



Investigación en
Educación Médica

www.elsevier.es



METODOLOGÍA DE INVESTIGACIÓN EN EDUCACIÓN MÉDICA

Cálculo del tamaño de la muestra en investigación en educación médica

José Antonio García-García, Arturo Reding-Bernal, Juan Carlos López-Alvarenga

Departamento de Bioestadística y Bioinformática, Dirección de Investigación, Hospital General de México "Dr. Eduardo Liceaga", México D.F., México.

Recepción 7 de junio de 2013; aceptado 13 de agosto de 2013

PALABRAS CLAVE

Tamaño de muestra; cálculo; error estadístico; investigación en educación médica; México.

Resumen

Un aspecto importante en la metodología de la investigación, es el cálculo de la cantidad de participantes que deben incluirse en un estudio. El tamaño de muestra permite a los investigadores saber cuántos individuos son necesarios estudiar, para poder estimar un parámetro determinado con el grado de confianza deseado, o el número necesario para poder detectar una determinada diferencia entre los grupos de estudio, suponiendo que existiese realmente. El cálculo del tamaño de la muestra es una función matemática que expresa la relación entre las variables, cantidad de participantes y poder estadístico.

La muestra de un estudio debe ser representativa de la población de interés. El objetivo principal de seleccionarla es hacer inferencias estadísticas acerca de la población de la que proviene. La selección debe ser probabilística.

Los factores estadísticos que determinan el tamaño de la muestra son: hipótesis, error alfa, error beta, poder estadístico, variabilidad, pérdidas en el estudio y el tamaño del efecto.

Se revisan las fórmulas utilizadas para el cálculo del tamaño de la muestra en las situaciones más frecuentes en investigación, así como la revisión de fórmulas para un cálculo más rápido. Se incluyen ejemplos de investigación en educación médica. También se revisan aspectos importantes como: tamaño de la muestra para estudios piloto, estrategias para disminuir el número necesario de sujetos, y *software* para el cálculo del tamaño de muestra.

KEYWORDS

Sample size; calculation; population; statistical error; research in medical education; Mexico.

Sample size calculation in medical education research

Abstract

An important aspect in the research methodology, is the calculation of the number of participants that must be included in a study, since the sample size allows the researchers to

Correspondencia: José Antonio García García. Dr. Balmis N° 148, Colonia Doctores, Delegación Cuauhtémoc, C.P. 06726, México D.F., México. Teléfonos: 5004 3842, 5004 3843. Conmutador: 2789 2000, ext. 1164. Correo electrónico: drjagarcia2@prodigy.net.mx

know how many individuals it is necessary to study in order to estimate a parameter with the desired degree of confidence, or the number needed in order to detect a certain difference between the study groups, assuming that exist actually.

The calculation of the sample size is a mathematical function that expresses the relationship between the variables, amount of participants and statistical power.

A sample from a study should be representative of the population of interest. The main goal of selecting a sample is to make statistical inferences about the population from which comes from. The selection must be probabilistic.

Statistical factors that determine the sample size are: assumptions, error alpha, beta error, statistical power, and variability, losses in the study and size effect.

We review the formulas used for calculating the sample size in the most common situations in research, as well as the revision of formulas for a faster calculation.

It's included examples of research in medical education.

Also reviewed are important issues such as: sample size for pilot studies, strategies to reduce the required number of subjects, and software for the sample size calculation.

Introducción

Un aspecto relevante en la metodología de la investigación, es la estimación o cálculo de la cantidad de participantes que deben incluirse en un estudio. La primera reflexión que surge es ¿para qué sirve el cálculo del tamaño de la muestra? Permite a los investigadores saber cuántos individuos son necesarios estudiar, para estimar un parámetro determinado con el grado de confianza deseado o el número necesario para detectar una determinada diferencia entre los grupos de estudio, suponiendo que existiese realmente.

La inclusión de un número excesivo de sujetos encarece el estudio en varios aspectos. Un estudio con un tamaño insuficiente de la muestra estimará un parámetro con poca precisión o será incapaz de detectar diferencias entre los grupos, conduciendo a conclusiones erróneas.

En este documento se revisan los aspectos sobresalientes del tema, incluyendo los matemáticos utilizados para estimar el tamaño de la muestra.

