TAMAÑO DE LA MUESTRA Y TEORÍA ESTADÍSTICA

CARACTERÍSTICAS DE LA POBLACIÓN. Una pregunta práctica en gran parte de la investigación de mercado tiene que ver con el tamaño de la muestra. La encuesta, en principio, no puede ser aplicada sin conocer el tamaño de la muestra.

Supongamos que estamos interesados en las actitudes de los poseedores de boletos para la temporada operística en el Teresa Carreño, al cambiar la hora de inicio de las ejecuciones de los días de la semana. La población se compone de 10000 poseedores de boletos de la temporada. De los poseedores de los boletos, 3000 respondieron “definitivamente si” (lo cual se codifica como +2). Otros 2000 “preferirían si” (codificado como + 1), y así sucesivamente. La información necesaria es la respuesta promedio o la respuesta media en la población (los 10000 poseedores de los boletos para la temporada), la cual se denomina como:

$$\mu = \text{media de la población} = 0.3$$

Pero esta media de la población generalmente es desconocida, y nuestra meta consiste en determinar su valor lo más aproximadamente posible tomando una muestra de la población.

TAMAÑO DE LA MUESTRA Y TEORÍA ESTADÍSTICA

Otra característica poblacional de interés es la varianza de la población, σ^2, y su raíz cuadrada, la desviación estándar de la población, σ. La varianza de la población es una medida de la dispersión de la población, es el grado en el que los distintos poseedores de boletos de temporada difieren entre sí en términos de actitudes. Se basa en el grado en el que una respuesta difiere de la respuesta promedio de la población. En nuestro ejemplo, la varianza de la población es:

$$\sigma^2 = \text{la varianza de la población} = 2.22$$

$$\sigma = \text{la desviación estándar de la población} = 1.49$$

Resultados de la codificación de las respuestas de los poseedores de boletos (población)

Respuesta R	Frecuencia de la respuesta f	Promedio ponderado Rf	Media de la Población μ	Diferencia entre la respuesta $R - \mu$	Diferencia elevada al cuadrado $(R-\mu)^2$	Promedio ponderado $(R-\mu)^2 f$
+2	0.3	0.6	0.3	1.7	2.89	0.87
+1	0.2	0.2	0.3	0.7	0.49	0.10
0	0.2	0	0.3	0.3	0.09	0.02
-1	0.1	-0.1	0.3	1.3	1.69	0.17
-2	0.2	-0.4	0.3	2.3	5.29	1.06
Total		0.3 = μ				2.22 = σ^2

TAMAÑO DE LA MUESTRA Y TEORÍA ESTADÍSTICA

CARACTERÍSTICAS DE LA POBLACIÓN. Pero el problema realmente es que la media de la población no es conocida y debe ser estimada a partir de una muestra. Considere una muestra aleatoria sencilla de tamaño 10 la cual es tomada de la población. La diez personas seleccionadas y sus actitudes respectivas se muestran a continuación.

Resultados de la codificación de las respuestas de los poseedores de boletos (muestra)

María Virginia	$X_1 = +1$
Luís	$X_2 = +2$
Mayra	$X_3 = +2$
Idmare	$X_4 = 0$
David	$X_5 = +1$
Yohama	$X_6 = +1$
Rubén	$X_7 = -1$
Vanesa	$X_8 = +1$
Jean	$X_9 = -2$
El profe	$X_{10} = 0$

TAMAÑO DE LA MUESTRA Y TEORÍA ESTADÍSTICA

El cálculo correspondiente a la media, varianza, y desviación estándar pero ahora con respecto a la muestra será mediante la aplicación de las siguientes fórmulas.

Media: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = 0.5$

Varianza: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2 = 1.61$

Desviación estándar: $S = \sqrt{s^2} = 1.27$

S^2 será pequeña si las respuestas de la muestra son simulares, y grandes si se encuentran dispersas. Nuevamente, es importante hacer una distinción entre la varianza de la población (σ^2) y la varianza de la muestra (S^2).

TAMAÑO DE LA MUESTRA Y TEORÍA ESTADÍSTICA

CONFIABILIDAD DE LA MUESTRA. Desde luego, todas las muestras no generaran el mismo valor $\bar{x}$ (o *S*). Si otra muestra aleatoria simple con tamaño de 10 fuera tomada de la población, $\bar{x}$ podrá ser de 0.3, 1.2, 0.4, o cualquier otra cantidad. El punto importante es que $\bar{x}$ variara de muestra en muestra. Intuitivamente será lógico pensar que la variación en $\bar{x}$ será más grande a medida que la varianza de la población, σ^2 sea más grande. En un extremo, si no hay población, no habrá tampoco variación en $\bar{x}$. También es razonable pensar que, a medida que aumenta el tamaño de la muestra, la variación en $\bar{x}$ disminuirá.

Cuando la muestra es pequeña, se necesitan sólo uno o dos valores extremos para afectar sustancialmente la media muestral, generando de este modo una $\bar{x}$ relativamente grande o pequeña. A medida que aumenta el tamaño de la muestra, estos valores extremos tendrán un menor impacto cuando aparezcan, porque serán promediados con más valores. La variación en $\bar{X}$ es medida por su error estándar, es cual es:

$$\sigma_{\bar{x}} = \text{el error estándar de } \bar{x} = \sigma_x / \sqrt{n} = 1.49 / \sqrt{10} = 0.47$$

TAMAÑO DE LA MUESTRA Y TEORÍA ESTADÍSTICA

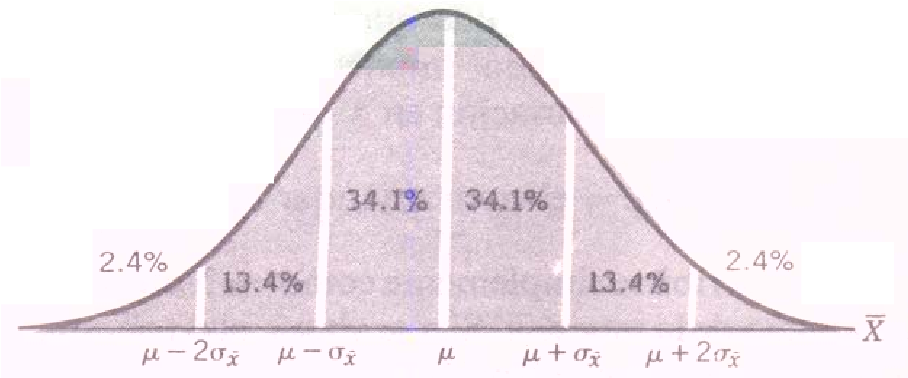
(σ_x puede escribirse simplemente como σ). Note que el error estándar $\bar{x}$ depende de *n,* el tamaño de la muestra. Si *n* se altera, el error estándar cambiará de acuerdo a ello, como se demuestra en el siguiente cuadro:

Tamaño de la muestra	$\boldsymbol{\sigma_x}$	$\sigma_{\bar{x}} = \boldsymbol{\sigma_x / \sqrt{n}}$
10	**1.49**	**0.470**
40	**1.49**	**0.235**
100	**1.49**	**0.149**
500	**1.49**	**0.067**

La variable X tiene una distribución de probabilidad similar a la forma acampanada de la distribución normal. Con la media de la muestra $\bar{x}$ pasa lo mismo. En otras palabras, indica que $\bar{x}$ generalmente se acercará a la media de la población (μ) y que tiene iguales posibilidades de ser más grande que μ o más pequeña. En la siguiente figura se muestra como el área bajo la curva normal es dividida. Por ejemplo, la probabilidad de que $\bar{x}$ se encuentre dentro $2\sigma_{\bar{x}}$ de la media de la población (μ) es de 0.95. Similarmente la probabilidad de que $\bar{x}$ se encuentre dentro de una $\sigma_{\bar{x}}$ de la media de la población (μ) es del 68,20%.

TAMAÑO DE LA MUESTRA Y TEORÍA ESTADÍSTICA

La distribución normal en función de $\bar{X}$



En nuestro ejemplo de la sinfonía, suponga que se obtuvieron 100 muestras diferentes de 10 sujetos. Aproximadamente el 95% de las medias muestrales resultantes $\bar{x}$ estarían dentro de ± 2 errores estándar ($\sigma_{\bar{x}}$ = 0.47) de la media de la población (μ = 0.3).

TAMAÑO DE LA MUESTRA Y TEORÍA ESTADÍSTICA

ESTIMACIÓN DEL INTEVALO DE CONFIANZA. La media de la muestra, $\bar{x}$, es usada para estimar la media de la población desconocida (μ). Debido a que $\bar{x}$ varía de muestra en muestra, no es desde luego, igual a la media de la población (μ). Ello porque como se explicó anteriormente hay un error de la muestra. Es útil proporcionar una estimación de intervalo en torno a el cual refleje nuestro juicio acerca del alcance del este error muestral.

$\bar{X}$ ± error de la muestra = estimación del intervalo de μ

El tamaño del intervalo dependerá de qué tan confiados queramos estar de que el intervalo contenga la media de la población verdadera y desconocida. Si fuera necesario tener una confianza del 95% de que la estimación del intervalo contuviera a la media poblacional verdadera, la estimación del intervalo seria:

$\bar{X} \pm 2\sigma_{\bar{X}} = \bar{X} \pm \frac{2\sigma_X}{\sqrt{n}}$ = estimación de intervalo del 95% de μ

En nuestro ejemplo el intervalo seria = 0.5 ± 2 * 0.47 = 0.5 ± 0.94

TAMAÑO DE LA MUESTRA Y TEORÍA ESTADÍSTICA

El tamaño del intervalo se basa en $2\sigma_{\bar{x}}$ porque, tal como se muestra en la figura de campana, la probabilidad de que X esté entre $\sigma_{\bar{x}}$ de la media poblacional es 0.95 (note que el intervalo anterior incluye la media poblacional μ. Además:

Interpretación de z

(Puntaje z, número de desviaciones estándar normales)

Como la distribución de la muestra es normal, se puede afirmar que:

a) El 68,27% de las veces, los promedios de la muestra caerán dentro de $\pm 1\sigma_{\bar{x}}$
b) El 95,45% de las veces, los promedios de la muestra caerán dentro de $\pm 2\sigma_{\bar{x}}$
c) El 99,73% de las veces, los promedios de la muestra caerán dentro de $\pm 3\sigma_{\bar{x}}$

Ejercicio: construya el intervalo de confianza para $1\sigma_{\bar{x}}$ y $3\sigma_{\bar{x}}$ y establezca el porcentaje de probabilidad correspondiente.

Importante: Si la desviación estándar ($\sigma_{\bar{x}} = \sigma$) es desconocida es necesario estimarla con la desviación estándar de la muestra, *S*.

TAMAÑO DE LA MUESTRA Y TEORÍA ESTADÍSTICA

Estimación del intervalo de confianza cuando ($\sigma_{\bar{x}} = \sigma$) es desconocida

$\bar{X} \pm 1S/\sqrt{n}$ ► estimación del intervalo de confianza de 68,27 % con σ desconocida

$\bar{X} \pm 2S/\sqrt{n}$ ► estimación del intervalo de confianza de 95,45 % con σ desconocida

$\bar{X} \pm 3S/\sqrt{n}$ ► estimación del intervalo de confianza de 99,73 % con σ desconocida

En nuestro ejemplo, sería ► $0.5 \pm 2\ (1.27/\sqrt{10}) = 0.50 \pm 0.80$

Intervalo con 95% de confianza, una z de 2 y un tamaño de muestra de 10.

