

Juan O. Talavera,
Rodolfo Rivas-Ruiz,
Laura Paola Bernal-Rosales

Investigación clínica V. Tamaño de muestra

Centro de Adiestramiento en Investigación Clínica,
Coordinación de Investigación en Salud,
Centro Médico Nacional Siglo XXI,
Instituto Mexicano del Seguro Social, Distrito Federal, México

Comunicación con: Juan O. Talavera
Correo electrónico: jotalaverap@uaemex.mx

Resumen

En investigación clínica resulta imposible e ineficiente estudiar a todos los pacientes con una patología específica, por lo que es necesario abocarse a una muestra. La estimación del tamaño de muestra garantiza la estabilidad de los resultados y permite prever la factibilidad del estudio dependiendo del costo y de la disponibilidad de los pacientes. La estructura básica de estimación del tamaño de muestra parte de la premisa que busca demostrar —entre otras cosas— que la diferencia entre dos o más maniobras en el estado subsecuente es real. Para ello se requiere conocer el valor de la diferencia esperada (δ) y la medida de dispersión de los datos que le dieron origen (desviación estándar), que suelen obtenerse de estudios previos. Posteriormente se consideran otros componentes: α , porcentaje de error aceptado en la aseveración de que la diferencia entre las medias es real, generalmente de 5 %; y β , porcentaje de error aceptado en la aseveración de que la no diferencia entre las medias es real, generalmente de 15 a 20 %. Estos valores se sustituyen en la fórmula o en algún programa electrónico de estimación del tamaño de muestra. Si bien las medidas de resumen y dispersión pueden variar conforme la medición del resultado y consecuentemente la fórmula, el principio es el mismo.

Palabras clave

tamaño de muestra
intervalo de confianza

Summary

In clinical research it is impossible and inefficient to study all patients with a specific pathology, so it is necessary to study a sample of them. The estimation of the sample size before starting a study guarantees the stability of the results and allows us to foresee the feasibility of the study depending on the availability of patients and cost. The basic structure of sample size estimation is based on the premise that seeks to demonstrate, among other cases, that the observed difference between two or more maneuvers in the subsequent state is real. Initially, it requires knowing the value of the expected difference (δ) and its data variation (standard deviation). These data are usually obtained from previous studies. Then, other components must be considered: α (alpha), percentage of error in the assertion that the difference between means is real, usually 5 %; and β , error rate accepted the claim that the no-difference between the means is real, usually ranging from 15 to 20 %. Finally, these values are substituted into the formula or in an electronic program for estimating sample size. While summary and dispersion measures vary with the type of variable according to the outcome, the basic structure is the same.

Key words

sample size
confidence interval

Introducción

En investigación clínica es imposible e ineficiente estudiar a todos los sujetos enfermos con una patología específica, en consecuencia cuando leemos un artículo los resultados que nos muestran corresponden a una porción del total de la población. El número de sujetos incluidos en un estudio está determinado por una serie de características que se men-

cionarán más adelante, pero cuyo objetivo principal es responder a una pregunta con la certeza de que el resultado encontrado es real. Aunado a esto, la estimación del tamaño de muestra antes de iniciar un estudio permite considerar la factibilidad del mismo dependiendo de la disponibilidad de pacientes y del costo. La ausencia del cálculo en el tamaño de muestra provoca una dispensa innecesaria de recursos tanto económicos como humanos. Es posible que se

incrementen los gastos del estudio en forma innecesaria por el número excedente de sujetos que se incluyan en el mismo, o que la inversión realizada sea infructuosa al incluir un número de sujetos insuficiente para contestar la pregunta de investigación.

La estructura básica de estimación del tamaño de muestra parte de la premisa que busca demostrar que la diferencia observada entre las mediciones realizadas antes y después de una maniobra, o entre dos maniobras en el estado subsecuente, es real y no se debe a efectos del azar. Esta estructura es la misma independientemente del tipo de variables necesarias para contestar la pregunta de investigación. En otros casos no se desea demostrar la veracidad de una diferencia sino obtener el valor promedio de una determinada característica dentro de una población con una precisión marcada por el límite superior e inferior del intervalo de confianza (IC), que en la mayoría de los casos se solicita de 95 o 99 %.