Preámbulo

Grosso modo, puede considerarse que el objetivo de una investigación puede ser:

1. *Estimación de un parámetro*. Se pretende hacer inferencias a valores poblacionales (medias, proporciones), a partir de los resultados en una muestra. Por ejemplo, el porcentaje de estudiantes de pregrado con obesidad o el de alumnos que son aceptados para hacer una residencia médica.
2. *Contraste de hipótesis*. Aquí se tiene como propósito comparar si las medidas (medias, proporciones) de las muestras son diferentes. Por ejemplo, evaluar qué intervención educativa consigue un mayor porcentaje de éxitos.^{1,2}

El cálculo del tamaño de la muestra como una función matemática

El cálculo del tamaño de la muestra no es una simple operación aritmética que nos proporcione un valor. Es una función matemática, por lo tanto, el cambio de una variable,

necesariamente se acompaña del cambio de la otra considerada en la ecuación. Permite una mejor aproximación al número que se requiere, ajustando a su vez el poder estadístico con otros parámetros.

Se denota por: $y = f(x)$

donde:

y = variable dependiente (atributo o característica cuyo cambio es el que interesa medir, también se le denomina resultante o desenlace. En el cálculo del tamaño de la muestra, es el número de participantes que se necesitan).

x = variable independiente (atributo o característica que explica o predice el cambio en la variable dependiente. En el cálculo del tamaño de la muestra, un ejemplo es el poder estadístico que se requiere y que el investigador fija con antelación).

f = función (es una colección de pares de valores ordenados, que pertenecen a diferentes conjuntos. En el cálculo del tamaño de la muestra, los conjuntos se pueden ejemplificar con el poder estadístico y el número muestral resultante).

$f(x)$ = regla de correspondencia (expresa que para cada elemento de un conjunto se relaciona solamente con un elemento de otro conjunto. En el cálculo del tamaño de la muestra, para un elemento del poder estadístico se relaciona solamente con un número muestral).

En la **Figura 1** se ilustran dos ejemplos hipotéticos para la representación gráfica del concepto de función para la estimación del tamaño de muestra. Se utilizaron datos para modelos con diferencia de medias (gráfica izquierda) y para diferencia de promedios (gráfica derecha). La gráfica de la función es una línea, y sobre ella, los seguidores del método tradicional solicitan el resultado de las fórmulas aritméticas empleadas para el cálculo, que representa solamente un punto sobre la línea. Se utilizó el *software* Statistica® versión 8, para las estimaciones y representación gráfica del tamaño muestral.^{3,4}

Rigor en el cálculo del tamaño de muestra en ciencias

En las diferentes áreas de la investigación científica se debe tener rigor metodológico tanto para la elaboración

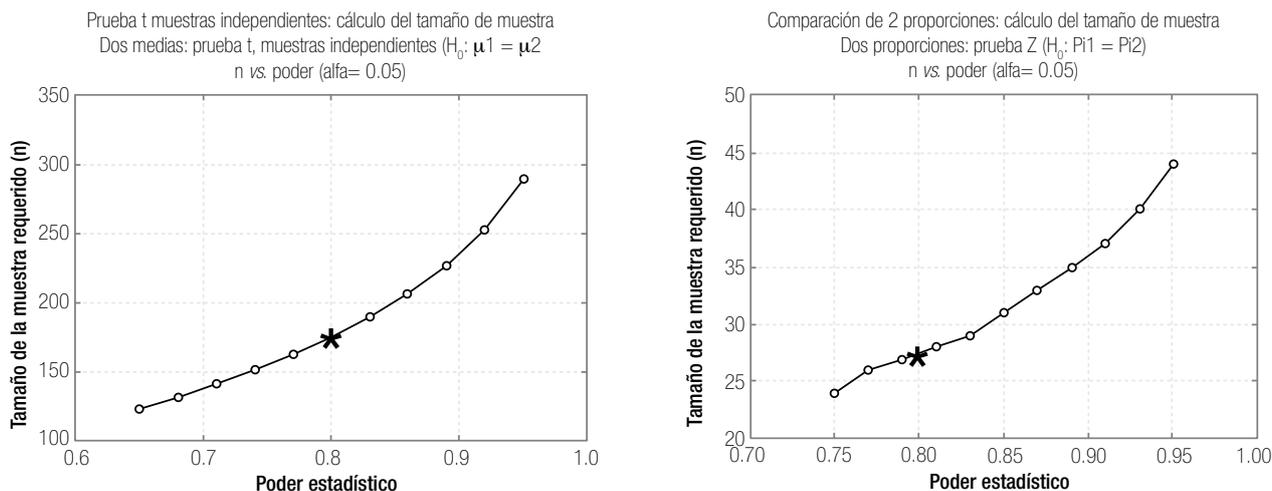


Figura 1. Representación gráfica de la función para el cálculo del tamaño de la muestra, tanto para comparar dos medias como dos proporciones. En el eje de las ordenadas se muestra el número de integrantes de la muestra y en el eje de las abscisas el poder estadístico. La función está representada por la línea, y el asterisco sobre ella representa el valor resultante de la fórmula matemática correspondiente, obteniendo así el tamaño de la muestra para un poder estadístico del 80%, que es utilizado habitualmente.

del protocolo, como para el desarrollo de las diferentes fases de la investigación. En este orden de ideas, es exigible la misma severidad para estimar el tamaño de la muestra en investigación en educación médica, que en otras áreas del conocimiento.⁵ Lo anterior aplica para la mayoría de los estudios contenidos en la brújula o compás de la investigación en educación médica.⁶

¿En dónde se anota el desarrollo del cálculo del tamaño de la muestra?

Los sitios en donde se desglosa este proceso son: el protocolo de la investigación, también aparece en las tesis de Maestrías y Doctorados en Ciencias Médicas y de la Salud y eventualmente en las de licenciatura. Pero no aparece en los artículos publicados, se da como un valor entendido que se realizó con rigor metodológico. Lo que aparece en los artículos científicos es la muestra en el estudio, pero no las variables y sus valores que se consideraron para la estimación del número.

Aspectos básicos en el proceso de muestreo. De población a muestra y viceversa

Población (cantidad representada en las fórmulas como N), es el conjunto total de elementos del que se puede seleccionar la muestra y está conformado por elementos denominados unidades de muestreo o unidades muestrales, con cierta ubicación en espacio y tiempo. Las unidades de muestreo pueden ser individuos, familias, universidades, grupos de alumnos, profesores, etc. Una muestra (cantidad representada en las fórmulas como n), no es más que un subconjunto de la población que se obtiene por un proceso o estrategia de muestreo.^{4,7}

El objetivo fundamental para seleccionar una muestra es hacer inferencias estadísticas (estimaciones de uno o más parámetros acerca de una población de interés). Esta

población es la que se desea investigar y se le denomina población de interés, blanco, objeto o diana. Para que la extrapolación (inferencia estadística) tenga validez, la muestra debe ser representativa, y alude a que el estimador muestral de las variables de interés debe tener una distribución similar a las de la población de donde proviene. Para cumplir este supuesto de representatividad es deseable que la muestra sea probabilística (Figura 2).⁸

Abraham Flexner, en su trascendental documento, incluyó al 100% de la población diana que fueron todas las escuelas de medicina de Estados Unidos de Norteamérica y Canadá. La muestra fue igual en número a la población, un hecho muy difícil de emular.⁹

Un aspecto diferente de muestreo es el caso de los estudios para determinar la eficacia y seguridad de algún medicamento, comparado con los tratamientos estándares o contra placebo. En ellos, el interés reside en contrastar hipótesis sobre una intervención (tratamiento o maniobra) que interesa al investigador. En este caso, el muestreo suele ser a conveniencia.¹⁰

Este artículo se enfoca en el diseño y la determinación del tamaño de la muestra para obtener representatividad o validez externa en las conclusiones.

Muestras y proceso de aleatorización en los estudios

Una vez que los sujetos de estudio son seleccionados, se hace una aleatorización para asignar la intervención que recibirá cada uno. En este caso es adecuado que la aleatorización se haga por bloques. Si el investigador conoce de antemano la existencia de factores que modifican la variable dependiente, es recomendable hacer estratos para controlar a la variable confusora, que es una variable predictora del cambio en la variable dependiente, externa a la relación principal que se analiza pero simultáneamente relacionada con la variable independiente. Cada estrato

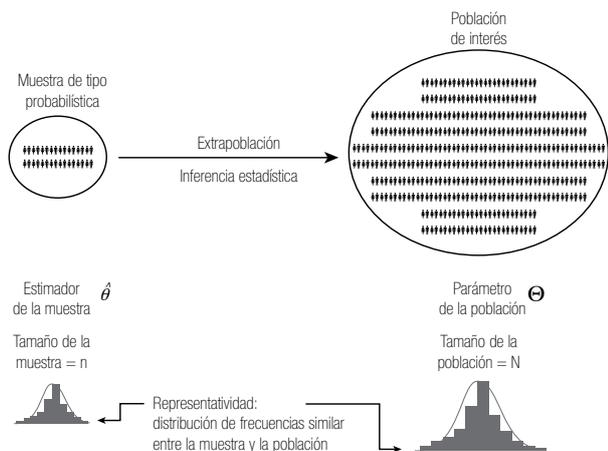


Figura 2. Interrelación entre población y muestra. La representatividad de una muestra probabilística permite hacer inferencias estadísticas a la población de interés. Modificado de López-Alvarenga JC, et al.⁸

se aleatoriza en forma independiente para lograr grupos balanceados en la intervención o tratamiento.¹¹

