

8. Muestra estadística, Muestreo en poblaciones finitas, Error muestral, Estrategias de muestreo y Muestreo estratificado.

8.1 Muestra estadística

En estadística, una muestra es un subconjunto de casos o individuos de una población estadística.

Las muestras se obtienen con la intención de inferir propiedades de la totalidad de la población, para lo cual deben ser representativas de la misma. Para cumplir esta característica la inclusión de sujetos en la muestra debe seguir una técnica de muestreo. En tales casos, puede obtenerse una información similar a la de un estudio exhaustivo con mayor rapidez y menor coste (véanse las ventajas de la elección de una muestra, más abajo).

Por otra parte, en ocasiones, el muestreo puede ser más exacto que el estudio de toda la población porque el manejo de un menor número de datos provoca también menos errores en su manipulación. En cualquier caso, el conjunto de individuos de la muestra son los sujetos realmente estudiados.

El número de sujetos que componen la muestra suele ser bastante inferior a la población total, aunque suficiente grande como para que la estimación de los parámetros determinados tenga un nivel de confianza adecuado. Para que el tamaño de la muestra sea idóneo es preciso recurrir a su cálculo.

8.1.1 Otras definiciones relacionadas

8.1.1.1 Espacio muestral

El espacio muestral del que se toma una muestra concreta está formado por el conjunto de todas las posibles muestras que se pueden extraer de una población mediante una determinada técnica de muestreo.

8.1.1.2 Parámetro o Estadístico muestral

Un parámetro estadístico o simplemente un estadístico muestral es cualquier valor calculado a partir de la muestra, como por ejemplo la media, varianza o una proporción, que describe a una población y puede ser estimado a partir de una muestra. Un estadístico muestral es un tipo de variable aleatoria, y que como tal, tiene una distribución de probabilidad concreta, frecuentemente caracterizada por un conjunto finito de parámetros.

8.1.1.3 Estimación

Una estimación es cualquier técnica para conocer un valor aproximado de un parámetro referido a la población, a partir de los estadísticos muestrales calculados a partir de los elementos de la muestra. Si se estima el suficiente número de parámetros puede aproximarse de manera razonable la distribución de probabilidad de la población para ciertas variables aleatorias.

8.1.1.4 Nivel de confianza

El nivel de confianza de una aseveración basada en la inferencia estadística es una medida de la bondad de la estimación realizada a partir de estadísticos muestrales. Usualmente se usan niveles de confianza para intervalos de confianza o bien p-valores que miden la probabilidad de errores de tipo I (probabilidad de rechazar una cierta hipótesis siendo que esta era correcta).

8.1.1.5 Ejemplo

Se tiene una población de 222.222 habitantes y se quiere conocer cuantos de ellos son hombres y cuantos de ellos son mujeres. Se conjetura que cerca del 50% son mujeres y el resto hombres, pero se quiere seleccionar una muestra para determinar cuantos hombres y mujeres hay en la muestra y a partir de ahí inferir el porcentaje exacto de hombres y mujeres en la población total. La descripción de una muestra, y los resultados obtenidos sobre ella, puede ser del tipo mostrado en el siguiente ejemplo:

Dimensión de la población:	222.222 habitantes
Probabilidad del evento:	Hombre o Mujer 50%
Nivel de confianza:	90%
Desviación tolerada:	5%
Resultado	196
Tamaño de la muestra:	270

La interpretación de esos datos sería la siguiente:

1. La **población** a investigar tiene 222.222 habitantes y queremos saber cuántos son hombres o mujeres.
2. Estimamos en un 50% para cada sexo y para el propósito del estudio es suficiente un 90% de seguridad con un nivel entre $90 - 5$ y $90 + 5$.

3. Generamos una tabla de 270 números al azar entre 1 y 222.222 y en un censo numerado comprobamos el género para los seleccionados.

8.1.1.6 Ventajas de la elección de una muestra

El estudio de muestras es preferible, en la mayoría de los casos, por las siguientes razones:

1. Si la población es muy grande (en ocasiones, infinita, como ocurre en determinados experimentos aleatorios) y, por tanto, imposible de analizar en su totalidad.
2. Las características de la población varían si el estudio se prolonga demasiado tiempo.
3. **Reducción de costos:** al estudiar una pequeña parte de la población, los gastos de recogida y tratamiento de los datos serán menores que si los obtenemos del total de la población.
4. **Rapidez:** al reducir el tiempo de recogida y tratamiento de los datos, se consigue mayor rapidez.
5. **Viabilidad:** la elección de una muestra permite la realización de estudios que serían imposible hacerlo sobre el total de la población.
6. La población es suficientemente homogénea respecto a la característica medida, con lo cual resultaría inútil malgastar recursos en un análisis exhaustivo (por ejemplo, muestras sanguíneas).
7. El proceso de estudio es destructivo o es necesario consumir un artículo para extraer la muestra (ejemplos: vida media de una bombilla, carga soportada por una cuerda, precisión de un proyectil, etc.).