Estimación para dos grupos

Este propósito queda ejemplificado cuando se desea demostrar que las cifras de presión sanguínea son diferentes con cierto fármaco u otro y que esa diferencia no se debe al azar. Para estimar el tamaño de muestra lo primero que se necesita es el promedio (\bar{x}), en este ejercicio, de las cifras de tensión arterial diastólica (TAD) de los pacientes que tomaron un fármaco (grupo A) u otro (grupo B): en el supuesto de que el promedio de TAD en el grupo A sea de 90 mm Hg y en el

grupo B de 85 mm Hg, entonces la diferencia entre las medias será de 5 mm Hg, valor que representa el primer componente, el cual se identifica como *delta* (δ).

En segundo término será necesario disponer de alguna medida de la variación de los valores al interior de cada grupo dado que habrá pacientes con presiones mucho menores y mucho mayores del promedio, por ejemplo, 60 a 112 mm Hg. Con este valor se podrá observar la variación al interior de cada grupo y, al mismo tiempo, conocer si los valores entre grupos se sobrepone en demasía en relación con la diferencia de promedio. En una variable cuantitativa, como en el modelo descrito, la medida de dispersión se conoce como desviación estándar (DE).

Como se observa en la figura 1, el promedio de TAD para la población total es de 87 mm Hg, con una desviación estándar de 9 mm Hg; mientras que en la figura 2a, el promedio de TAD del grupo A es 90 ± 9 mm Hg ($\bar{x} \pm DE$) y el promedio de TAD (figura 2b) del grupo B es 85 ± 8 mm Hg ($\bar{x} \pm DE$). Esto significa que la población general tiene un promedio de 87 mm Hg pero que sus valores a dos desviaciones estándar van de 69 a 105 mm Hg ($\bar{x} \pm 2 DE$). En el grupo A con promedio de 90 mm Hg sus valores van de 72 a 108 mm Hg ($\bar{x} \pm 2 DE$), y en el B con promedio de 85 mm Hg sus valores van de 69 a 101 mm Hg ($\bar{x} \pm 2 DE$). Los valores promedios y de dispersión de la variable de interés suele obtenerse de información existente en estudios previos ya publicados o preliminares.

Una vez con una medida de resumen (promedio) y su correspondiente medida de dispersión (DE), hay que considerar:

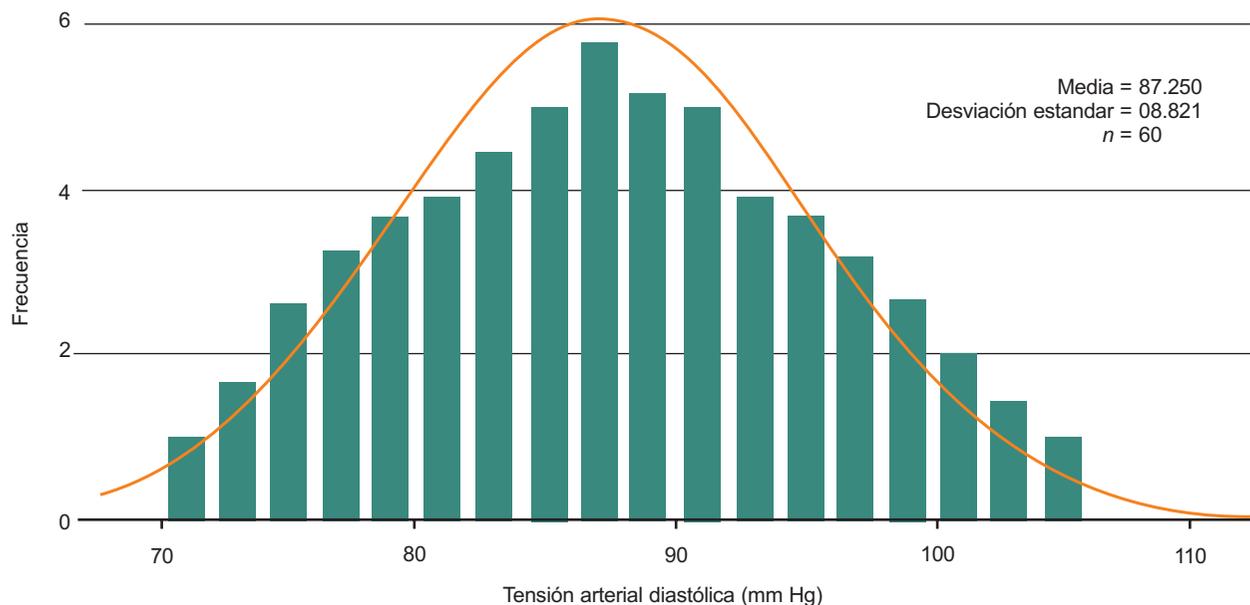


Figura 1 | Grupo total de pacientes hipertensos bajo tratamiento farmacológico

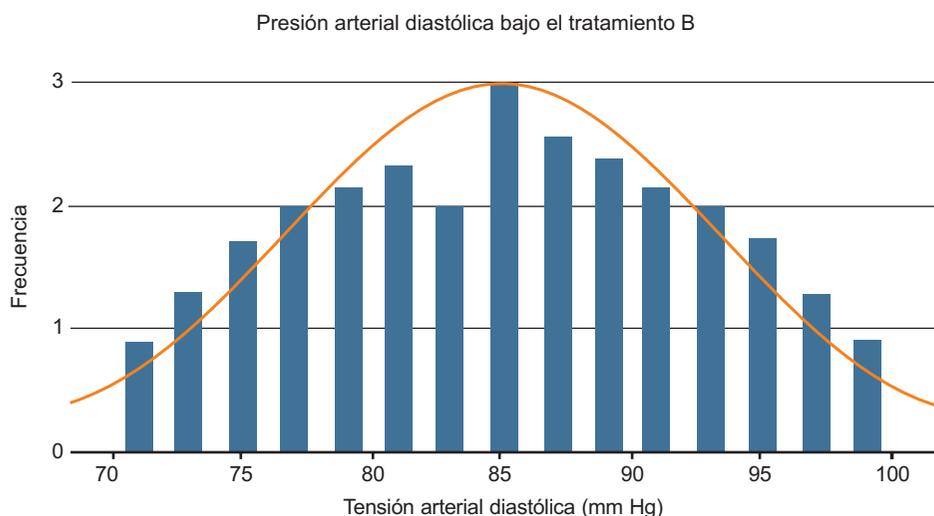
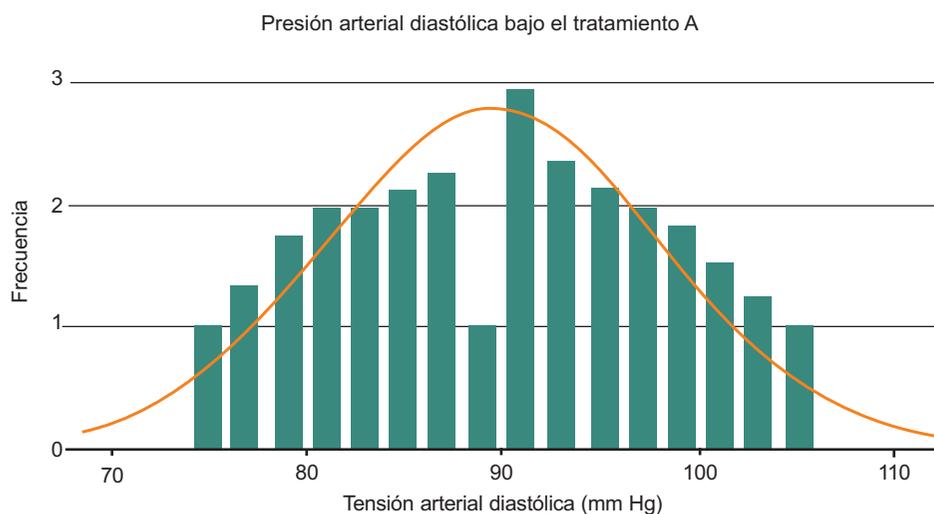


Figura 2 Pacientes hipertensos bajo tratamientos A y B, respectivamente

1. ¿Con qué certeza queremos demostrar que la diferencia de TAD entre grupos es real? Cuando este punto no se toma en cuenta podemos incurrir en lo que se conoce como error tipo I: aceptar que la diferencia es real sin que lo sea.
2. ¿Con qué certeza queremos demostrar que la no-diferencia es real? Cuando este punto no se toma en cuenta podemos caer en lo que se conoce como error tipo II: aceptar que la no-diferencia es real.