Recientemente se publicó un ensayo controlado y aleatorizado en investigación en educación médica, en donde la intervención o tratamiento fue un curso sobre medicina basada en evidencia de seis meses de duración, la muestra incluyó a los alumnos del quinto año de la licenciatura en medicina, los cuales fueron aleatorizados en dos grupos balanceados; es decir, con el mismo número de participantes. Como variables dependientes se midieron las actitudes, conocimientos y habilidades auto-reportadas, en ambos grupos.¹²

Factores para la determinación del tamaño de la muestra

Los factores que condicionan el tamaño de muestra, son de orden logístico o estadístico. Entre los primeros se encuentran las limitantes financieras o la disponibilidad de participantes. Los siguientes son los factores de orden estadístico que se desglosarán a continuación:⁸

1. Hipótesis.
2. Error tipo I o error α .
3. Error tipo II o error β .
4. Poder estadístico.
5. Variabilidad.

6. Pérdidas en el seguimiento del estudio.
7. Relevancia del tamaño del efecto y significancia estadística.

1. Hipótesis

De acuerdo con el tipo de estudio de investigación, puede ser necesario formular una o más hipótesis. Si se trata de un estudio tipo descriptivo, ésta no es necesaria. En los estudios de tipo comparativo es necesario establecerlas. En ambos casos, es necesario contrastar las hipótesis y determinar si se aceptan o se rechazan. Para ese contraste, las hipótesis toman el nombre de nula (H_0) o alternativa (H_1). El investigador desea probar la hipótesis alternativa, que significa rechazar la hipótesis nula. Al valor α (error tipo I) se le conoce como la probabilidad de que se rechace H_0 (se acepte H_1) cuando H_0 es cierta. Al valor β se le conoce como la probabilidad de que se acepte H_0 cuando es falsa (H_1 es cierta) (Tabla 1).^{7,13}

El tipo de contraste de hipótesis puede ser unilateral (una cola) o bilateral (dos colas). Una hipótesis unilateral especifica la dirección de la asociación (mayor o menor) de las variables; en la bilateral se puede afirmar la asociación entre las variables, pero no especifica la dirección. En el contraste bilateral el tamaño de muestra es más grande, estos contrastes también poseen mayor robustez y se prefieren a los de una cola. Cabe mencionar que el valor de Z de una distribución normal (distribución en el que el valor de la media igual a 0 y desviación estándar igual a 1) cambia dependiendo el tipo de contraste de hipótesis. En la Tabla 2 se muestran los valores frecuentemente utilizados de la distribución normal para $Z_{\alpha/2}$ (2 colas) o para Z_{β} (1 cola).^{2,14}

2. Error tipo I o error α

En un contraste de hipótesis, al valor α (error tipo I) se le conoce como la probabilidad de que se rechace H_0 (se acepte H_1) cuando H_0 es cierta. Es decir, $P(\text{aceptar } H_1 \mid H_0 \text{ es cierta}) = \alpha$. Al valor $(1 - \alpha) \cdot 100$ se le conoce como el nivel de confianza de la prueba. El valor de α varía dependiendo del nivel de confianza que se quiera de la prueba; el criterio más usado en la literatura biomédica es aceptar un riesgo de $\alpha < 0.05$.^{4,15}

3. Error tipo II o error β

A la probabilidad de que se acepte H_0 cuando ésta es falsa (H_1 es cierta), se le conoce como error tipo II o error β , es decir: $P(\text{aceptar } H_0 \mid H_1 \text{ es cierta}) = \beta$. El valor de β

Tabla 1. Interpretación de los posibles errores estadísticos en el contraste de hipótesis.

		Realidad en la población	
		Existen diferencias (H_0 falsa)	No existen diferencias (H_0 cierta)
Resultado de la investigación	Hay diferencia significativa (se rechaza H_0)	Correcto	Error tipo I (α)
	No hay diferencia significativa (se acepta H_0)	Error tipo II (β)	Correcto

Tabla 2. Valores frecuentemente utilizados de la distribución normal para Z.

α	$Z_{\alpha/2}$ (2 colas)	β	Z_{β} (1 cola)
0.1	1.65	0.2	0.84
0.05	1.96	0.1	1.28
0.01	2.58	0.05	1.65

tolerable de mayor aceptación en la comunidad científica va de 0.1 a 0.2.⁴

4. Poder estadístico

Es la probabilidad de que un estudio de un determinado tamaño detecte como estadísticamente significativa una diferencia que realmente existe.