Ojo: De este modo, el tamaño de la estimación del intervalo dependerá de tres factores. El primero es el nivel de confianza. Si estamos dispuestos a tener menos confianza de que la estimación del intervalo incluya a la media de la población verdadera y desconocida, entonces el intervalo será más pequeño. El segundo factor es la desviación estándar de la población. Si hay poca variación de la población, entonces la estimación del intervalo de la media poblacional será más pequeña. El tercero el es tamaño de la muestra. Conforme esta aumenta, el error de la muestra se ve reducido y el intervalo se volverá más pequeño.

TAMAÑO DE LA MUESTRA Y TEORÍA ESTADÍSTICA

LA PREGUNTA DEL TAMAÑO DE LA MUESTRA. Ahora, estamos finalmente listos para usar estos conceptos para que nos ayuden a determinar el tamaño de la muestra. Para proceder, el análisis debe especificar:

- El tamaño del error de la muestra que se desea
- El nivel de confianza, por ejemplo, un nivel de confianza del 95%.

Ojo: Esta especificación dependerá de las intercomposición entre el valor de la información más exacta y el costo por un incremento en el tamaño de la muestra. Para un nivel de confianza dado, un error muestral más pequeño tendrá un “costo” en términos de un tamaño de muestra más grande. Similarmente, para un error de la muestra dado, un nivel de confianza más alto tendrá un “costo” en términos de un tamaño muestral más grande. Estas afirmaciones se entenderán más fácilmente con algunos ejemplos.

TAMAÑO DE LA MUESTRA Y TEORÍA ESTADÍSTICA

Usando la fórmula general para la estimación del intervalo (recuerde que σ y σ_x son las mismas)

► $x \pm$ error muestral, ó $x \pm z\sigma/\sqrt{n}$

Asu vez sabemos que ► error muestral = $z\sigma/\sqrt{n}$

Dividiendo toda la expresión entre el error de la muestra y multiplicando por $\sqrt{n}$, nos queda

► $\sqrt{n} = z\sigma$ / error muestral

Despejando n tenemos ► $n = z^2\sigma^2$ / (error muestral)^2

Esta n representa la fórmula del tamaño de la muestra.

En nuestro ejemplo n es 99 (con desviación estándar de la población (σ = 1.49) conocida, una z = 2, y un 95% de confianza teniendo un error muestral de 0.3.

TAMAÑO DE LA MUESTRA Y TEORÍA ESTADÍSTICA

Cuando la desviación estándar de la población (σ) no es conocida, se utilizan varios enfoques. a) utilizar una σ proveniente de estudios anteriores, b) usar la desviación estándar de la muestra (*S*), tomada de una encuesta anterior o de una encuesta piloto, c) ponerse en el "peor de los casos".

- **Usar la desviación estándar de la muestra (*S*)**

► $n = z^2 S^2 / (\text{error muestral})^2$

Esta n representa la fórmula del tamaño de la muestra cuando no se conoce la desviación estándar de la población.

$$n = (2^2 * 1.27^2) / 0.80^2$$

En nuestro ejemplo n es 10 (con desviación estándar de la población (*S* = 1.27) conocida, una z = 2, y un 95% de confianza teniendo un error muestral de 0.80.

TAMAÑO DE LA MUESTRA Y TEORÍA ESTADÍSTICA

- **Ponerse en el "peor de los casos"**

En nuestro ejemplo, la varianza poblacional más grande ocurrirá si la mitad de la población respondiera con un +2 y la otra mitad con un -2. La varianza de la población seria entonces 4, y el tamaño de la muestra recomendado, a un nivel de confianza del 95% y un error permitido de 0.3, seria 178. Note que el tamaño de la muestra sería más grande que lo deseado, y por lo tanto la exactitud deseada se vería excedida. Lo lógica indica que es admisible cometer errores cuando se ha optado por ser demasiado exacto.

$\boldsymbol{\sigma^2_P = (\pi * (1 - \pi)) / n}$ donde

La varianza de la población sería de $0.5(2\text{-}0)^2 + 0.5(\text{-}2\text{-}0)^2 = 0.5 \times 4 + 0.5 \times 4 = 4$ Puesto que el 0.5 de la población respondió con +2, y la media de la población, o el promedio poblacional, sería cero.

π = la proporción de la población

P = la proporción de la muestra (correspondiente a $\bar{x}$), usada para estimar la proporción de la población desconocida)

σ^2_P = la varianza de la población de P

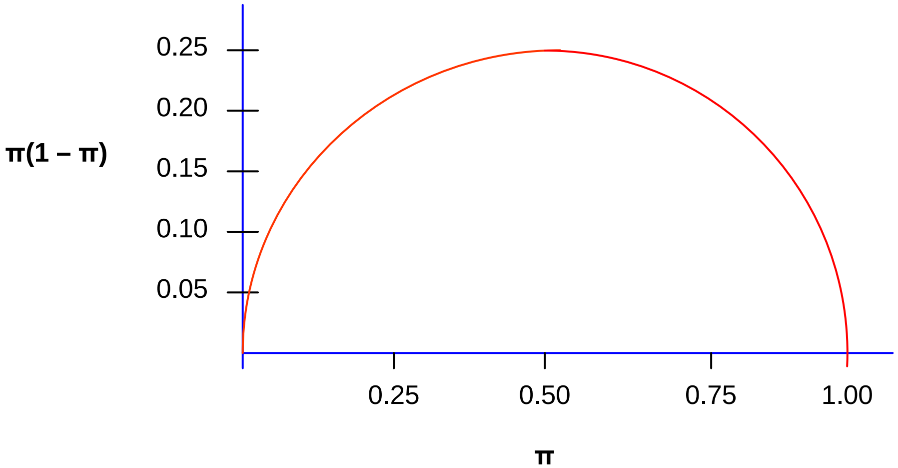
La fórmula para el tamaño de la muestra es entonces

$$\mathbf{n = (z^2\, \pi * (1 - \pi)) / (error\ muestral)^2}$$

TAMAÑO DE LA MUESTRA Y TEORÍA ESTADÍSTICA

Para nuestro ejemplo la varianza de la población sería de 0.5 (2-0)2 + (-2-0)2 = 0.5 * 4 + 0.5 * 4 = 4, puesto que el 0.5 de la población respondió con un +2, y la media de la población, o el promedio poblacional, seria cero.

El cálculo del error muestral en nuestro caso sería:



Como lo muestra la figura, el peor de los casos, es donde la varianza de la población está a su máximo, y esto ocurre cuando la proporción poblacional es igual a 0.50

$$\pi(1 - \pi) = 0.25$$

$$\pi = 0.50$$

TAMAÑO DE LA MUESTRA Y TEORÍA ESTADÍSTICA

Debido a que la proporción poblacional es desconocida, un procedimiento común consiste en supone el peor de los casos. La fórmula para el tamaño de muestra se simplifica a

$$n = (z^2 \times 0.25) / (\text{error muestral})^2$$

De este modo, si la proporción de la población debe ser estimada dentro de un error, de 0.05 (o 5 puntos porcentuales) a un nivel de confianza del 95%, el tamaño necesario de la muestra es

$$n = (2^2 \times 0.25) / (0.05)^2 = 400$$

Puesto que z = 2, correspondiente a un nivel de confianza del 95%, y el error muestral permitido es igual a 0.05.

En general tenemos un pequeño resumen de fórmulas:

- Tamaño de la muestra = $n = z^2\sigma / (\text{error muestral})^2$ ► Cuando σ se conoce
- Para las proporciones = $n = z^2(0.25) / (\text{error muestral})^2$ ► Cuando σ se desconoce