Manejo de la información para sistemas de gestión

Tamaño de la muestra y cómo calcularlo

Servicio Nacional de Aprendizaje SENA

Tamaño de la muestra y cómo calcularlo

Tamaño de la Muestra: El tamaño de muestra es fundamental para garantizar representatividad estadística de los datos y de esta forma asegurarse de la validez de los mismos a la hora de analizarlos y tomar decisiones.

Ante una muestra demasiado pequeña: podría incluir una cantidad desproporcionada de encuestados con características atípicas, hecho que podría generar valores anómalos, distorsionando los resultados, impidiendo obtener una visión realista respecto a la población objeto de estudio.

Ante una muestra demasiado grande: la investigación se vuelve demasiado compleja, costosa y lenta. Incluso aunque los resultados obtenidos sean de mayor precisión, los beneficios obtenidos pueden no compensar los costos del estudio.

Calcular el tamaño la muestra

Con el fin de calcular el tamaño adecuado para la muestra es necesario tener en cuenta los factores que afectan el estudio, además es necesario comprender algunos elementos básicos de los cálculos estadísticos a aplicar. Posteriormente, a través de una fórmula de tamaño de muestra, sumado a un adecuado muestreo, es posible garantizar adecuación de los datos del estudio.

Los siguientes pasos permiten calcular el tamaño de la muestra para datos continuos, es decir, datos cuantitativos. Por tal motivo, no se aplica a datos categóricos, es decir, datos que se agrupan en categorías, cualitativas, tales como grande, pequeño, mujer, hombre, niño, anciano, entre otros.

Determinar las variables del tamaño de la muestra como instancia previa al cálculo de la muestra se requiere determinar algunos aspectos relacionados con la población, además del grado de precisión necesario para el estudio, dichos elementos son abordados a continuación.

- **Tamaño de la población** ¿Cuál es el total de individuos que componen el segmento que desea estudiar? Con el fin de determinarlo, es imprescindible tener claras las características de los individuos que componen el grupo poblacional. Por ejemplo, si el estudio implica explorar los hábitos de consumo de los clientes de un local, debe incluir a todas las personas que alguna vez hayan sido clientes del local (dependiendo del objetivo del estudio, el investigador podría incluir o no a aquellos que fueron clientes anteriormente, pero no en la actualidad). En caso de no conocer el número exacto, es posible realizar el estudio partiendo de un tamaño de población desconocido o un rango estimado.

- **Margen de error (intervalo de confianza)** Es inevitable que se produzcan errores. Por tal motivo es necesario determinar el grado de error aceptable para el estudio. El margen de error, también llamado intervalo de confianza, se expresa en valores medios. Es necesario establecer qué diferencia se aceptará entre la media de la muestra y la media de la población a estudiar. Si alguna vez vio una encuesta política, entonces ha visto un intervalo de confianza y cómo se expresa. Se ve como algo parecido a esto: “El 68 % de los votantes aprobaron la propuesta Z, con un margen de error de +/– 5 %”.

- **Nivel de confianza** esto es lo que sigue al intervalo de confianza del paso 2. Se refiere a cuánta confianza se desea tener en que la media real se encontrará dentro del margen de error. Los intervalos de confianza más comunes son de un 90 %, un 95 % y un 99 % de confianza.

- **Desviación estándar** En este paso, se debe estimar el grado de variación que tendrán las respuestas recibidas entre sí y en relación con la media. Una desviación estándar baja significa que todos los valores se agruparán en torno a la media, mientras que una desviación estándar alta significa que se distribuirán en un rango mucho más amplio, con valores atípicos muy pequeños y muy grandes. Dado que aún no se ha realizado la encuesta, lo más seguro es optar por una desviación estándar de 0,5, que garantizará que el tamaño de la muestra sea lo suficientemente grande.

Determinar las variables del tamaño de la muestra

Como instancia previa al cálculo de la muestra se requiere determinar algunos aspectos relacionados con la población, además del grado de precisión necesario para el estudio, dichos elementos son abordados a continuación.

El cálculo

Una vez determinados los valores explicados anteriormente, procederá a calcular el tamaño de la muestra. Para lo cual podrá seguir las siguientes instrucciones:

- **Determinar la puntuación Z**

Lo siguiente que debe hacer es convertir el nivel de confianza en una puntuación Z.