La certeza con la que habitualmente se acepta que una diferencia es real está en 95 % y corresponde a un valor de alfa (α) de 0.05, lo cual indica que al determinar que hay diferencia en la cifra de TAD entre los grupos existe 95 % de seguridad de que esa diferencia sea real y solo se acepta un error de 5 %.

Para aceptar que la no-diferencia encontrada es real, se debe iniciar con una capacidad preestablecida de encontrar significancia cuando hay una diferencia, lo que se conoce como poder y se representa por la diferencia de $1 - \beta$. El valor del poder aceptado puede variar de 80 a 95 % y corresponde a una β de 5 a 20 %.

Para este momento se dispone ya de todos los componentes necesarios para estimar el tamaño de muestra:

- δ : diferencia entre las medidas de resumen (en el ejemplo es la diferencia entre las medias).
- DE: medida de dispersión, que en el ejemplo es desviación estándar.
- α : error aceptado en la aseveración de que la diferencia entre las medias es real, generalmente de 5 % (0.05).

- β : error aceptado en la aseveración de que la no-diferencia entre las medias es real, generalmente va de 5 a 20 %.

Ignorar estos distintos componentes comúnmente provoca que al final del estudio el tamaño de muestra sea insuficiente, con lo que a pesar de contar con una diferencia clínica significativa ($\geq 10\%$) no se encuentre una diferencia estadística ($p < 0.05$), lo que significa un poder bajo ($< 80\%$) y por lo tanto un error tipo II.

Diferencia de medias

Con los componentes anteriores se estima el tamaño de muestra mediante la fórmula para diferencia de medias:

$$n = 2 \left[\frac{(Z_\alpha - Z_\beta) DE}{\mu_1 - \mu_2} \right]^2$$

Donde:

Z_α = valor de z relacionado con $\alpha = 0.05$ (se extrae de tablas de referencia)

Z_β = valor de z relacionado con un $\beta = 0.20$ (poder de 80 %).

DE = desviación estándar

μ_1 = media de grupo A

μ_2 = media de grupo B

De acuerdo con el ejemplo, la sustitución de los valores sería de la siguiente forma:

$$Z_\alpha = 1.96$$

$$Z_\beta = -0.84$$

$$DE = 9 \text{ mm Hg}$$

$$\mu_1 = 90 \text{ mm Hg}$$

$$\mu_2 = 85 \text{ mm Hg}$$

Y sustituyendo en la fórmula:

$$n = 2 \left[\frac{(1.96 - (-0.84))9}{90 - 85} \right]^2 = 50.80 \quad 51$$

Por lo tanto, es necesario incluir 51 pacientes en cada grupo si se desea obtener 80 % de posibilidad (80 % de poder) para detectar una diferencia de medias de 5 mm Hg o más entre los dos grupos de tratamiento.

Diferencia de proporciones

Se utiliza cuando el resultado de interés está expresado en términos de proporciones. Ejemplo: comparación de dos gru-

pos de pacientes con sobrepeso. El primer grupo de pacientes recibe fármaco y el segundo, orientación alimentaria. Si el evento de resultado se evalúa después de seis meses y se mide como la proporción de pacientes que logran normalizar su peso (índice de masa corporal menor a 25), ¿qué se necesita?

$$\alpha = 0.05$$

$$\beta = 0.10$$

$\pi_1 - \pi_2$ = (diferencia de proporciones) proporción del grupo 1 menos la proporción del grupo 2, que sea clínicamente significativa

DE = la fórmula para su determinación es 1 - proporción de grupo, que queda incluida dentro de la fórmula global

