TAMAÑO DE LA MUESTRA

Recordemos los siguientes conceptos básicos

Población.- Llamado también universo o colectivo, es el conjunto de todos los elementos que tienen una característica común. Una población puede ser finita o infinita. Es *población finita* cuando está delimitada y conocemos el número que la integran, así por ejemplo: Estudiantes de la Universidad UTN. Es *población infinita* cuando a pesar de estar delimitada en el espacio, no se conoce el número de elementos que la integran, así por ejemplo: Todos los profesionales universitarios que están ejerciendo su carrera.

Muestra.- La muestra es un subconjunto de la población. Ejemplo: Estudiantes de 2do Semestre de la Universidad UTN.

Sus principales características son:

Representativa.- Se refiere a que todos y cada uno de los elementos de la población tengan la misma oportunidad de ser tomados en cuenta para formar dicha muestra.

Adecuada y válida.- Se refiere a que la muestra debe ser obtenida de tal manera que permita establecer un mínimo de error posible respecto de la población.

Para que una muestra sea fiable, es necesario que su tamaño sea obtenido mediante procesos matemáticos que eliminen la incidencia del error.

Elemento o individuo.- Unidad mínima que compone una población. El elemento puede ser una entidad simple (una persona) o una entidad compleja (una familia), y se denomina unidad investigativa.

A) DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA PARA LA MEDIA

Para desarrollar una fórmula que permita determinar el tamaño apropiado de la muestra para construir una estimación del intervalo de confianza para la media, recuerde la ecuación

$$\bar{X} - Z\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z\frac{\sigma}{\sqrt{n}}$$

La cantidad sumada o sustraída de $\bar{X}$ es igual a la mitad de la amplitud del intervalo. Esta cantidad representa la cantidad de imprecisión en la estimación que resulta del error de muestreo. El error de muestreo e (también conocido como margen de error) se define como:

$$e = Z\frac{\sigma}{\sqrt{n}}$$

Al despejar n se obtiene el tamaño de muestra necesario para construir una estimación del intervalo de confianza adecuado para la media. "Adecuado" significa que el intervalo resultante tendrá una cantidad aceptable de error de muestreo.
El proceso de despejar n se muestra a continuación:

Elevando al cuadrado ambos miembros de la igualdad:

$$(e)^2 = \left(Z\frac{\sigma}{\sqrt{n}}\right)^2$$

Resolviendo la potencia:

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

Transponiendo n a la izquierda de la igualdad:

$$e^2 n = Z^2 \sigma^2$$

Transponiendo e^2 a la derecha de la igualdad y ordenando se obtiene la fórmula buscada:

$$\boldsymbol{n = \frac{\sigma^2 Z^2}{e^2}}$$

Donde:

$\sigma =$ Desviación estándar de la población que rara vez conoce su valor. En algunas ocasiones es posible estimar la desviación estándar a partir de datos pasados. En otras situaciones, puede hacer una cuidadosa conjetura tomando en cuenta el rango y la distribución de la variable. Por ejemplo, si supone que hay una distribución normal, el rango es aproximadamente igual a 6σ (es decir, $\pm 3\sigma$ alrededor de la media). Generalmente cuando no se tiene su valor, suele utilizarse un valor constante de 0,5.

Z = Valor obtenido mediante niveles de confianza. Es un valor constante que, si no se tiene su valor, se lo toma en relación al 95% de confianza equivale a 1,96 (como más usual) o en relación al 99% de confianza equivale 2,58, valor que queda a criterio del investigador.

e = Límite aceptable de error muestral que, generalmente cuando no se tiene su valor, suele utilizarse un valor que varía entre el 1% (0,01) y 9% (0,09), valor que queda a criterio del encuestador, sin embargo, se aconseja emplear el 5% o el 1%, por ser valores que guardan relación con el 95% y 99% de confianza, respectivamente.

Cuando en los datos se tiene el tamaño de la población se utiliza la siguiente fórmula:

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

La fórmula anterior se obtiene de la fórmula de la estimación del intervalo de confianza para la media, la cual es:

$$\bar{X} - Z\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \leq \mu \leq \bar{X} + Z\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

De donde el error es:

$$e = Z\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

De esta fórmula del error de la estimación del intervalo de confianza para la media se despeja la n, para lo cual se sigue el siguiente proceso:

Elevando al cuadrado a ambos miembros de la fórmula se obtiene:

$$(e)^2 = \left(Z\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right)^2 \Longrightarrow e^2 = Z^2\frac{\sigma^2}{n}\frac{N-n}{N-1}$$