8.1.1.7 Descripción matemática de una muestra aleatoria

El uso de muestras para deducir fiablemente características de la población requiere que se trate con **muestras aleatorias**. Si la muestra estadística considerada no constituye una muestra aleatoria las conclusiones basadas en dicha muestra no son fiables y en general estarán sesgadas en algún aspecto.

En términos matemáticos, dada una variable aleatoria X con una distribución de probabilidad F , una muestra aleatoria de tamaño N es un conjunto finito de N variables independientes, con la misma distribución de probabilidad F .

Otra forma más intuitiva, de entender una muestra es considerar que una muestra es una sucesión de N experimentos independientes de una misma cantidad. Es importante diferenciar una muestra de tamaño N , o más exactamente un muestreo de

tamaño N , del resultado concreto de los N experimentos (que como conjunto de valores fijos, en sí mismo, no es una muestra). El concepto de muestra incluye de alguna manera el procedimiento escogido para obtener los datos (es decir, si las variables aleatorias consideradas son independientes entre sí, y si tienen la misma distribución).

En general, resulta muy difícil comprobar si una determinada muestra es o no aleatoria, cosa que sólo puede hacerse considerando otro tipo de muestreos aleatorios robustos que permitan decir si la primera muestra era aleatoria o no.

8.2 Muestreo en poblaciones finitas

El **muestreo en poblaciones finitas** o encuesta por muestreo consiste en la selección de una parte de los elementos de una población estadística (U), con el objetivo de sacar conclusiones de dicha población.

8.2.1 Estimación de población

Por lo tanto una población es un conjunto de N elementos, los cuales se numeran de 1 a N , por lo tanto:

$U = (1, 2, \dots, i, \dots, N)$, cada elemento i posee un valor x_i asociado a la variable de interés x .

El objetivo es estimar el total de la variable x o sea:

$$t_x = \sum_U x_i, \text{ donde } \sum_U \text{ indica la suma en toda la población } (U)$$

o la media poblacional de dicha variable.

$$\bar{x}_U = \frac{\sum_U x_i}{N} = \frac{t_x}{N}$$

por lo tanto el estimador HT para la variable x es:

$$\hat{t}_x = \sum_s x_i d_i$$

donde d_i es el inverso de la probabilidad de inclusión del elemento i en la muestra.

8.3 Error muestral

En estadística, error muestral o error de estimación es el error a causa de observar una muestra en lugar de la población completa.

La estimación de un valor de interés, como la media o el porcentaje, estará generalmente sujeta a una variación entre una muestra y otra.¹ Estas variaciones en las posibles muestras de una estadística pueden, teóricamente, ser expresadas como errores muestrales, sin embargo, normalmente, en la práctica el error exacto es desconocido. El error muestral se refiere en términos más generales al fenómeno de la variación entre muestras. Cuando este no es mencionado se considera que el margen de error base es el 0.02% (0.2 para muestreo paralelo y 2 para muestreo directo).

El error muestral deseado, generalmente puede ser controlado tomando una muestra aleatoria de la población, suficientemente grande, sin embargo, el costo de esto puede ser limitante. Si las observaciones son tomadas de una muestra aleatoria, la teoría estadística brinda cálculos probabilísticos del tamaño deseado del error muestral para una estadística en particular o estimación. Estos usualmente son expresados en términos del error estándar.

El error muestral puede ser contrastado con el error no muestral, el cual se refiere al conjunto de las desviaciones del valor real que no van en función de la muestra escogida, entre los cuales se encuentran varios errores sistemáticos y algunos errores aleatorios. Resultan mucho más difíciles de cuantificar que el error muestral.

8.4 Estrategias de muestreo

8.4.1 Muestreo probabilístico

En este tipo de muestreo los arqueólogos intentan que las generalizaciones que realizan a partir de las muestras, sean correctas. Se basan en la probabilidad. En arqueología cuanto más precisa y más amplia sea la muestra, más probabilidades habrá de que los resultados sean óptimos.

Se conocen cuatro tipos de estrategias de muestreo: muestreo aleatorio simple, muestreo aleatorio estratificado, muestreo sistemático y muestreo sistemático estratificado.

8.4.2 Muestreo aleatorio simple

Es considerado el método más sencillo. Mediante una tabla de números al azar se eligen las zonas que se quieren muestrear. Este tipo de muestreo posee algunos inconvenientes. Por un lado, supone definir de antemano los límites de un yacimiento, y no siempre se conocen con certeza. Por otro lado, el carácter aleatorio de las tablas numéricas provoca que en algunas áreas se acumulen las muestras, mientras que en otras permanecen intactas.

8.4.3 Muestreo aleatorio estratificado

En este tipo de muestreo lo que se hace es dividir una región o un yacimiento teniendo en cuenta sus estratos naturales, como puede ser una tierra cultivada y un bosque. A continuación se seleccionan las cuadrículas, con base en números aleatorios, como en el muestreo aleatorio simple, pero con la diferencia de que se asigna a cada zona un número de cuadros proporcional a su superficie.

8.4.4 Muestreo sistemático

Este tipo de muestreo se basa en el empleo de una red de lugares equidistantes, como por ejemplo eligiendo un cuadrado cada dos. Uno de los problemas que tiene es que es muy probable errar, ya que este método de espaciado regular corre el riesgo de errar (o acertar) todas las muestras sin excepción si la distribución misma es también constante.

8.4.5 Muestreo sistemático estratificado

Este tipo de muestreo se caracteriza por la combinación de elementos de los otros tipos de muestreo: muestreo aleatorio simple, aleatorio estratificado y sistemático. Es un intento de reducir la arbitrariedad en la toma de muestras.

8.4.6 Muestreo no probabilístico

Esta estrategia de muestreo se basa en la intuición, la documentación histórica o una gran experiencia de campo de la zona. El muestreo no probabilístico no sería tan científico como pudiera ser el probabilístico, ya que si se quiere saber de forma cuantitativa la importancia de un yacimiento o región es más factible utilizar el muestreo probabilístico. Véase muestreo accidental, uno de los nombres del también llamado muestreo de oportunidad o muestreo de conveniencia. Desarrollado por RRS.

8.5 Muestreo estratificado

Es una forma de representación estadística que muestra como se comporta una característica o variable en una población a través de hacer evidente el cambio de dicha variable en sub-poblaciones o estratos.

Consiste en la división previa de la población de estudio en grupos o clases que se suponen homogéneos respecto a característica a estudiar y que no se solapan.

Según la cantidad de elementos de la muestra que se han de elegir de cada uno de los estratos, existen dos técnicas de muestreo estratificado:

1. Asignación proporcional: el tamaño de cada estrato en la muestra es proporcional a su tamaño en la población.

2. Asignación óptima: la muestra recogerá más individuos de aquellos estratos que tengan más variabilidad. Para ello es necesario un conocimiento previo de la población.

Por ejemplo, para un estudio de opinión, puede resultar interesante estudiar por separado las opiniones de hombres y mujeres pues se estima que, dentro de cada uno de estos grupos, puede haber cierta homogeneidad. Así, si la población está compuesta de un 55% de mujeres y un 45% de hombres, se tomaría una muestra que contenga también esa misma proporción.

8.5.1 Otro ejemplo práctico

En general, el tamaño de la muestra en cada estrato se toma en proporción con el tamaño del estrato. Eso se llama asignación proporcional. Supóngase que en una empresa se encuentran los siguientes funcionarios :

hombre, jornada completa: 90

hombre, media jornada: 18

mujer, jornada completa: 9

mujer, media jornada: 63

Total: 180

Se pide tomar una muestra de 40 personas, estratificada según las categorías anteriores.

El primer paso es encontrar el número total de funcionarios (180) y calcular el porcentaje de cada grupo.

$$\% \text{ hombre, jornada completa} = 90 / 180 = 50\%$$

$$\% \text{ hombre, media jornada} = 18 / 180 = 10\%$$

$$\% \text{ mujer, jornada completa} = 9 / 180 = 5\%$$

$$\% \text{ mujer, media jornada} = 63 / 180 = 35\%$$

Esto dice que nuestra muestra de 40:

50% debe ser hombre, jornada completa

10% debe ser hombre, media jornada

5% debe ser mujer, jornada completa

35% debe ser mujer, media jornada

50% de 40 es 20

10% de 40 es 4

5% de 40 es 2

35% de 40 es 14

Otra manera fácil sin necesidad de calcular el porcentaje es multiplicar cada tamaño de grupo por el tamaño de la muestra y se dividen por el tamaño total de la población (tamaño de todo el personal):

$$\text{hombre, jornada completa} = 90 \times (40 / 180) = 20$$

$$\text{hombre, media jornada} = 18 \times (40 / 180) = 4$$

mujer, jornada completa = $9 \times (40 / 180) = 2$

mujer, media jornada = $63 \times (40 / 180) = 14$

8.5.2 Herramientas de la Mejora Continua

Este tipo de muestreo forma parte de las siete nuevas herramientas de mejora continua de la calidad. Son herramientas que permiten al área gerencial recabar y organizar datos que les permitan crear un plan estratégico e implementar acciones de mejora en una empresa.