DETERMINACIÓN DE TAMAÑOS DE MUESTRA

Cálculo del Tamaño de la Muestra para Estimar una Media

¿Qué tan grande debe ser una muestra si la media muestral se va a usar para estimar la media poblacional? La respuesta depende del error estándar de la media, si este fuera cero, entonces se necesitaría una sola media que será igual necesariamente a la media poblacional desconocida μ, porque $\sigma = 0$. Este caso extremo no se encuentra en la práctica, pero refuerza el hecho de que mientras menor sea el error estándar de la media, menor es el tamaño de muestra necesario para lograr un cierto grado de precisión.

Se estableció antes que una forma de disminuir el error de estimación es aumentar el tamaño de la muestra, si éste incluye el total de la población, entonces $|\bar{x} - \mu|$ sería igual a cero. Con esto en mente, parece razonable que para un nivel de confianza fijo, sea posible determinar un tamaño de la muestra tal que el error de estimación sea tan pequeño como queramos, para ser mas preciso, dado un nivel de confianza y un error fijo de estimación ε, se puede escoger un tamaño de muestra *n* tal que $P(|\bar{x} - \mu| < \varepsilon)$ = Nivel de confianza. Con el propósito de determinar *n*. El error máximo de estimación esta dado por:

$$\varepsilon = \frac{z\sigma}{\sqrt{n}}$$

Si se eleva al cuadrado ambos lados de esta ecuación y se despeja *n* de la ecuación resultante, obtenemos:

$$n = \left(\frac{z\sigma}{\varepsilon}\right)^2$$

Como *n* debe de ser un número entero, redondeamos hacia arriba todos los resultados fraccionarios.

En el caso de que se tenga una población finita y un muestreo sin reemplazo, el error de estimación se convierte en:

$$\varepsilon = \frac{z\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

De nuevo se eleva al cuadrado ambos lados y se despeja la *n*, obteniendo:

$$n = \frac{z^2\sigma^2 N}{\varepsilon^2(N-1) + z^2\sigma^2}$$

Ejemplos:

1. Un biólogo quiere estimar el peso promedio de los ciervos cazados en el estado de Maryland. Un estudio anterior de diez ciervos cazados mostró que la desviación estándar de sus pesos es de 12.2 libras. ¿Qué tan grande debe ser una muestra para que el

biólogo tenga el 95% de confianza de que el error de estimación es a lo más de 4 libras?

Solución:

$$n = \left(\frac{z\sigma}{\varepsilon}\right)^2 = \left(\frac{(1.96)(12.2)}{4}\right)^2 = 35.736$$

En consecuencia, si el tamaño de la muestra es 36, se puede tener un 95% de confianza en que μ difiere en menos de 4 libras de $\bar{x}$.

2. Una empresa eléctrica fabrica focos que tienen una duración aproximadamente normal con una desviación estándar de 40 horas. ¿De qué tamaño se necesita una muestra si se desea tener 96% de confianza que la media real esté dentro de 10 horas de la media real?

$$n = \left(\frac{z\sigma}{\varepsilon}\right)^2 = \left(\frac{(2.053)(40)}{10}\right)^2 = 67.43$$

Se necesita una muestra de 68 focos para estimar la media de la población y tener un error máximo de 10 horas.

¿Qué pasaría si en lugar de tener un error de estimación de 10 horas sólo se requiere un error de 5 horas?

$$n = \left(\frac{z\sigma}{\varepsilon}\right)^2 = \left(\frac{(2.053)(40)}{5}\right)^2 = 269.74$$

Se puede observar como el tamaño de la muestra aumenta, pero esto tiene como beneficio una estimación más exacta.

3. Suponga que en el ejercicio anterior se tiene una población de 300 focos, y se desea saber de que tamaño debe de ser la muestra. El muestreo se realizará sin reemplazo.

Solución:

Como se tiene una población finita y un muestreo sin reemplazo es necesario utilizar la formula con el factor de corrección.

$$n = \frac{z^2\sigma^2 N}{\varepsilon^2(N-1) + z^2\sigma^2} = \frac{(2.053)^2(40)^2(300)}{(10)^2(300-1) + (2.053)^2(40)^2} = 55.21$$

Si se tiene una población finita de 300 focos sólo se tiene que extraer de la población una muestra sin reemplazo de 56 focos para poder estimar la duración media de los focos restantes con un error máximo de 10 horas.