Estas son las puntuaciones Z para los niveles de confianza más comunes:

- 90 % – Puntuación Z = 1,645
- 95 % – Puntuación Z = 1,96
- 99 % – Puntuación Z = 2,576

- **Use la fórmula del tamaño de la muestra** Una vez determinada la puntuación Z, la desviación estándar y el intervalo de confianza deberá hacer uso de la siguiente fórmula para calcularlo:

$$\text{Necessary Sample Size} = \frac{(\text{Z-score})2 \times \text{StdDev} \times (1\text{-StdDev})}{(\text{margin of error})2}$$

Figura 1: Fórmula de tamaño de muestra con población indeterminada

Fuente:(https://www.qualtrics.com/es-la/gestion-de-la-experiencia/investigacion/calcular-tomano-muestra)

Esta ecuación se utiliza para un tamaño de población desconocido o muy grande.
Llevándolo a la práctica, suponiendo que se elige un nivel de confianza del 95%, una desviación estándar de 0,5 y un margen de error (intervalo de confianza) de +/– 5 %, entonces:

((1,96)2 x 0,5(0,5)) / (0,05)2
(3,8416 x 0,25) / 0,0025
0,9604 / 0,0025
384,16

Figura 2: El tamaño de la muestra es de 385 individuos.

Fuente: (https://www.qualtrics.com/es-la/gestion-de-la-experiencia/investigacion/calcular-tomano-muestra/)

En caso de población finita, es decir se conoce el total de individuos que componen la población y se requiere establecer el tamaño de la muestra, es necesario aplicar la siguiente fórmula:

$$n = \frac{N * Z_{\alpha}^{2} p * q}{d^{2} * (N-1) + Z_{\alpha}^{2} * p * q}$$

Figura 3: Fórmula de tamaño de muestra con población finita

Fuente:

https://investigacionpediahr.files.wordpress.com/2011/01/formula-para-cc3a1lculo-de-la-muestra-poblaciones-finitas-var-categorica.pdf

- N = Total de la población
- Zα= 1.96 al cuadrado (si la seguridad es del 95%)
- p = proporción esperada (en este caso 5% = 0.05)
- q = 1 – p (en este caso 1-0.05 = 0.95)
- d = precisión (en su investigación use un 5%).

Ejemplo:

¿A cuántas personas tendría que estudiar de una población de 15.000 habitantes para conocer la Prevalencia de diabetes?

Seguridad = 95%; Precisión = 3% (recuerde, en su investigación use 5%, en este ejemplo usaremos 3%); proporción esperada = asumamos que puede ser próxima al 5% (0.05); si no tuviese ninguna idea de dicha proporción utilizaríamos el valor p = 0.5 (50%) que maximiza el tamaño muestral.

$$n = \frac{15.000 * 1.96^2 * 0.05 * 0.95}{0.03^2 (15.000 - 1) + 1.96^2 * 0.05 * 0.95} = 200$$

Figura 4: Aplicación Fórmula de tamaño de muestra con población finita

Fuente:

https://investigacionpediahr.files.wordpress.com/2011/01/formula-para-cc3a1lculo-de-la-muestra-poblaciones-finitas-var-categorica.pdf

Variables estadísticas:

Las variables estadísticas pueden ser clasificadas en dos tipos.

Variables cuantitativas: Se expresan de manera numérica y pueden ser:

- **Discretas:** Representan un valor finito de valores en un intervalo de datos, ejemplo: Número de prendas vendidas.
- **Continuas:** Representan un valor infinito de valores en un intervalo de datos, ejemplo: Tiempo que requiere un atleta en terminar una maratón.

Variables cualitativas: Son variables que expresan características y pueden ser:

- **Ordinales:** Representan elementos ordinales o niveles, ejemplo: ranking de estudiantes (primero, segundo, tercero, entre otras.)
- **Nominales:** Representan categorías claramente definidas, ejemplo: color de cabello (rojo, negro, rubio, castaño, entre otras.)

Distribución de frecuencias:

Consiste en presentar de manera ordenada los datos estadísticos a través de una tabla, en la que se asigna la frecuencia correspondiente a cada dato, las frecuencias pueden ser de varios tipos:

- **Frecuencia absoluta:** equivale al número de veces que se presenta un valor determinado en un estudio estadístico, puede ser representada, dependiendo del autor por *fi* o *ni*. Al final, la sumatoria de las frecuencias equivale al número total de datos N.

$$N = \sum_{i=1}^{n} f_i$$

- **Frecuencia acumulada:** Corresponde a la suma de las frecuencias absolutas de todos los valores inferiores o iguales al valor considerado. Se representa por *fai*.

- **Frecuencia relativa:** Corresponde al cociente entre la frecuencia absoluta de un determinado valor y el número total de datos y se representa por *fri*.

$$fr_i = \frac{f_i}{N}$$

La frecuencia relativa es un número comprendido entre 0 y 1. La suma de las frecuencias relativas es igual a 1.

- **Frecuencia relativa acumulada:** Corresponde a la suma de las frecuencias relativas de todos los valores inferiores o iguales al valor considerado. Se obtiene como resultado de dividir la frecuencia acumulada de un valor determinado entre el total de datos y se representa por *frai*

$$frai = \frac{fai.}{\text{N}}$$

- **Frecuencia porcentual:** Corresponde al porcentaje de la frecuencia absoluta en relación con el total de los datos, se obtiene al multiplicar el valor de la frecuencia relativa por 100, se expresa de manera porcentual y se representa por *fpi*.

$$fp_i = fr_i \times 100$$