Se define como $1 - \beta$. Es decir, $P(\text{aceptar } H_1 \mid H_1 \text{ es cierta}) = 1 - \beta$

Su valor depende del error tipo II que se acepte. Si $\beta = 0.2$, se tendrá una potencia de $1 - \beta = 0.8$. En términos porcentuales se dice que la prueba tiene una potencia del 80%, que es el mínimo aceptado en la literatura biomédica.

Cuanto menores sean los riesgos calculados para los errores alfa y beta, mayor será el tamaño muestral requerido. Cuanto menor sea la variabilidad, menor será la muestra estimada. A menor diferencia que se desea detectar, mayor será el número de participantes.^{2,16}

5. Variabilidad

Es la dispersión esperada de los datos. Se evalúa dependiendo de la variable de interés. Si éstas son numéricas continuas (grupo de valores infinitos que incluyen decimales), el tamaño de muestra estará determinado por la variable con el mayor coeficiente de variación (CV) [$CV(\bar{Y}) = (S_y/\bar{Y})$], donde S_y es la desviación estándar y \bar{Y} es la media. Por otra parte, cuando las variables de interés son categóricas, por convención se recomienda utilizar la estimación de la proporción que más se acerque a 0.5, ya que proporciona el mayor número muestral. Para determinar la variabilidad se debe recurrir a la literatura publicada de la variable de interés, cuando el dato no está disponible se usarán datos de pruebas piloto y en última instancia a estimaciones hechas por expertos.^{8,17}

6. Pérdidas en el seguimiento del estudio

Durante la realización del estudio, puede haber pérdidas de participantes por diversas razones. El tamaño mínimo de muestra necesario para obtener resultados estadísticamente significativos está pensado, de acuerdo con el número de sujetos al final del estudio y no con el inicial. Es recomendable adicionar al cálculo inicial, un 10% a 20% de participantes. Una forma sencilla de estimar el cálculo es: $n(1/1-R)$, donde n representa el número de participantes sin pérdidas, y R es la proporción de pérdidas esperadas.²

7. Relevancia del tamaño del efecto y significancia estadística

La magnitud de la diferencia del efecto que se desea detectar entre los grupos evaluados, es la condicionante más importante para el cálculo del tamaño de la muestra. Con frecuencia, la obtención de una diferencia estadísticamente significativa (diferencia en los resultados al contrastar dos o más valores o grupos con una prueba estadística, generalmente se fija un punto de corte para decir que si hay diferencias entre los valores. Por convención, lo más frecuente es aceptar la propuesta de Karl Pearson, que hay diferencias significativas cuando el valor de p es ≤ 0.05) no resulta relevante para el área en que se está investigación, práctica clínica, educación médica, etc. El investigador debe determinar si la magnitud de esa diferencia es relevante para el área de interés, independientemente de que haya sido estadísticamente significativa. Se espera que cualquier diferencia de relevancia también sea estadísticamente significativa.^{10,18}

Si en un estudio se han considerado los factores arriba descritos, pero no se ha anticipado que el resultado sea relevante en educación médica, pierde utilidad. Para ejemplificar: se realizó un estudio cuyo objetivo fue medir el conocimiento en medicina familiar de dos muestras de estudiantes que tomaron clases con profesores distintos, y el instrumento de medición del nivel de conocimiento fue un examen de opción múltiple de 100 ítems. Al momento de analizar estadísticamente los datos, se encontraron diferencias entre ambos grupos ($p < 0.05$), pero en el análisis se identificó que las diferencias fueron solamente del valor de dos respuestas, por lo anterior, se puede afirmar que hay diferencias estadísticamente significativas, pero carece de relevancia para la toma de decisiones educativas.

Al calcular el tamaño de la muestra se utilizan fórmulas matemáticas que consideran en forma simultánea varios de los siete factores estadísticos antes descritos, para la mayoría de ellos ya existen valores aceptados por convención o incluso asignados de manera arbitraria; al momento de sustituir valores en tales fórmulas nos encontramos que los rubros de variabilidad y tamaño del efecto requieren revisión bibliográfica, estudios piloto o la opinión de expertos para asignar un valor apropiado.

Tamaños de muestra de acuerdo a distintos diseños de muestreo

Para la determinación del tamaño de muestra, también hay que considerar el tipo de diseño empleado en la investigación. Existen diseños de tamaño fijo (los más usados en estudios clínicos, epidemiológicos y en investigación educativa) y de tamaño variable. En los de tamaño fijo, el tamaño de muestra se fija desde el inicio de la investigación; en los estudios de tamaño variable, el número de sujetos se irá incrementando hasta obtener un tamaño predeterminado (diseño secuencial) o el diseño experimental que involucra un solo caso. En el resto del documento sólo se hace referencia a los diseños de tamaño fijo.^{2,4}

La mayoría de las fórmulas utilizadas para el cálculo del tamaño de muestra, parten del supuesto de una distribución normal de los valores de las variables en cuestión;

sin embargo, existen herramientas estadísticas para analizar los datos cuando ese supuesto no se cumple.

1. Cálculo del tamaño de muestra de una media

El intervalo de confianza para estimar la media poblacional a partir de una muestra es el siguiente: $IC = \bar{y} \pm \delta$, donde \bar{y} es la media estimada a partir de la muestra,

$$\delta = Z_{\alpha/2} * \left(\frac{s}{\sqrt{n}} \right), \quad s = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}, \quad \text{y } Z_{\alpha/2} \text{ es el valor del eje de las abscisas de la función normal estándar en donde se acumula la probabilidad de } (1-\alpha).$$

Cuando n es muy pequeña, $Z_{\alpha/2}$ podría sustituirse por $t_{(n-1), \alpha/2}$. Entonces al despejar n se tiene $n = \frac{Z_{\alpha/2}^2 * S^2}{\delta^2}$. En muestras finitas

donde la población es inferior a un millón, la fórmula para el cálculo del tamaño de la muestra se suele multiplicar por el factor de corrección por finitud $\left(1 - \frac{n}{N}\right)$, quedando la estimación del tamaño de muestra $n = \frac{Z_{\alpha/2}^2 * S^2}{\delta^2} * \left(1 - \frac{n}{N}\right)$.

El error de estimación o absoluto (δ) se obtiene de una muestra piloto o de estudios previos.^{4,8}

2. Cálculo para determinar el tamaño de muestra de una proporción

El tamaño de muestra de una proporción se calcula como

$$\text{sigue: } n = \frac{Z_{\alpha/2}^2 * p(1-p)}{\delta^2} * \left(1 - \frac{n}{N}\right), \quad \text{donde } \delta = Z_{\alpha/2} * \sqrt{\frac{p(1-p)}{n}}, \text{ se conoce como "precisión" del muestreo o error de la estimación}$$

$\left(1 - \frac{n}{N}\right)$, es el factor de corrección por finitud de la población, p es la proporción estimada del parámetro poblacional y $Z_{\alpha/2}$ es el valor del eje de las abscisas de

la función normal estándar, en donde se acumula la probabilidad de $(1-\alpha)$. El error absoluto (δ) se obtiene de una muestra piloto o estudios previos. Si no puede determinarse esta proporción, se tomará a $p = 0.5$, porque este valor garantizará el mayor tamaño de muestra. El nivel de confianza $(1-\alpha) * 100$ que suele utilizarse en estas pruebas por lo general es del 95%. El intervalo de confianza para una proporción queda definido de la siguiente manera

$IC = p \pm \delta = p \pm Z_{\alpha/2} * \sqrt{\frac{p(1-p)}{n}}$.^{10,19}

3. Cálculo para el tamaño de muestra de la diferencia de dos medias independientes

La fórmula es: $n_c = n_e = \frac{2 * S^2}{D^2} * \left(Z_{\alpha/2} * Z_{\beta} \right)^2$, donde n_c es el tamaño de muestra para el grupo de referencia y n_e es el del

grupo con una intervención alternativa, $D = (M_c - M_e)$, M_c es la media del primer grupo y M_e es la media del segundo, S^2 es la varianza de ambas distribuciones, que se suponen iguales, Z_{β} es el valor del eje de las abscisas de la función normal estándar en donde se acumula la probabilidad de $(1-\beta)$. Esta fórmula para estimar $n_c = n_e$ se emplea cuando se trata de un contraste de hipótesis bilateral; en caso de un contraste unilateral, se sustituirá $Z_{\alpha/2}$ por Z_{α} .^{2,10}

4. Cálculo para el tamaño de muestra de la comparación de dos medias repetidas (pareadas) en un solo grupo

La fórmula es: $n_c = n_e = \frac{(Z_{\alpha/2} + Z_{\beta})^2 * S^2}{d^2}$, donde d es el promedio

de las diferencias individuales entre los valores basales y posteriores, S^2 es la varianza de ambas distribuciones, que se suponen iguales. $Z_{\alpha/2}$ es el valor del eje de las abscisas de la función normal estándar, en donde se acumula la probabilidad de $(1-\alpha)$ para un contraste de hipótesis bilateral y Z_{β} es el valor del eje de las abscisas de la función normal estándar, en donde se acumula la probabilidad de $(1-\beta)$.¹⁹

