra aleatoria a otra, se le puede considerar como una variable aleatoria con su correspondiente distribución de frecuencias.

La distribución de frecuencia de un estadístico muestral se denomina **distribución muestral**.

En general, la distribución muestral de un estadístico es la de todos sus valores posibles calculados a partir de muestras del mismo tamaño.

Distribución Muestral de Medias

Suponga que se han seleccionado muestras aleatorias de tamaño 20 en una población grande. Se calcula la media muestral \bar{x} para cada muestra; la colección de todas estas medias muestrales recibe el nombre de **distribución muestral de medias**, lo que se puede ilustrar en la siguiente figura:

