



Metodología

¿Cómo puedo calcular el tamaño de la muestra? Importancia en la calidad y la validez en la investigación en ciencias de la salud

How can I calculate the sample size? Importance of quality and validity in health sciences research

Dra. Martha A. Sánchez Rodríguez¹

¹ Profesora de Carrera Titular C, TC de la carrera de Química Farmacéutico Biológica, FES Zaragoza, UNAM

INTRODUCCIÓN

El tamaño de muestra es uno de los puntos fundamentales en una investigación clínica-epidemiológica para garantizar la validez externa de los estudios. En este sentido, la validez externa se refiere a la representatividad de la muestra o la posibilidad de generalizar los resultados, por lo que es completamente dependiente de cuántos sujetos son incluidos en una investigación y cómo fueron seleccionados.¹ Si el tamaño de muestra es mayor del calculado, será mejor, pero si es menor se corre el riesgo de no encontrar ningún efecto pues la variabilidad puede ser tan alta que impida lograr el objetivo.² Este cálculo es indispensable cuando el diseño del estudio es de tipo analítico, esto es, si se pretende realizar una comparación entre grupos o se busca la asociación entre variables, no importando si la escala de medición de las variables es cuantitativa o cualitativa. Para estudios descriptivos, es necesario para asegurar la presencia de las variables de estudio.

Lo ideal sería realizar una investigación con poblaciones, pero logística y económicamente la mayoría de las veces no es viable, por lo que se utilizan muestras. La elección y número de participantes tiene como propósito asegurar que los hallazgos observados representen con exactitud lo que sucede en la población y esto se logra cuando se lleva a cabo un muestreo probabilístico, es decir, utilizar métodos aleatorios en la selección y cálculos estadísticos para estimar el mínimo de sujetos que se requieren en función de la variabilidad del fenómeno a estudiar.^{2,3}

Aunque la elección y determinación de cuántos sujetos deben participar en una investigación puede hacerse de una forma no probabilística, esta manera de plantear un proyecto tiene la limitación de carecer de validez externa, por lo que depende del alcance que se pretenda en un estudio para establecer que forma de muestreo es más adecuada para llevarlo a cabo. El uso de métodos no probabilísticos no descalifica un estudio si está bien planteado y llevado a cabo, sólo limita el alcance.³

El planteamiento del tipo de muestreo a utilizar y el cálculo del tamaño de muestra para una investigación debe hacerse desde la elaboración del proyecto (plan de la investigación). Si lo que se desea es emplear un muestreo probabilístico es necesario partir de una serie de consideraciones que desde el punto de vista estadístico son supuestos. La elección de los datos que conforma-

Correspondencia: Dra. Martha A. Sánchez Rodríguez

E.mail: masanrod@comunidad.unam.mx

Artículo recibido: 21 de noviembre de 2022

Artículo aceptado: 30 de noviembre de 2022

Sánchez-Rodríguez MA ¿Cómo puedo calcular el tamaño de la muestra? Importancia en la calidad y la validez en la investigación en ciencias de la salud. *CyRS*. 2022; 4(2): 113-126
<https://doi.org/10.22201/fesz.26831422e.2022.4.2.8>



rán los supuestos para llevar a cabo el cálculo es de suma importancia, y si son incorrectos, el cálculo será erróneo comprometiendo la viabilidad del estudio y la posible validez de los resultados.^{4,5} Por ello es de suma importancia conocer los elementos que componen el cálculo del tamaño de muestra, cómo los debemos seleccionar y cómo lo podemos calcular; siendo el objetivo de este trabajo resumir estos elementos para facilitar esta tarea, además de incluirse el uso programas estadísticos como una herramienta de utilidad para realizar el cálculo.

ELEMENTOS NECESARIOS PARA UN CÁLCULO DEL TAMAÑO DE MUESTRA

Los elementos que se deben considerar para estimar el tamaño de la muestra son: (i) escala de medición de la variable dependiente; (ii) variabilidad o magnitud esperada; (iii) diferencia mínima esperada, tamaño del efecto o precisión; (iv) nivel de significancia; (v) poder estadístico; y (vi) planteamiento de la prueba o contraste de hipótesis, uni- o bilateral.^{4,6}

Escala de medición de la variable dependiente (lo que se quiere medir)

La escala de medición implica la manera en que se recolectan los datos para su análisis. Recordemos que una característica de las variables cuantitativas es que expresan la magnitud de la medición numéricamente, siendo más objetivas y reproducibles; en cuanto a las cualitativas, como su nombre lo dice, el valor es una característica o cualidad que permite diferenciarlo en el sujeto/objeto que la posee, pudiendo indicar una magnitud limitada cuando la escala ordinal, la nominal no tiene implicación numérica, por lo que las variables cualitativas son esencialmente subjetivas.⁷ De esta manera, la escala cuantitativa permite discriminar mejor los fenómenos que las cualitativas, por lo que, en el cálculo, el tamaño de muestra será menor para una escala cuantitativa por su mayor posibilidad de discriminación.⁶

La variabilidad o magnitud

La dispersión esperada de los datos justifica que la estimación y selección del tamaño muestral no sea subjetivo, por lo que se requiere de la experiencia obtenida previamente del problema investigado, de ahí que es necesario recurrir a estudios anteriores reportados en la literatura o llevar a cabo un estudio piloto.^{4,5,8} Si esto no es posible, se asigna una máxima probabilidad a la

variable de interés. Cuando la variable está en escala cualitativa el valor propuesto es de 0.5, que corresponde al 50% de variabilidad; en el caso de variables cuantitativas se determina una aproximación al valor de la desviación estándar obteniendo la diferencia entre el mínimo y el máximo valor esperado dividido entre cuatro.^{5,9}

Obtener el dato de la literatura puede ser un desafío importante cuando el fenómeno a estudiar es reportado con variaciones muy amplias. Si este es el caso, algunos tips para una correcta elección son: los datos deben provenir de poblaciones o muestras que se parezcan a la población objetivo; si la variable es cuantitativa conviene tomar los resultados con el mayor coeficiente de variación, el cual se obtiene dividiendo la desviación estándar (DE) entre la media; y si es cualitativa, como fue mencionado la proporción que más se acerque a 0.5.^{4,10}

Cuando el cálculo, es para estudios analíticos, otro punto a considerar será cuántos grupos se van a comparar, ya que se requerirá un tamaño muestral por cada grupo. Por ejemplo, si se pretende llevar a cabo un estudio en donde se quiere comparar el efecto de un hipoglucemiante nuevo contra el líder en el mercado, se requerirá de dos valores de media y DE, uno del mejor resultado con el mejor tratamiento (éxito) y otro de los que no se recuperan con el tratamiento, o que se recuperan con placebo o control (fracaso), en el caso de variables cuantitativas. Si la variable en cuestión es cualitativa, los datos a obtener son de proporciones, siguiendo la misma lógica (Cuadro 1).^{2,11}

Tamaño del efecto

El tamaño del efecto se refiere a la diferencia medida más pequeña entre los grupos comparados que el investigador quiere detectar o la magnitud del efecto bajo la hipótesis alterna; también se conoce como diferencia mínima esperada o precisión y se representa en las fórmulas como d o delta (δ). Prácticamente es la más pequeña diferencia que puede tener significancia clínica o biológica y con ello se asegura el cumplimiento de la hipótesis de trabajo. De acuerdo con los cálculos, cuando la diferencia mínima esperada es pequeña se requerirá de un tamaño de muestra mayor para poder detectar una significancia estadística.^{2,5} La determinación del tamaño del efecto es un punto crítico en el cálculo del tamaño de muestra, dependiendo del valor que se utilice será el resultado pudiendo ser insuficiente para demostrar la hipótesis planteada.¹²

Cuadro 1. Clave para la obtención de la literatura de los valores de los parámetros de la variabilidad esperada y realizar el cálculo de la diferencia mínima esperada. Modificado de Mejía-Aranguré y cols.¹¹

Parámetro	Clave
\bar{X}_1 o p_1	<p>Media o proporción del fracaso. Si se obtiene de un estudio de:</p> <ul style="list-style-type: none"> - Cohorte: media o proporción de individuos no expuestos que desarrollan la enfermedad. - Casos y Controles: media o proporción de controles que estuvieron expuestos. - Ensayo clínico: son los individuos con el peor tratamiento o placebo que se recuperan, o con el mejor tratamiento que no se recuperan.
\bar{X}_2 o p_2	<p>Media o proporción del éxito. Si es un estudio de:</p> <ul style="list-style-type: none"> - Cohorte: media o proporción de individuos expuestos que desarrollan la enfermedad. - Casos y Controles: media o proporción de casos que están expuestos. - Ensayo clínico: media o proporción de individuos con el mejor tratamiento que se recuperan.

De acuerdo con el diseño de estudio y la escala de medición de la variable dependiente, la diferencia mínima esperada se obtiene de medias en el caso de variables cuantitativas (diferencia de medias) o proporciones cuando son cualitativas (diferencia de proporciones), aunque también puede ser un parámetro de asociación como la correlación, regresión o algún tipo de riesgo, cuando los diseños son de tipo analítico.^{5,10,12} Estos datos pueden ser obtenidos de la literatura, de tal manera que las diferencias se calculan a partir de los resultados de la variabilidad esperada descrita anteriormente.

En el caso de estudios descriptivos, la diferencia mínima esperada refleja la diferencia entre los límites inferior y superior de un intervalo de confianza (IC) esperado, generalmente al 95%, por lo que sólo se necesita de un valor de proporción como tamaño del efecto.⁹ Si la variable es de escala cuantitativa se requiere de un valor de varianza poblacional o muestral, dependiendo de los datos con los que se cuente.^{4,5,10} También es posible

establecer una precisión *a priori*, propuesta por el investigador en función a la experiencia, aunque el dato no debe ser mayor de 10% para variables cualitativas; para variables cuantitativas es muy riesgoso establecer un valor de precisión *a priori*, por lo cual no se recomienda.

Nivel de significancia

Se refiere a la certeza con la que se quiere demostrar que la muestra proviene de la población, también si la diferencia o asociación entre los grupos a comparar es real; el significado depende del objetivo del estudio. Representa el grado de confianza de que el valor verdadero del parámetro en la población se encuentre en la muestra calculada.^{2,5} Este nivel de significancia se conoce como α corresponde al error tipo I y su fundamento está referido en la prueba de hipótesis estadística.^{2,6,10,13}

En este artículo no ahondaremos en el tema, solo tomaremos los conceptos, si requiere repasarlo se sugiere la



lectura del artículo que previamente fue presentado en esta revista.¹³ De ahí que el error tipo I es la probabilidad de detectar un efecto cuando realmente no existe, por lo que se puede definir como un falso positivo.^{2,12,13}

El valor de α se obtiene de la distribución normal, y como el parámetro que describe esta distribución es z , éste debe escribirse como z_{α} . El criterio más utilizado es una certeza o confianza del 95% que corresponde a un valor de $\alpha = 0.05$, pero puede cambiarse a 0.1 o 0.01, lo que significa una confianza del 90% o 99%, respectivamente, dependiendo del nivel de confianza con el que se quiera hacer el cálculo.^{4,5,14} Un criterio del 95% indica que hay una posibilidad de error en el cálculo del 5%. El error tipo I o α es más grave en el cálculo del tamaño de muestra, por ello la posibilidad de cometerlo debe ser baja y el valor ideal es 5% o menos; además, mientras más pequeño es el nivel el tamaño de muestra será más grande.⁶ Un nivel de significancia superior al 5% no es conveniente porque se incrementa la posibilidad de cometer el error tipo I.

Poder o potencia de la muestra

Se define como la probabilidad de que en un estudio con un determinado tamaño de muestra y nivel de significancia se encuentre un resultado estadísticamente significativo, es decir, la probabilidad de rechazar correctamente la hipótesis nula (H_0) cuando el valor de la medida en la población es el tamaño del efecto (d)¹¹ y se define como $1 - \beta$.^{10,12,14} Beta (β) es la probabilidad de cometer el error tipo II y se refiere a la posibilidad de no rechazar H_0 a pesar de que exista una diferencia o asociación real, es decir, no detectar un efecto cuando realmente lo hay o falso negativo.^{2,6,12} Como este error es dependiente de la muestra se conoce como el error de muestreo. El poder sólo aplica a estudios comparativos.

El valor de β también se obtiene de la distribución normal, por lo que se representa como z_{β} y el poder como $z_{1-\beta}$; y a diferencia de α , se aceptan valores que oscilan entre 5% y 20% que representa un poder del 80% al 95%,¹⁴ aunque frecuentemente se utiliza una potencia del 90% para la mayoría de los estudios y 80% para los ensayos clínicos.^{2,4} Un aspecto importante a considerar es que mientras más grande es la potencia utilizada en el cálculo, mayor será el tamaño muestral requerido; además, cuando el poder es $< 80\%$ se corre el riesgo de cometer error tipo II.

Contraste de hipótesis

Hay dos posibilidades: i) unilateral o de una cola, o ii) bilateral o de dos colas. En el estadígrafo z , al igual que los demás estadígrafos, la distribución se divide en un área de aceptación y otra de rechazo de H_0 ; por lo que, si se hace una prueba de hipótesis de dos colas o bilateral hay dos áreas de rechazo (una de cada lado de la distribución), de tal manera que no importa la dirección del rechazo. Cuando se hace un planteamiento bilateral el estadígrafo de z debe representarse como $z_{\alpha/2}$. En el caso de un planteamiento unilateral, o de una cola, el área de rechazo está en un solo lado, izquierdo (inferior) o derecho (superior), representándose como z_{α} . Los valores críticos de z que dividen la distribución serán diferentes dependiendo del tipo de planteamiento y el nivel de confianza seleccionado,^{2,6} los más utilizados para ambos planteamientos se encuentran en el cuadro 2.

¿De qué depende la elección del contraste de hipótesis? De lo que se espera en la investigación. Se puede suponer que la variable a medir tendrá una respuesta en una dirección; por ejemplo, retomando la propuesta de comparación entre dos hipoglucemiantes, si esperamos que el fármaco nuevo será mejor que el tradicional y tenemos evidencia de ello, es un planteamiento unilateral (hipoglucemiante nuevo $>$ hipoglucemiante tradicional). En el caso que no haya seguridad o se espera que el comportamiento entre ambos fármacos sea semejante, el planteamiento debe ser bilateral (hipoglucemiante nuevo \neq hipoglucemiante tradicional).⁴

El planteamiento unilateral llevará a un cálculo de tamaño muestral menor, pero es riesgoso puesto que, si no se cumple en la investigación la hipótesis propuesta, en nuestro ejemplo que el fármaco nuevo es mejor, el tamaño de muestra no será suficiente para demostrar ningún otro efecto, disminuyendo el poder; por lo que no es frecuente un cálculo de tamaño muestral unilateral, sino bilateral, que asegura un mayor número de elementos a estudiar.^{4,6,10}

El planteamiento para el contraste de hipótesis aplica para z_{α} , en el caso de z_{β} siempre se seleccionan los valores críticos unilaterales, en ambos casos el valor del estadígrafo cambia según el nivel de significancia elegido (Cuadro 2). Si se desea un nivel de confianza diferente al de este cuadro deberá consultarse una tabla de z que podrá encontrar en los diferentes libros de estadística.

Cuadro 2. Valores críticos de z, uni- y bilateral, para el nivel de confianza y poder más utilizados para el cálculo del tamaño de muestra. Modificado de Zodpey SP.¹⁵

α	$z_{\alpha/2}^*$	α o β	z_{α} o z_{β}^{\dagger}
0.10	1.65	0.20	0.84
0.05	1.96	0.15	1.04
0.01	2.58	0.10	1.28
		0.05	1.65

*Valor de z bilateral; †valor de z unilateral.

¿CÓMO SE REALIZA EL CÁLCULO?

Una vez que ya elegimos todos los elementos señalados se procede a seleccionar la fórmula adecuada para realizar el cálculo del tamaño de la muestra. Hay muchas fórmulas propuestas para hacerlo, así que se debe utilizar la que corresponda de acuerdo con el objetivo y diseño del estudio, la prueba estadística que sirve de base para obtener los datos, el número de grupos que se compararán (si es el caso) y la razón de asignación o desigualdad entre los grupos (1:1, 1:2, 1:3).

Cuando la investigación es epidemiológica de tipo descriptivo o un estudio exploratorio en donde principalmente se realizará el cálculo de prevalencias, el tamaño muestral se obtiene con fórmulas para una sola población, y por ello no es necesario un dato de poder o β .^{4,5} Para los estudios de tipo analítico, ya sea comparativos o correlacionales, observacionales o experimentales, es necesario contar con todos los elementos señalados

anteriormente, y la diferencia en la forma de calcular es el parámetro estadístico por utilizar. Por ejemplo, si se compararán dos medias, se debe utilizar el estadígrafo de t , y puede hacerse para dos grupos independientes o para grupos dependientes, según sea el caso. Si la comparación es de proporciones, el cálculo se hace a través de la prueba de χ^2 con las diferentes formas de obtenerla, lo cual es dependiente de lo que se pretende comparar. Cuando es una relación entre variables cuantitativas el parámetro a utilizar es el coeficiente de correlación (r); para comparar tres o más medias, el cálculo es a través del análisis de la varianza (ANOVA); por mencionar algunas.^{16,17} Diseños de cohorte para análisis de supervivencia o con múltiples variables a relacionar hacen que la estimación sea más compleja lo mismo que las fórmulas a utilizar.

Existen libros dedicados exclusivamente a este tema, lo que rebasa el objetivo de este trabajo, por lo que

Cuadro 3. Fórmulas comunes para el cálculo del tamaño muestral de estudios descriptivos y comparativos con variables cuantitativas y cualitativas en un contraste bilateral.^{2,5,18,19}

Estudios descriptivos	Estudios comparativos
Proporciones con población conocida (finita) $n = \frac{Nz_{\alpha/2}^2pq}{[d^2(N-1)] + [pqz_{\alpha/2}^2]}$	Proporciones para dos grupos en razón 1:1 $n = \frac{2(z_{\alpha/2} + z_{\beta})^2 \bar{p}\bar{q}}{(p_2 - p_1)^2} \quad \bar{p} = \frac{p_1 + p_2}{2}$
Proporciones con población desconocida (infinita) $n = \frac{z_{\alpha/2}^2 pq}{d^2}$	Proporciones para dos grupos en razón 1:j $n = \frac{(j+1)(z_{\alpha/2} + z_{\beta})^2 \bar{p}\bar{q}}{d^2 j} \quad \bar{p} = \frac{j p_1 + p_2}{1+j}$
Variable cuantitativa con población conocida (finita) $n = \frac{Nz_{\alpha/2}^2 s^2}{[d^2(N-1)] + [z_{\alpha/2}^2 s^2]}$	Diferencia de medias para dos grupos razón 1:1* $n = \frac{2(z_{\alpha/2} + z_{\beta})^2 s^2}{(\bar{X}_1 - \bar{X}_2)^2}$
Variable cuantitativa con población desconocida (infinita) $n = \frac{z_{\alpha/2}^2 s^2}{d^2}$	Diferencia de medias para tres grupos razón 1:1:1* $n = \frac{\lambda}{\Delta} \quad \Delta = \frac{\Sigma(\bar{X}_i - \bar{X}_{total})^2}{s^2}$
Coeficiente de correlación $n = \left[\frac{z_{\alpha/2} + z_{\beta}}{0.5 \ln[(1+r)/(1-r)]} \right]^2 + 3$	

En donde: n: tamaño muestral; N: tamaño poblacional; p: proporción del efecto; q: 1 - p; $z_{\alpha/2}$: valor crítico de z_{α} bilateral; d: diferencia mínima esperada; s^2 : varianza; z_{β} : valor crítico de z_{β} unilateral; \bar{p} : tamaño del efecto estimado o ponderado entre las dos proporciones esperadas; \bar{q} : 1 - \bar{p} ; p_1 : proporción de fracaso; p_2 : proporción de éxito; j: razón de desigualdad o asignación de grupos; \bar{X}_1 : media del grupo 1; \bar{X}_2 : media del grupo 2; λ : constante = 9.64 con $\alpha = 0.05$ y $\beta = 0.20$; Δ : media ponderada; \bar{X}_i : media i (i = 1, 2, 3); \bar{X}_{total} : promedio de las medias 1, 2, 3; ln: logaritmo natural; r: coeficiente de correlación de Pearson; *se asume que las varianzas son iguales.

solo se presentan en el cuadro 3 las fórmulas más comunes para estudios descriptivos y comparativos de variables cualitativas y cuantitativas en la investigación clínica. El cómo utilizar las fórmulas lo puede consultar en el capítulo XV del libro Metodología de la investigación. Un enfoque práctico.²

Otra opción es el uso de programas computacionales y aplicaciones para telefonía móvil o tabletas. Estos programas utilizan diferentes algoritmos, pero el fundamento es el mismo y la mayoría de las veces será necesario contar con 4 o 5 de los elementos que se han tratado en este trabajo para poder efectuar el cálculo (Figura 1).

La escala de medición de la variable dependiente no es un cuestionamiento para el cálculo, pero es un

concepto necesario para saber qué tipo de cálculo se realizará.

Como los algoritmos son diferentes, es posible que haya una discrepancia entre el resultado cuando se hace el cálculo utilizando una fórmula o con cualquiera de estos programas o aplicaciones, aunque generalmente son pequeñas,¹⁰ a menos que el algoritmo esté construido con una finalidad específica más allá del tamaño muestral, como obtener el tamaño del efecto y mantener el poder de la muestra,¹⁶ o se incluyan ajustes de continuidad, en cuyo caso la diferencia en el cálculo puede ser más grande.

En internet podemos encontrar programas de acceso libre para sólo el cálculo del tamaño muestral tales como:



Home

Calculators

CI for proportion

CI for mean

Means - effect size

Means - sample size

Correlation - sample size

Proportions - sample size

CI for proportion - sample size

Survival analysis - sample size

Prevalence

CI for risk ratio

More calculators...

Calculator finder

About calculating sample size

About us

Sample size - Means

Compare the mean of a continuous measurement in two samples. The sample sizes are calculated in two different ways: first using the T statistic (with a non-centrality parameter), then using the Z statistic. The Z statistic approximates the T statistic, but provides sample sizes that are slightly too small. (We provide the Z statistic calculation to allow comparison with other calculators which use the Z approximation.)

Instructions: Enter parameters in the green cells. Answers will appear in the blue box below.

α (two-tailed) = 0.05 ¹ Threshold probability for rejecting the null hypothesis. Type I error rate.

β = 0.2 ² Probability of failing to reject the null hypothesis under the alternative hypothesis. Type II error rate.

q_1 = 0.5 Proportion of subjects that are in Group 1 (exposed)

q_0 = 0.5 Proportion of subjects that are in Group 0 (unexposed); 1- q_1 .

E = 0.5 ³ Effect size (if μ_1 = mean in Group 1 and μ_0 = mean in Group 0, then $E = \mu_1 - \mu_0$.)

S = 1 ⁴ Standard deviation of the outcome in the population

Calculate

Este programa hace el cálculo bilateral

1. Nivel de significancia
2. Poder (potencia) o β
3. Diferencia entre los grupos o tamaño del efecto
4. Variabilidad o magnitud de la dispersión.
5. Contraste de hipótesis
 - a. Razón de asignación de grupos o desigualdad.

Figura 1. Elementos que solicita el módulo del cálculo del tamaño de muestra del programa estadístico STATA® y la aplicación Sample Size Calculators para un cálculo a partir de las medias de dos grupos.

**G*Power**

(<https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower>)

Sample size calculator

(<https://www.sample-size.net/>)

EPIDAT

(<https://www.sergas.es/Saude-publica/EPIDAT-4-2?idioma=es>)

Epi Info

(https://www.cdc.gov/epiinfo/support/esp/es_downloads.html)

También varios programas estadísticos comerciales contienen un módulo semejante como STATA®, SAS®, STATISTICA® y Sigma Plot®10, por mencionar algunos.

Los elementos comunes en los cálculos, independientemente del software o si el cálculo es manual son: nivel de significancia, poder o valor de β , tamaño y variabilidad del efecto. Algunos solicitan seleccionar el criterio del contraste de hipótesis y la razón de asignación de grupos, si son iguales o desiguales. Si el software no solicita el contraste de hipótesis, el cálculo es bilateral.

Un último punto que se debe tomar en cuenta son las posibles pérdidas o abandonos en una investigación longitudinal (ensayos clínicos o cohortes), debido a que, como se indicó en un inicio, el tamaño de muestra calculado es el mínimo para encontrar una diferencia o asociación a partir del tamaño del efecto supuesto, y en los estudios de seguimiento es frecuente que quienes participan abandonen el estudio por diferentes razones, provocando una disminución en la potencia. Para evitar esta pérdida en el poder del estudio se sugiere aumentar entre el 5 y el 20% el tamaño de muestra calculado, recomendándose específicamente en los ensayos clínicos aumentar el 20% debido a la naturaleza de la investigación.^{6,18}

Uno de los programas más completos para el cálculo del tamaño muestral es el *G*Power*, por la sencillez de su uso y la variedad de pruebas estadísticas con las que se puede hacer el cálculo, lo cual es aplica-

ble a investigaciones en los campos del conocimiento social, biomédico o del comportamiento. El programa fue desarrollado por investigadores de la Universidad de Düsseldorf e incluye el ajuste del poder, una calculadora para mejorar el tamaño del efecto y con ello el tamaño muestral, además de representaciones gráficas.^{12,16,17} Otros programas son un poco más complejos para comprender qué datos son los que se deben colocar para hacer el cálculo, un punto crucial para tener un resultado confiable, por lo que, no es sólo que el software realice el cálculo, sino saber cuáles son los datos correctos para alimentarlo.

EJEMPLO DE CÁLCULO

Para observar la aplicación de lo que hemos revisado, realizaremos el cálculo de un ejemplo en forma manual, con un programa comercial (STATA®) y uno libre (*G*Power*) para ver su uso y utilidad, además de señalar la diferencia.

Ejemplo: Se desea calcular el tamaño muestral para llevar a cabo un ensayo clínico con la finalidad de determinar el efecto de la terapia estrogénica sobre los niveles de colesterol sanguíneo en mujeres posmenopáusicas con síndrome metabólico. Encontramos un artículo que reporta el efecto sobre los niveles de triglicéridos y se toman estos datos por analogía. En este estudio se reporta que por efecto de esta terapia la proporción de mujeres con síndrome metabólico e hipertrigliceridemia disminuyó 26% y de las que estuvieron con placebo el 9% se mantuvo con triglicéridos normales después de 6 meses.²⁰ Se requiere un cálculo al 95% de confianza bilateral y poder del 80%, con una asignación igual a cada grupo (1:1).

Comenzaremos con el cálculo manual. Lo primero que hay que hacer es identificar a qué proporción corresponde cada porcentaje. Revisando el cuadro 1, p_1 en un ensayo clínico es la proporción de los sujetos en placebo que se recuperan, que en nuestro caso es el 9%; y p_2 es la proporción de sujetos con el mejor tratamiento que se recuperan, es decir 26%. Posteriormente, hay que pasar estos porcentajes a proporciones dividiendo entre 100, por lo que $p_1 = 0.09$ y $p_2 = 0.26$. A continuación debemos obtener los valores de z que correspondan buscándolos en el cuadro 2. De acuerdo con ello, para el nivel de confianza al 95% $Z_{\alpha/2} = 1.96$, y un poder del 80% indica $1 - \beta = 0.20$ por lo que $Z_{\beta} = 0.84$.

Finalmente, seleccionar la fórmula que corresponde en el cuadro 3. En este sentido, vamos a comparar dos proporciones en razón 1:1, por lo que las fórmulas son:

$$n = \frac{2(Z_{\alpha/2} + Z_{\beta})^2 pq}{d^2} \quad \bar{p} = \frac{p_1 + p_2}{2}$$

Calcular las proporciones estimadas y la diferencia entre ellas:

$$\bar{p} = \frac{0.09 + 0.26}{2} = 0.175 \quad \bar{q} = 1 - 0.175 = 0.825 \quad d = 0.26 - 0.09 = 0.17$$

Sustituir en la fórmula:

$$n = \frac{2(1.96 + 0.84)^2 (0.175)(0.825)}{(0.17)^2} = \frac{2.2638}{0.0289} = 78.33 \sim 78$$

Se requieren 78 mujeres en el grupo tratamiento y 78 en el placebo para encontrar una diferencia del 17% entre ellos, con una confianza del 95% y poder del 80%.

Si realizamos el mismo cálculo con el programa STATA® se procede como se muestra en la figura 2. Debe seleccionarse el submenú "Potencia y tamaño de muestra" y de ahí la opción "Test chi-cuadrado para comparar dos proporciones independientes", se colocan los datos de α , poder, p_1 , p_2 y se deja la indicación de "Reportar tamaño del efecto como Diferencia". Los datos que se deben colocar son proporciones por lo que $\alpha = 0.05$ y poder = 0.80; y p_1 y p_2 son los resultados de la transformación del porcentaje obtenida de manera manual. Es importante señalar que el cuadro 1 es una herramienta para identificar en los artículos las diferentes proporciones o medias, no importa la forma en que se realizará el cálculo, y que hay que transformar el porcentaje a proporción. En los resultados, observamos que el cálculo es el mismo que por el método manual y reporta la diferencia (Figura 3).

Ahora haremos el cálculo con el programa G*Power. Como primer paso, hay que seleccionar la "familia de pruebas" estadísticas, en este caso, al ser una comparación de proporciones debe utilizarse la χ^2 y este programa utiliza el cálculo con la prueba "exacta" de Fisher. Después hay que seleccionar la prueba estadística: "Proporciones, desigualdad, dos grupos independientes", y finalmente dejar la opción "A priori, dado α , poder y tamaño del efecto" (Figura 4).

Una vez seleccionadas las condiciones del análisis, se procede a colocar los datos de p_1 y p_2 , α y poder como se indicó en el ejemplo anterior, dejando la "asignación de grupos" ya que por default está 1. Debe hacerse la selección del contraste de hipótesis ("Tail(s)") a "Two" para que sea bilateral porque por default indica unilateral ("One"). En este programa no es necesario incluir la diferencia. Como resultado nos indica 8 mujeres más (86) que en los cálculos anteriores (Figura 4), esto es porque hace un ajuste para mejorar tanto el nivel de significancia como el poder. En este caso mejoró el nivel de significancia ($\alpha = 0.033$) aumentando el tamaño de muestra.

Frecuentemente, el programa G*Power nos proporciona un resultado mayor al obtenido de forma manual o con otros softwares por el ajuste que hace del poder y la significancia. Esto no es malo, al contrario, puede ser un ajuste beneficioso porque asegura la diferencia o asociación que se pretende buscar con una muestra un poco mayor que la calculada sin ajuste.

Importante no olvidar que, como el cálculo es para un ensayo clínico, deberá aumentarse el 20% a la muestra. Para el caso de 78 sería:

$$78 \times 0.20 = 15.6 \sim 16 \quad 78 + 16 = 94$$

Así, se requieren 94 mujeres con síndrome metabólico por grupo para llevar a cabo la investigación. Para el cálculo de 86:

$$86 \times 0.20 = 17.2 \sim 17 \quad 86 + 17 = 103$$


```

STATA (R)
-----
Statistics/Data Analysis 14.2

Special Edition

. power twoproportions 0.09 0.26, test(chi2)

Performing iteration ...

Estimated sample sizes for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1 versus Ha: p2 != p1

Study parameters:

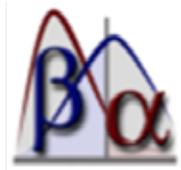
      alpha =    0.0500
      power =    0.8000
      delta =    0.1700 (difference)
      p1 =     0.0900
      p2 =     0.2600

Estimated sample sizes:

      N =      156
      N per group =    78
    
```

Se reportan los datos que colocamos y nos muestra el cálculo de la diferencia (recuadro azul). Al final encontramos el total de sujetos que requerimos y cuántos son por grupo (recuadro rojo).

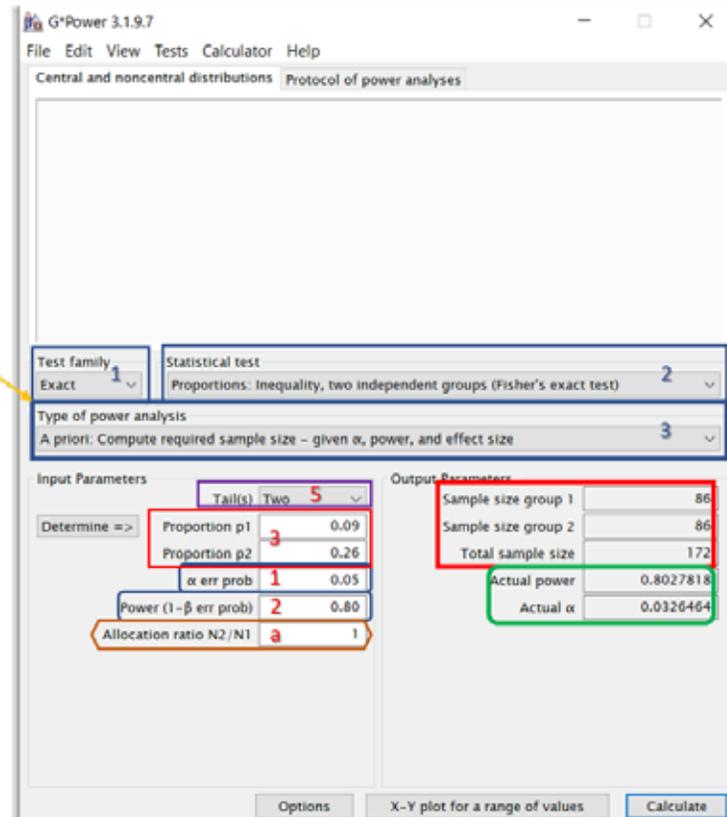
Figura 3. Resultados del cálculo del tamaño de muestra con el programa STATA®.



Recuadros azules

1. Seleccionar la familia de pruebas
2. Seleccionar la prueba estadística
3. Seleccionar el tipo de análisis, en general A priori.

1. Nivel de significancia
2. Poder o potencia
3. Diferencia entre los grupos o tamaño del efecto
5. Contraste de hipótesis
 - a. Razón de asignación de grupos o desigualdad.



Resultados:

Se requieren 86 sujetos por grupo. Se hace un ajuste en el nivel de significancia, por eso el tamaño muestral es más grande, porque el nivel de significancia es menor.

Figura 4. Cálculo del tamaño de muestra del ejemplo con el programa G*Power. Lado izquierdo recuadros azules, procedimiento de selección de la forma de análisis y pruebas. Lado izquierdo inferior, datos a colocar, no se requiere de la diferencia. Lado derecho recuadro rojo, resultados; recuadro verde, resultados del ajuste de α y poder.

Si se prefiere considerar el cálculo de *G*Power* se requieren 103 mujeres por grupo, con la ventaja de que aseguraría la posible diferencia al aumentar el nivel de significancia manteniendo la potencia.

Finalmente, este ha sido un breve recorrido de cómo se realiza el cálculo del tamaño de muestra para una investigación clínica o epidemiológica, revisando los elementos necesarios para realizarlo y diferentes maneras de hacerlo. Para un uso eficiente del programa *G*Power* se puede revisar el manual de usuario que está disponible en la página web donde se obtiene el programa y para profundizar en el tema, los capítulos del libro previamente recomendado.^{2,3,17}

AGRADECIMIENTOS

El artículo es parte del material de apoyo didáctico elaborado por la Red Académica Asesora de Revisiones Sistemáticas de la FES Zaragoza, UNAM (RAARS), financiada por la Dirección General de Asuntos del Personal Académico, UNAM, proyecto PAPIME PE203421. Programa para la Investigación Bibliográfica Científica sobre Salud de la FES Zaragoza, UNAM (PIBCIS).

REFERENCIAS

1. Mendoza-Núñez VM, Romo-Pinales MR, Sánchez-Rodríguez MA, Hernández Zavala MS. Investigación. Introducción a la metodología. México: Facultad de Estudios Superiores Zaragoza, UNAM; 2004. p. 84-94.
2. Sánchez-Rodríguez MA. Cuántos habrán de participar. Tamaño de la muestra. En: Metodología de la investigación. Un enfoque práctico. [ebook] México: FES Zaragoza; 2018. Disponible en: <https://www.zaragoza.unam.mx/area-quimico-biologicas/>
3. Sánchez-Rodríguez MA. Cómo seleccionar a los participantes. Muestreo. En: Sánchez-Rodríguez MA, Rosas-Barrientos JV, Mendoza-Núñez VM. Metodología de la investigación. Un enfoque práctico. [ebook] México: FES Zaragoza; 2018.
4. Eng J. Sample size estimation: how many individuals should be studied? *Radiology*. 2003; 227:309-13. doi: 10.1148/radiol.2272012051.
5. Aguilar-Barojas S. Fórmulas para el cálculo de la muestra en investigaciones de salud. *Salud Tabasco*. 2005; 11(1-2):333-8.
6. Rendón-Macías ME, Villasís-Keever MÁ. El protocolo de investigación V: el cálculo del tamaño de muestra. *Rev Alergia Mex*. 2017; 64(2):220-7.
7. Zacarías-Flores M, Sánchez-Rodríguez MA. Que tanto medir y como medir o lo que es lo mismo: las variables. En: Sánchez-Rodríguez MA, Rosas-Barrientos JV, Mendoza-Núñez VM. Metodología de la investigación. Un enfoque práctico. [ebook] México: FES Zaragoza; 2018.
8. Naduvilath TJ, John RK, Dandona L. Sample size for ophthalmology studies. *Ind J Ophthalmol*. 2000; 48:245-50.
9. Cortés-Cortés M, Mur-Villar N, Iglesias-León M, Cortés-Iglesias M. Algunas consideraciones para el cálculo del tamaño muestral en investigaciones de las Ciencias Médicas. *Medisur [revista en Internet]*. 2020; 18(5): [5 p.]. Disponible en: <http://medisur.sld.cu/index.php/medisur/article/view/4558>
10. García-García JA, Reding-Bernal A, López-Alvarenga JC. Cálculo del tamaño de la muestra en investigación en educación médica. *Inv Ed Med*. 2013; 2(8):217-24.
11. Mejía-Aranguré JM, Fajardo-Gutiérrez A, Gómez-Delgado A, Cuevas-Urióstegui ML, Hernández-Hernández DM, Garduño-Espinosa J, et al. El tamaño de muestra: un enfoque práctico en la investigación clínica pediátrica. *Bol Med Hosp Infant Mex*. 1995; 52:381-91.
12. Leongómez JD. Análisis del poder estadístico y cálculo del tamaño de muestra en R: guía práctica (Versión 3). *Inv Abierta*; 2020. Zenodo. Disponible en: <https://doi.org/10.5281/zenodo.4019906>
13. Sánchez-Rodríguez MA. La significancia estadística y los intervalos de confianza ¿qué me indican y cómo puedo interpretarlos? *CyRS*. 2021; 3(1):74-82. doi: 10.22201/fesz.26831422e.2021.3.1.8.
14. Talavera JO, Rivas-Ruiz R, Bernal-Rosas LP. Investigación clínica V. Tamaño de muestra. *Rev Med Inst Mex Seguro Soc* 2011; 49 (5):517-22.
15. Zodpey SP. Sample size and power analysis in medical research. *Indian J Dermatol Venereol Leprol*. 2004; 70: 123-8.

16. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*; 2007; 39: 175-91. doi: 10.3758/bf03193146.

17. Buchner A, Erdfelder E, Faul F, Lang A-G. G*Power 3.1 manual. 2021. Available from: <https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower>.

18. Browner WS, Newman TB, Hulley SB. Cálculo de la potencia y el tamaño de muestra: aplicaciones y ejemplos. En: Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. *Diseño de investigaciones clí-*

nicas. 4a ed. Barcelona: Lippincott Williams & Wilkins; 2014. p. 59-87.

19. Wang H, Chow S. Sample size calculation for comparing means. *Wiley Encyclopedia of Clinical Trials*. Wiley Online Library; 2007. p. 1-10. Disponible en: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780471462422.eoct005>

20. Sánchez-Rodríguez MA, Zacarías-Flores M, Castrejón-Delgado L, Ruiz-Rodríguez AK, Mendoza-Núñez VM. Effects of hormone therapy on oxidative stress in postmenopausal women with metabolic syndrome. *Int J Mol Sci*. 2016; 17: 1388. doi:10.3390/ijms17091388.