Cálculo del Tamaño de la Muestra para Estimar una Proporción

Se desea saber que tan grande se requiere que sea una muestra para asegurar que el error al estimar P sea menor que una cantidad específica ε.

$$\varepsilon = z\sqrt{\frac{pq}{n}}$$

Elevando al cuadrado la ecuación anterior se despeja *n* y nos queda:

$$n = \frac{z^2 pq}{\varepsilon^2}$$

Esta fórmula está algo engañosa, pues debemos utilizar *p* para determinar el tamaño de la muestra, pero *p* se calcula a partir de la muestra. Existen ocasiones en las cuales se tiene una idea del comportamiento de la proporción de la población y ese valor se puede

sustituir en la fórmula, pero si no se sabe nada referente a esa proporción entonces se tienen dos opciones:

- Tomar una muestra preliminar mayor o igual a 30 para proporcionar una estimación de P. Después con el uso de la fórmula se podría determinar de forma aproximada cuántas observaciones se necesitan para proporcionar el grado de precisión que se desea.
- Tomar el valor de p como 0.5 ya que sustituyendo este en la fórmula se obtiene el tamaño de muestra mayor posible. Observe el siguiente ejemplo:

Se desconoce el valor de P, por lo que se utilizarán diferentes valores y se sustituirán en la fórmula para observar los diferentes tamaños de muestras. El nivel de confianza que se utilizará es del 95% con un error de estimación de 0.30.

p	$\frac{z^2 pq}{\varepsilon^2}$	n
0.10	$\frac{(1.96)^2(0.10)(0.90)}{(0.30)^2}$	3.84
0.20	$\frac{(1.96)^2(0.20)(0.80)}{(0.30)^2}$	6.82
0.30	$\frac{(1.96)^2(0.30)(0.70)}{(0.30)^2}$	8.96
0.40	$\frac{(1.96)^2(0.40)(0.60)}{(0.30)^2}$	10.24
0.50	$\frac{(1.96)^2(0.50)(0.50)}{(0.30)^2}$	10.67
0.60	$\frac{(1.96)^2(0.60)(0.40)}{(0.30)^2}$	10.24
0.70	$\frac{(1.96)^2(0.70)(0.30)}{(0.30)^2}$	8.96
0.80	$\frac{(1.96)^2(0.80)(0.20)}{(0.30)^2}$	6.82
0.90	$\frac{(1.96)^2(0.90)(0.10)}{(0.30)^2}$	3.84

Como se puede observar en la tabla anterior cuando P vale 0.5 el tamaño de la muestra alcanza su máximo valor.

En el caso de que se tenga una población finita y un muestreo sin reemplazo, el error de estimación se convierte en:

$$\varepsilon = z\sqrt{\frac{pq}{n}}\sqrt{\frac{N-n}{N-1}}$$

De nuevo se eleva al cuadrado ambos lados y se despeja la *n*, obteniendo:

$$n = \frac{z^2 pqN}{\varepsilon^2 (N-1) + z^2 pq}$$

Ejemplos:

1. En una muestra aleatoria de 500 familias que tienen televisores en la ciudad de Hamilton, Canadá, se encuentra que 340 están suscritas a HBO. ¿Qué tan grande se requiere que sea una muestra si se quiere tener 95% de confianza de que la estimación de P esté dentro de 0.02?

Solución:
Se tratarán a las 500 familias como una muestra preliminar que proporciona una estimación de p=340/500=0.68.

$$n = \frac{z^2 pq}{\varepsilon^2} = \frac{(1.96)^2 (0.68)(0.32)}{(0.02)^2} = 2090$$

Por lo tanto si basamos nuestra estimación de P sobre una muestra aleatoria de tamaño 2090, se puede tener una confianza de 95% de que nuestra proporción muestral no diferirá de la proporción real por más de 0.02.

2. Una legisladora estatal desea encuestar a los residentes de su distrito para conocer qué proporción del electorado conoce la opinión de ella, respecto al uso de fondos estatales para pagar abortos. ¿Qué tamaño de muestra se necesita si se requiere un confianza del 95% y un error máximo de estimación de 0.10?

Solución:
En este problema, se desconoce totalmente la proporción de residentes que conoce la opinión de la legisladora, por lo que se utilizará un valor de 0.5 para *p*.

$$n = \frac{z^2 pq}{\varepsilon^2} = \frac{(1.96)^2 (0.50)(0.50)}{(0.10)^2} = 96.04$$

Se requiere un tamaño de muestra de 97 residentes para que con una confianza del 95% la estimación tenga un error máximo de 0.10.

Cálculo del Tamaño de la Muestra para Estimar la Diferencia de Medias

Si se recuerda a la distribución muestral de diferencia de medias se tiene que error esta dado por:

$$\varepsilon = z\sqrt{\frac{\sigma^2{}_1}{n_1} + \frac{\sigma^2{}_2}{n_2}}$$

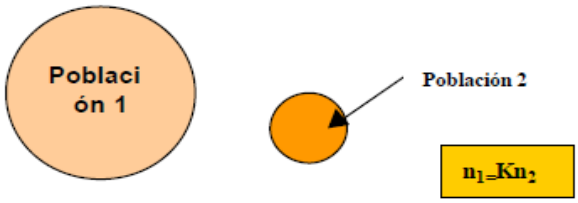
En esta ecuación se nos pueden presentar dos casos:

- Los tamaños de muestra son iguales.
- Los tamaño de muestra son diferentes .

Para el primer caso no se tiene ningún problema, se eleva al cuadrado la ecuación y se despeja n ya que n_1 es igual a n_2.

$$n = \frac{z^2(\sigma^2{}_1 + \sigma^2{}_2)}{\varepsilon^2}$$

Para el segundo caso se pondrá una n en función de la otra. Este caso se utiliza cuando las poblaciones son de diferente tamaño y se sabe que una es K veces mayor que la otra.



$$n_2 = \frac{z^2(\sigma^2{}_1 + k\sigma^2{}_2)}{k\varepsilon^2}$$

Ejemplo:

Un director de personal quiere comparar la efectividad de dos métodos de entrenamiento para trabajadores industriales a fin de efectuar cierta operación de montaje. Se divide un número de operarios en dos grupos iguales: el primero recibe el método de entrenamiento 1, y el segundo, el método 2. Cada uno realizará la operación de montaje y se registrará el tiempo de trabajo. Se espera que las mediciones para ambos grupos tengan una desviación estándar aproximadamente de 2 minutos. Si se desea que la estimación de la diferencia en tiempo medio de montaje sea correcta hasta por un minuto, con una probabilidad igual a 0.95, ¿cuántos trabajadores se tienen que incluir en cada grupo de entrenamiento?

$$n = \frac{z^2(\sigma^2{}_1 + \sigma^2{}_2)}{\varepsilon^2} = \frac{(1.96^2)(2^2 + 2^2)}{1^2} = 31$$

Cada grupo debe contener aproximadamente 31 empleados.

Cálculo del Tamaño de la Muestra para Estimar la Diferencia de Proporciones

Si se recuerda a la distribución muestral de diferencia de medias se tiene que error esta dado por:

$$\varepsilon = z\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

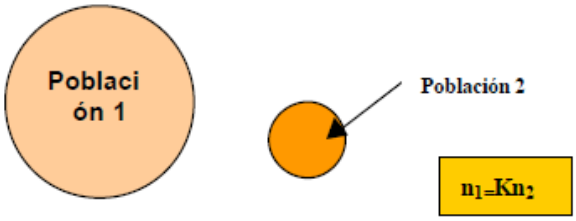
En esta ecuación se nos pueden presentar dos casos:

- Los tamaños de muestra son iguales.
- Los tamaño de muestra son diferentes .

Para el primer caso no se tiene ningún problema, se eleva al cuadrado la ecuación y se despeja n ya que n_1 es igual a n_2.

$$n = \frac{z^2(p_1q_1 + p_2q_2)}{\varepsilon^2}$$

Para el segundo caso se pondrá una *n* en función de la otra. Este caso se utiliza cuando las poblaciones son de diferente tamaño y se sabe que una es *K* veces mayor que la otra.



$$n_2 = \frac{z^2(p_1q_1 + kp_2q_2)}{k\varepsilon^2}$$

Ejemplo:

Una compañía de productos alimenticios contrató a una empresa de investigación de mercadotecnia, para muestrear dos mercados, I y II, a fin de comparar las proporciones de consumidores que prefieren la comida congelada de la compañía con los productos de sus competidores. No hay información previa acerca de la magnitud de las proporciones P_1 y P_2. Si la empresa de productos alimenticios quiere estimar la diferencia dentro de 0.04, con una probabilidad de 0.95, ¿cuántos consumidores habrá que muestrear en cada mercado?

$$n = \frac{z^2(p_1q_1 + p_2q_2)}{\varepsilon^2} = \frac{(1.96^2)[(0.5)(0.5) + (0.5)(0.5)]}{0.04^2} = 1200.5$$

Se tendrá que realizar encuestas a 1201 consumidores de cada mercado para tener una estimación con una confianza del 95% y un error máximo de 0.04.