La fórmula para la determinación del tamaño de muestra para diferencia de proporciones es:

$$n = \left[\frac{Z_\alpha \sqrt{2\pi_1(1-\pi_1)} - Z_\beta \sqrt{\pi_1(1-\pi_1) + \pi_2(1-\pi_2)}}{\pi_1 - \pi_2} \right]^2$$

Donde:

$$Z_\alpha = (\alpha = 0.05) \quad 1.96$$

$$Z_\beta = (\beta \approx 0.10-0.20) \approx -1.645, -0.84$$

π_1 = proporción de grupo 1

π_2 = proporción de grupo 2

$\pi_1 - \pi_2$ = diferencia entre proporción de grupo 1 - proporción de grupo 2, que sea clínicamente significativa

Suponiendo que para el problema de estudio se esperara que a seis meses el grupo que recibe tratamiento farmacológico tenga éxito en 70 % de los casos, mientras que el grupo con orientación alimentaria tenga éxito en 50 % de los casos, se sustituirán los valores en la fórmula de la siguiente manera:

$$n = \left[\frac{1.96 \sqrt{2 \times 0.70 \times 0.30} - (-1.645) \sqrt{(0.70 \times 0.30) + (0.50 \times 0.50)}}{0.70 - 0.50} \right]^2$$

$$n = \left[\frac{2.435}{0.20} \right]^2 = 12.18^2 = 148.35 \text{ sujetos por grupo}$$

El resultado anterior se tiene que redondear al dígito superior, de tal forma la muestra debe incluir 149 sujetos en cada grupo de estudio si se quiere tener 90 % de posibilidad (90 % de poder) para detectar al menos una diferencia de 20 % en el porcentaje de éxito de pérdida de peso entre los dos grupos de tratamiento que se ejemplifican.

Estimación para un grupo

Por otro lado, cuando el objetivo es obtener el valor promedio de una determinada característica dentro de una población, lo que se requiere para estimar el tamaño de muestra es el valor promedio (proporción o media) y su límite superior e inferior marcado por el IC, que en la mayoría de las ocasiones se solicita de 95 o 99 %.

Para una proporción

Para estimar el tamaño de muestra de una prevalencia o proporción de un evento o característica se deben identificar distintos componentes, empezando por la medida de resumen (p_0), que corresponde a la proporción esperada y la precisión de la misma (d), que equivale a la mitad de la amplitud del IC. Entendiendo este apartado podemos despegar la fórmula de tamaño de muestra a partir de la fórmula de la precisión, que a su vez proviene de la estimación de la desviación estándar de una proporción:

$$d = Z_{\alpha} \frac{\sqrt{p_0 \times q_0}}{n}$$

Despejando n se obtiene:

$$n = \frac{Z_{\alpha} \times p_0 \times q_0}{d^2}$$

En este caso, $q_0 = (1 - p_0)$, por lo tanto si queremos buscar una prevalencia (p_0) de 20 % el valor de q_0 sería $1 - 0.2 = 0.8$. Por tanto, para hacer el cálculo del tamaño de muestra para una proporción debe considerarse lo siguiente:

- La precisión (d , igual a la 1/2 de la amplitud del IC), cuyo valor en general es conferido por el investigador y se corresponde con el grado de error que pudiera ser tolerado hacia cada lado de la media; por ejemplo, para un error de 8 % a partir de la media, su d^2 sería de 0.0064 ($0.08^2 = 0.0064$).
- La confianza, también llamada Z_{α} , corresponde a $1 - \alpha$.
- El valor de p_0 que se pretende estimar.

Ejemplo: ¿cuántos prematuros será necesario estudiar para verificar si la prevalencia estimada de enfermedad ósea metabólica en una población de la unidad de cuidados intensivos neonatales es de 20 %, considerando una precisión de 8 % y un α de 0.05 %?

Con un nivel de confianza de 95 % ($\alpha = 0.05$; $Z_{\alpha} = 1.96$), $Z_{\alpha}^2 = 3.8416$, que al despejar:

$$N = 3.8416 \times 0.8 \times 0.2 / 0.0064$$

$$N = 96.04$$

Por lo tanto, el tamaño de muestra necesario será de 97 niños para una prevalencia esperada de 20 % con un IC que va de 12 a 28 %.