5. Cálculo para el tamaño de muestra de la comparación de dos medias repetidas en dos grupos distintos de participantes

Se utiliza cuando se quiere comparar el cambio entre una medida basal y otra posterior de dos grupos distintos de sujetos. La fórmula para la estimación del tamaño de muestra de los grupos es la siguiente:

$$n_c = n_e = \frac{2 * (Z_{\alpha} + Z_{\beta})^2 * (1-\rho) * S^2}{|M_{dc} - M_{de}|^2},$$

donde M_{dc} es la diferencia entre los valores iniciales y los finales en el grupo de los controles y M_{de} es la diferencia entre los valores iniciales y finales en el grupo con tratamiento.^{8,19}

6. Cálculo para estimar el tamaño de muestra de la diferencia de dos proporciones

La fórmula es: $n_c = n_e = \frac{p_1(1-p_1) + p_2(1-p_2)}{(p_1 - p_2)^2} * (Z_{\alpha/2} + Z_{\beta})^2$, don-

de p_1 es la proporción del primer grupo y p_2 es la proporción del segundo grupo a comparar y $(p_1 - p_2)$ es la diferencia de las proporciones entre ambos grupos, $Z_{\alpha/2}$ es el valor del eje de las abscisas de la función normal estándar en donde se acumula la probabilidad de $(1-\alpha)$ para un contraste de hipótesis bilateral y Z_{β} es el valor del eje de las abscisas de la función normal estándar, en donde se acumula la probabilidad de $(1-\beta)$.² Un ejemplo es el Reporte Nacional del Estatus de la Educación Médica en EUA.²⁰

7. Cálculo para el tamaño de muestra de la comparación de dos proporciones independientes

Cuando se tiene una tabla de contingencia de 2×2 y las condiciones se cumplen para aplicar una prueba *ji cuadrada*,

se puede utilizar esta aproximación para el cálculo del tamaño de la muestra de la comparación de proporciones independientes. La fórmula que Marragat y colaboradores

proponen es: $n_{c-e} = n = \frac{[Z_{\alpha} * \sqrt{2 * P * Q} + Z_{\beta} * \sqrt{P_c * Q_c + P_e * Q_e}]^2}{(P_e - P_c)^2}$, donde

P es la proporción media de la proporción de eventos de interés del grupo control (c) y en el grupo en tratamiento (e), $Q_c = 1 - P_c$, P_c es la proporción de eventos de interés en el grupo control, $Q_e = 1 - P_e$, P_e es la proporción de eventos de interés en el grupo expuesto o en tratamiento, $Q_e = 1 - P_e$, y $(P_e - P_c)$ es la diferencia de las proporciones entre el grupo control y la proporción del grupo de expuestos.^{8,19}

8. Opción rápida y aceptable para el cálculo del tamaño muestral

Existe una fórmula simplificada para el cálculo del tamaño muestral para comparar dos medias, cuando se acepta un error bilateral alfa del 5% y una potencia del 80%.² Si se denomina diferencia estandarizada (DE) al cociente entre las diferencias de medias d y la desviación estándar s , tenemos: $DE = d/s$, por lo que, una fórmula abreviada, que sirve para estimar muy aproximadamente el tamaño de la muestra, es:

$$n = 16 / (DE)^2$$

Cuando esta fórmula es utilizada para comparar dos proporciones,² la expresión es:

$$n = 16 p_m q_m / d^2$$

Consideraciones especiales

1. Tamaño de muestra para estudios piloto

Se recomienda incluir entre 30 y 50 participantes, los cuales deben poseer los atributos que se desean medir en la población objetivo.²¹

2. Estrategias para minimizar el número necesario de participantes

Se basan en conseguir una población homogénea (desde los criterios de selección), disminuir la variabilidad de las medidas (aleatorizando, formando bloques) y aumentar la frecuencia de aparición del fenómeno de interés, por lo que deben aplicarse siempre que sea posible.^{2,17}

3. Software de utilidad

El uso de internet facilita obtener el tamaño de muestra empleando programas en línea. Los programas utilizan diferentes algoritmos matemáticos para efectuar el cálculo, y aunque esencialmente utilizan los mismos elementos, puede haber ligeras diferencias en el número de la muestra.

Entre los programas más utilizados están EPIDAT[®], G*Power[®] y Epi Info[®],⁸ de acceso libre. Hojas de cálculo como Excel^{®2}, también son de utilidad. Entre los *software* de paga destacan Stata[®],¹⁶ SAS[®],²² STATISTICA^{®3} y Sigma-Plot[®],²³ por mencionar sólo algunos. Los dos últimos tienen la ventaja de poder graficar las funciones de estimaciones del tamaño de la muestra.

Hay que usarlos críticamente, siendo necesario comprender bien los principios del cálculo.

Conclusiones

La investigación educativa debe tener el mismo rigor metodológico que otras áreas científicas, incluido el cálculo del tamaño de la muestra. Hay que practicar una y otra vez, es decir, ser activos, para poder ser competente en la conceptualización de cómo estimar la función matemática del tamaño muestral. Al respecto, Abraham Flexner escribió “la medicina moderna, como toda enseñanza científica, está caracterizada por la actividad. Las conferencias y los libros no son sustitutos de las experiencias”.

Contribución de los autores

JAGG, generador de la propuesta, búsqueda, recuperación y análisis de la información relacionada con el tema y redacción del manuscrito.

ARB, aportación de ideas para la estructura del documento, redacción del documento.

JCLA, asesoría continua, aportó comentarios y revisión del manuscrito.

Financiamiento

Ninguno.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Presentaciones previas

Ninguna.

Referencias

- Argimon PJM, Jiménez VJ. Métodos de investigación clínica y epidemiológica. 4ª edición. España: Elsevier; 2012. p. 140-158.
- Martínez GMA, Sánchez VA, Faulín FJ. Bioestadística amigable. 2ª edición. España: Díaz de Santos; 2006. p. 373-417.
- Consultado el 22 de marzo de 2013. <http://www.statsoft.com/textbook/power-analysis/>
- Fox N, Hunn A, Mathers N. Sampling and sample size calculation. The National Institutes for Health Research. USA: NIHR RDS EM/ YH; 2009. p. 12-24.
- Cook DA, Beckman TJ. Reflections on experimental research in medical education. Adv Health Sci Edu Theory Pract 2010;15(3):455-464.
- Ringsted C, Hodges B, Scherpbier A. “The research compass”: An introduction to research in medical education: AMEE Guide No 56. Med Teach 2011;33:695-709.
- Bennett JO, Briggs WL, Triola MF. Razonamiento estadístico. México: Pearson Educación; 2011. p. 333-361.
- López-Alvarenga JC, Reding-Bernal A. Cálculo del tamaño de la muestra: enfoque práctico de sus elementos necesarios. En: García-García JA, Jiménez-Ponce F, Arnaud-Viñas MR (eds.). Introducción a la metodología de la investigación en ciencias de la salud. México: McGraw-Hill Interamericana; 2011. p. 67-76.
- Consultado el 27 de febrero de 2013. http://www.carnegiefoundation.org/sites/default/files/elibrary/Carnegie_Flexner_Report.pdf (
- López-Alvarenga JC, Reding-Bernal A, Pérez-Navarro M, et al. Cómo se puede estimar el tamaño de la muestra de un estudio. Dermatol Rev Mex 2010;54(6):375-379.

11. Box GE, Hunter JS, Hunter WG. Estadística para investigadores. Diseño, innovación y descubrimiento. 2ª edición. España: Editorial Reverté; 2008. p. 133-172.
12. Sánchez-Mendiola M, Kieffer-Escobar LF, Marín-Beltrán S, et al. Teaching of evidence-based medicine to medical students in Mexico: a randomized controlled trial. *BMC Med Educ* 2012;12: 107.
13. Landero HR, González RMT. Estadística con SPSS y metodología de la investigación. México: Trillas; 2007. p. 67-75.
14. Cobo E, Muñoz P, González JA. Bioestadística para no estadísticos. España: Elsevier; 2007. p. 212-228.
15. Elorza PTH. Estadística para las ciencias sociales, del comportamiento y de la salud. 3ª edición. México: CENGAGE Learning; 2008. p. 319-338.
16. Acock AC. A gentle introduction to Stata. 3th edition. Texas: Stata Press; 2012. p.170-177.
17. Hulley SB, Cummings SR, Browner WS, et al. Design clinical research. 3th edition. Philadelphia, USA: Lippincott, Williams & Wilkins; 2007. p. 65-69.
18. Sullivan GN, Feinn R. Using effect size - or why the p value is not enough. *J Grad Med Educ* 2012;4:279-282.
19. Marrugat J, Vila J, Pavesi M, et al. Estimación del tamaño de muestra en la investigación clínica y epidemiológica. *Med Clin* 1998;111:267-276.
20. Sullivan AM, Lakoma MD, Block SD. The status of medical education in end-of-life care. A National Report. *J Gen Intern Med* 2003;18:685-695.
21. Babbie E. Fundamentos de la investigación social. 3ª edición. México: Thomson editores; 2000. p. 232-256.
22. Consultado el 08 de mayo de 2013. <http://www.sas.com/technologies/analytics/statistics/stat/index.html>
23. Consultado el 02 de mayo de 2013. <http://www.sigmaplot.com/products/sigmaplot/sigmaplot-details.php#sa>.