Multiplicando fracciones y eliminando denominadores:

$$e^2 = \frac{Z^2\sigma^2(N-n)}{n(N-1)} \Longrightarrow e^2n(N-1) = Z^2\sigma^2(N-n)$$

Eliminando paréntesis y transponiendo n a la izquierda:

$$e^2nN - e^2n = Z^2\sigma^2N - Z^2\sigma^2n \Longrightarrow e^2nN - e^2n + Z^2\sigma^2n = Z^2\sigma^2N$$

Factor común de n y Despejando n:

$$n(e^2N - e^2 + Z^2\sigma^2) = Z^2\sigma^2N \Longrightarrow \ n = \frac{Z^2\sigma^2N}{e^2N - e^2 + Z^2\sigma^2}$$

Sacando factor común e^2 y ordenando se obtiene la fórmula para calcular el tamaño de la muestra para la media cuando se conoce el tamaño de la población:

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

B) DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA PARA LA PROPORCIÓN

Para determinar el tamaño de la muestra necesario para estimar la proporción poblacional (π), se utiliza un método similar al método para calcular la media poblacional. Al desarrollar el tamaño de la muestra para un intervalo de confianza para la media, el error de muestreo se define por:

$$e = Z\frac{\sigma}{\sqrt{n}}$$

Cuando estima la proporción, se remplaza $\sigma = \sqrt{\pi(1-\pi)}$. Así, el error de muestreo es:

$$e = Z\frac{\sqrt{\pi(1-\pi)}}{\sqrt{n}} \Rightarrow e = Z\sqrt{\frac{\pi(1-\pi)}{n}}$$

Al despejar n se obtiene el tamaño de la muestra necesario para desarrollar la estimación del intervalo de confianza para la proporción. El tamaño de la muestra n es igual al cuadrado del valor Z multiplicado por la proporción poblacional π, multiplicado por 1 menos la proporción poblacional π, y dividido por el cuadrado del error de muestreo e

$$\boldsymbol{n} = \frac{\boldsymbol{\pi(1-\pi)Z^2}}{\boldsymbol{e^2}}$$

Al determinar π se tiene dos alternativas. En muchas situaciones se cuenta con información anterior o experiencias relevantes que proporcionan una estimación cuidadosa de π. 0 también, si no se cuenta con información anterior o experiencias relevantes, se trata de proporcionar un valor para π que nunca subestime el tamaño de muestra necesario. Con referencia a la ecuación anterior, observe que la cantidad $\pi(1-\pi)$aparece en el numerador. Por eso, se requiere determinar el valor de π que haga la cantidad $\pi(1-\pi)$lo más grande posible. Cuando $\pi = 0{,}5$, el producto $\pi(1-\pi)$ logra su resultado máximo. Para mostrar esto, los valores de π junto con los productos que los acompañan de $\pi(1-\pi)$ son como sigue:

$$Cuando\ \pi = 0{,}9 \Rightarrow \pi(1-\pi) = 0{,}9(1-0{,}9) = 0{,}09$$
$$Cuando\ \pi = 0{,}7 \Rightarrow \pi(1-\pi) = 0{,}7(1-0{,}7) = 0{,}21$$

$Cuando\ \pi = 0{,}5 \Rightarrow \pi(1-\pi) = 0{,}5(1-0{,}5) = 0{,}25$
$Cuando\ \pi = 0{,}3 \Rightarrow \pi(1-\pi) = 0{,}3(1-0{,}3) = 0{,}21$
$Cuando\ \pi = 0{,}1 \Rightarrow \pi(1-\pi) = 0{,}1(1-0{,}1) = 0{,}09$

Por lo tanto, cuando no se tiene conocimiento previo o una estimación de la proporción poblacional π, se debería usar $\pi = 0{,}5$ para determinar el tamaño de la muestra. Esto produce el tamaño de muestra más grande posible y deriva en el mayor costo posible del muestreo.

Cuando en los datos se tiene el tamaño de la población se utiliza la siguiente fórmula:

$$\boldsymbol{n = \frac{N\pi(1-\pi)Z^2}{(N-1)e^2 + \pi(1-\pi)Z^2}}$$

Ejemplo ilustrativo

Calcular el tamaño de la muestra de una población de 500 elementos con un nivel de confianza del 95%

Solución: Se tiene N=500, para el 95% de confianza Z = 1,96, y como no se tiene los demás valores se tomará $\sigma = 0{,}5$, y e = 0,05.

Remplazando valores de la fórmula se tiene:

$$n = \frac{N\sigma^2 Z^2}{(N-1)e^2 + \sigma^2 Z^2} \Longrightarrow n = \frac{500 \cdot 0{,}5^2 \cdot 1{,}96^2}{(500-1)0{,}05^2 + 0{,}5^2 \cdot 1{,}96^2} = 217$$

Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F
1	N	500				
2	e	0,05				
3	σ	0,5				
4	Confianza	95				
5	Área a la izquierda de -Z	0,025	=(100-B4)/200			
6	-Z	-1,96	=INV.NORM.ESTAND(B5)			
7	Z	1,96	=B6*-1			
8	$n = \frac{N\sigma^2 Z^2}{(N-1)e^2 + \sigma^2 Z^2}$	217,49	=B1*B3^2*B7^2/((B1-1)*B2^2+B3^2*B7^2)			
9						

TAREA DE INTERAPRENDIZAJE

1) Realice un organizador gráfico sobre el tamaño de la muestra

2) Proponga 3 ejemplos de población, de muestra y de elemento.

3) Contestar
3.1) ¿Qué significa “adecuado” para una estimación del intervalo de confianza que permite calcular el tamaño de la muestra?

3.2) ¿A partir de qué ecuación se obtiene el error de muestreo para la media cuando se conoce el tamaño de la población?
3.3) ¿Con qué otro nombre se le conoce al error de muestreo?
3.4) ¿Por qué se emplea la distribución Z en lugar de la distribución t para determinar el nivel de confianza deseado del tamaño de la muestra?
3.5) ¿Qué valores suelen utilizarse para Z y e cuando no se conocen sus valores?
3.6) ¿Por qué cuando no se tiene conocimiento previo o una estimación de la proporción poblacional π se debe empelar $\pi = 0{,}5$?

4) Calcular el valor de σ si se sabe $\pi = 0{,}5$

0,5

5) Escriba el proceso para despejar n de la siguiente fórmula para obtener la fórmula que permite calcular el tamaño de la muestra

$$e = Z\sqrt{\frac{\pi(1-\pi)}{n}}$$

6) Siguiendo el razonamiento del ejercicio anterior, escriba el proceso para despejar n de la fórmula del error de muestreo que permita obtener la siguiente fórmula:

$$n = \frac{N\pi(1-\pi)Z^2}{(N-1)e^2 + \pi(1-\pi)Z^2}$$

7) Un departamento de la dirección aérea civil del Ecuador define que el vuelo de una línea llega "a tiempo" si aterriza menos de 15 minutos después del tiempo programado en el Sistema de Reservación Computarizado. Los vuelos cancelados y desviados se consideran con retraso. Un estudio de las aerolíneas encontró que Icaro tiene la menor proporción de retrasos. Suponga que se le pide realizar un estudio de seguimiento para Icaro cuyo objetivo sea actualizar la proporción estimada de retrasos. ¿Qué tamaño de muestra deberá emplear para estimar la proporción poblacional con un 95% de nivel de confianza y con un error de ± 0,06 ?. Calcule con lectura en la tabla, Excel y Winstats

267

8) Con lectura en la tabla, Excel y Winstats calcular el tamaño de muestra con un error de muestreo de ±0,05 para N =500, si se desea tener un nivel de confianza en la estimación de la media poblacional del:
8.1) 95%

218

8.2) 99 %

285

9) Cree y resuelva con Excel y Winstats y ejercicio similar al anterior.

10) Una muestra de 280 fue calculada de una población con un nivel de confianza del 95% y un error de muestreo del 5%. Calcule el tamaño de la población.

1030

11) Una muestra de 286 fue calculada de una población de 500 con un error de muestreo del 5%. Calcule el nivel de confianza empleado para calcular la mencionada muestra.

99%

12) Consulte en la biblioteca o en el internet sobre la importancia del cálculo del tamaño de la muestra, y presente la consulta a través de un organizador gráfico.

Fuente:

Suárez, Mario. (2014). *Probabilidades y Estadística empleando las TIC*, Ecuador: Imprenta GRAFICOLOR

Nota: Libros y artículos del Mgs. Mario Suárez sobre Aritmética, Álgebra, Geometría, Trigonometría, Lógica Matemática, Probabilidades, Estadística Descriptiva, Estadística Inferencial, Cálculo Diferencial, Cálculo Integral, Los Poliprismas, y Planificaciones Didácticas se encuentran publicados en:

http://repositorio.utn.edu.ec/handle/123456789/24

http://docentesinnovadores.net/Usuarios/Ver/29591

http://es.scribd.com/mariosuarezibujes

http://www.monografias.com/usuario/perfiles/mario_suarez_7/monografias

Mgs. Mario Suárez *mariosuarezibujes@gmail.com; mosuarez@utn.edu.ec*