Como podemos observar, el tamaño de la muestra dependerá de la precisión esperada del error a partir de la media, de modo que para un IC más estrecho se requiere una d menor, en general se usan los valores de 0.08 y de 0.04, siendo este último el de mayor precisión (o menor error) y, por lo tanto, se necesitará un mayor tamaño de la muestra. De la misma manera, si se desea cambiar el nivel de confianza, de 95 a 99 % como se exige en los estudios de determinantes genéticos, el tamaño de muestra se incrementará. En el cuadro I se muestran algunos ejemplos de variación dependiendo de estos parámetros.

Cuadro I | Diferentes tamaños de muestra según distintos valores de nivel de confianza (α), de la prevalencia (p) y la precisión (d)

α (Z_{α})	p	d	n
0.05 (1.960)	0.2	0.08	97
0.05 (1.960)	0.2	0.04	85
0.01 (2.576)	0.2	0.08	66
0.01 (2.576)	0.2	0.04	64
0.05 (1.960)	0.4	0.08	145
0.05 (1.960)	0.4	0.04	476

Para una media

Entendiendo lo anterior resulta sencillo comprender los componentes para estimar el tamaño de muestra de una media. De igual forma, la base es la fórmula del IC de la media:

$$IC \text{ de } 95 \% = \bar{x} \pm Z_{\alpha} \frac{DE}{n}$$

En este caso, la precisión (d) se calcula como sigue:

$$d = Z_{\alpha} \frac{DE}{n}$$

Por lo tanto, la fórmula para el cálculo de tamaño de muestra para estimar una media es:

$$n = \frac{Z_{\alpha}^2 \times DE^2}{d^2}$$

La cual requiere del conocimiento de Z_{α} , DE y la d deseada. De esta forma, el tamaño de muestra para una media esperada depende de Z_{α} (1.96 para $\alpha = 0.05$), de la desviación estándar observada en estudios previos y de la precisión deseada.

Consideraciones finales

Debe quedar claro que los anteriores supuestos no son los únicos para estimar el tamaño de muestra, de tal suerte que si queremos estimar el tamaño de muestra para demostrar diferencias en tasas de incidencia acumuladas (*Hazard risk ratio*), unidades obtenidas en modelos como las curvas de supervivencia de riesgos proporcionales de Cox, la estimación es más compleja por considerar el des-

enlace a través del tiempo, sin embargo, el concepto base es el mismo. Por otro lado, si se desea controlar múltiples variables confusoras o explorar múltiples factores de riesgo a través de un modelo de regresión logística múltiple será necesario utilizar el número de eventos por variable, para el que se requieren entre 10 y 20 sujetos por variable evaluada en el menor de los grupos de desenlace (si la mortalidad es de 30 %, éste es el menor de los grupos, ya que sobrevivirá el restante 70 %).

Bibliografía

1. Cohen J. Statistical power analysis for the behavioural sciences. Second edition. New Jersey: Lawrence Earlbaum; 1988.
2. Dawson B, Trapp R. Basic and clinical biostatistics. Fourth edition. Lange Medical Books-McGraw-Hill; New York, USA; 2004.
3. Hulley SP, Cummings SR. Diseño de la investigación clínica. Un enfoque epidemiológico. Barcelona: Doyma; 1993.
4. Portney LG, Watkins MP. Foundations of clinical research. Applications to practice. Appleton and Lange; Chicago, IL: 1993.
5. Feinstein AR. Principles of medical statistics. London, UK: Chapman and Hall-CRC; 2002.

Para el cálculo del tamaño de muestra

6. Brixton Health. [Sitio web]. Epicalc 2000. Disponible en <http://www.brixtonhealth.com/epicalc.html>
7. EpiInfo 2000. Disponible en <http://huespedes.cica.es/huespedes/epiinfo/espanol.htm>
8. Department of Biostatistics, Vanderbilt University. [Sitio web]. PS: Power and Sample Size Calculation. Disponible en <http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize>