

Hablemos de...

Cálculo de la muestra. ¿Cómo y por qué?

CRISTINA FERNÁNDEZ

Unidad de Epidemiología Clínica. Servicio de Medicina Preventiva.
Hospital Clínico San Carlos. Madrid. España.



Áxel Oliveres

Puntos clave

- El tamaño de la muestra se calcula para encontrar significativa ($p < 0,05$) la diferencia planteada en la hipótesis principal del estudio.
- El tamaño de la muestra tiene implicaciones éticas, económicas y estadísticas.
- La precisión (ausencia de error aleatorio) de un estudio, y no su validez (ausencia de error de diseño o de ejecución), viene determinada por el tamaño de la muestra.
- El tamaño de la muestra se basa en la diferencia planteada por el investigador, en la hipótesis principal del estudio. La diferencia depende del tipo de variable: si es cuantitativa, habitualmente, es la diferencia de medias, y si es cualitativa, en la diferencia de proporciones.
- El número de sujetos necesarios para la comprobación de una hipótesis es menor si ésta se basa en una medida cuantitativa (media) con una variabilidad razonable (varianza).
- En estudios observacionales es imprescindible construir modelos multivariantes, lo que exige un mayor tamaño muestral.

Hacer investigación puede llegar a ser fascinante, pero siempre debe ser un proceso sistemático. La planificación de un estudio es una tarea compleja que debe tener en cuenta multitud de aspectos. La forma más sencilla de abordarla es escribir el protocolo siguiendo un esquema estándar. Esto obliga a reflexionar sobre los puntos clave a tratar y facilita que se consideren y se discutan muchos matices que fácilmente podrían olvidarse si no se sigue un sistema predefinido¹.

Uno de los principales interrogantes que se plantea el investigador es ¿cuántos pacientes necesito para el estudio? Varios errores estadísticos se han introducido en la forma de realizar la investigación clínica. Todos

Cuando el investigador se enfrenta a una pregunta científica (hipótesis) está planteando una comparación de parámetros (variables evaluadas) entre los grupos de estudio.

La incertidumbre se genera a partir de la imposibilidad de acceder a toda la población objetivo. Para evaluar la incertidumbre se necesita realizar un contraste de hipótesis estadísticas.

hemos aprendido escuchando frases como “coge 30 que te llegran” o “es mi muestra estadísticamente significativa”. Afirmaciones como éstas representan las falacias existentes hoy día a la hora de utilizar las herramientas estadísticas.

La estadística puede clasificarse en 2 grandes tipos: descriptiva e inferencial.

La primera de ellas se encarga de recoger, organizar, clasificar, resumir y presentar los datos, cumpliendo una función informativa y descriptiva de la que no pueden extraerse conclusiones.

El segundo tipo es la estadística inferencial, que permite extraer conclusiones de las muestras que son generalizables, es decir, que pueden hacerse extensibles al total de la población. La estadística inferencial se basa en el cálculo de probabilidades y efectúa estimaciones, decisiones, predicciones y otras generalizaciones. Se basa fundamentalmente en la estimación de parámetros y en el contraste de hipótesis.

La estimación de parámetros consiste en estimar el valor de una determinada variable en la población

(parámetro) utilizando un menor número de observaciones, es decir, una muestra representativa de esa población² (p. ej., valorar si la media de edad de los pacientes con úlcera duodenal, visitados en la consulta de un centro se corresponde con la media de edad de todos los pacientes de la Comunidad de Madrid). En este caso, el

parámetro de estudio es la media de edad.

La estadística es una herramienta de apoyo fundamental de la investigación sanitaria, necesaria para poder manejar la incertidumbre (ignorancia de los procesos biológicos) de los datos clínicos y para tratar la ingente cantidad de información existente, información, por otra parte, cada vez más cuantificada. Sin embargo, los investigadores médicos recurren a la estadística de forma utilitarista, convirtiéndola en un instrumento y utilizándola de forma automática e incluso irreflexiva. Muchos investigadores la consideran el paso final del estudio, es decir, un simple proceso matemático que permite alcanzar el objetivo, cometiendo un craso error al eludir el razonamiento que reside en el fondo de un análisis estadístico^{1,3,4}.

Los conceptos estadísticos forman parte del método científico y deben planearse desde el comienzo del estudio y mantenerse hasta el final. Su adecuada utilización permitirá evaluar y cuantificar la variabilidad debida al azar, pudiendo incluso reducirla, con objeto de tener la máxima garantía de alcanzar el objetivo deseado. Pero hay una condición *sine qua non* para que la estadística sea de utilidad: un buen análisis estadístico comienza con un buen diseño del estudio y una correcta ejecución del mismo, partiendo inicialmente de una pregunta de investigación coherente y adecuada. Un estudio bien diseñado y pobremente analizado puede solucionarse a través de un segundo análisis estadístico correcto, mientras que un estudio mal diseñado está destinado al fracaso, aunque el análisis sea estadísticamente complejo.



Contraste de hipótesis⁴⁻¹⁰

Para poder plantear el tamaño de la muestra, el investigador debe conocer las bases estadísticas que se emplean.

Se plantea una hipótesis nula (H_0) y una alternativa (H_A). La primera apoya la no diferencia ($A = B$) (las medias o proporciones de las variables estudiadas) y la segunda la diferencia que se pretende detectar ($A \neq B$). Por lo tanto, un paso esencial en toda investigación es la formulación de una hipótesis que se intenta rechazar (hipótesis nula) mediante el uso de una prueba de significación estadística, y la formulación de otra hipótesis que se acepta si se rechaza la hipótesis nula (hipótesis alternativa). Esta prueba de hipótesis mostrará la fuerza de la evidencia en contra de la hipótesis nula en términos de probabilidad. En todo contraste de hipótesis se obtiene un valor p de significación estadística².

Al realizar un contraste de hipótesis pueden acontecer 4 situaciones diferentes. Pueden cometerse 2 errores: afirmar que hay diferencia cuando realmente no existe (p. ej., concluir que la obesidad se asocia con la calidad de vida del paciente con úlcera duodenal, cuando no existe ninguna asociación entre ambas variables) y se puede cometer el error de concluir que no hay diferencia cuando ésta existe (p. ej., concluir que la prevalencia de úlcera duodenal no se incrementa a medida que aumenta la edad, cuando realmente sí existe esa relación). En el primer caso se está cometiendo un error tipo I y su probabilidad es α . En el segundo caso el error es el denominado error tipo II y la probabilidad de cometerlo es β . Se asume por convenio una probabilidad aceptable de error tipo I de 0,05 (5%), es decir, α es el valor de p , que nos informa de la probabilidad de equivocarnos al decir que el tratamiento A tiene efecto sobre la curación de la úlcera duodenal, cuando eso no es cierto (falso positivo); por lo tanto, cuanto más pequeño sea el valor de α más probable es que la diferencia encontrada coincida con la verdad (resultado verdaderamente positivo). La p , sin embargo, no nos informa acerca de la probabilidad de cometer un error tipo II, es decir, de la probabilidad de decir que el consumo de alcohol no tiene efecto sobre la úlcera duodenal (falso negativo).

El objetivo de todo estudio es llegar a una conclusión verdaderamente positiva o verdaderamente negativa y la probabilidad de que la conclusión del estudio coincida con la verdad es igual a $1-\beta$. Esta probabilidad se denomina potencia o poder

Se plantea una hipótesis nula (H_0) y una alternativa (H_A). La primera apoya la no diferencia ($A = B$) y la segunda la diferencia que se pretende detectar ($A \neq B$).

Un paso esencial en toda investigación es la formulación de una hipótesis que se intenta rechazar (hipótesis nula) mediante el uso de una prueba de significación estadística, y la formulación de otra hipótesis que se acepta si se rechaza la hipótesis nula (hipótesis alternativa, que coincide con la pregunta del investigador).

En todo contraste de hipótesis se obtiene un valor de p , que es el grado de significación estadística.

Habitualmente, el investigador pretende comparar 2 o más grupos de estudio (diseño experimental u observacional). El cálculo del tamaño muestral es necesario para obtener el número de elementos de estudio (individuos, ratas) que le permitan alcanzar el grado de significación p que rechace la hipótesis nula de no asociación.

estadístico del estudio y es la capacidad que tiene el estudio de encontrar diferencias cuando éstas existen. Tal y como se ha descrito para el valor α , el valor para el riesgo β se establece con frecuencia en el 20%, o lo que es igual, se considera aceptable un poder estadístico del 80%, es decir, en 8 ocasiones de cada 10 se podrá detectar el efecto real, por ejemplo, en 8 de cada 10 pacientes tratados con omeprazol se podrá detectar un descenso del número de recurrencias de úlcera duodenal.

Es fundamental en el diseño del estudio decidir previamente el riesgo α y β que el investigador desea asumir, y no recurrir de forma rutinaria a los establecidos como estándar (α de 0,05 y β de 0,20).

La elección de un valor para el riesgo α debe depender de la magnitud de la evidencia que se requiere para rechazar la hipótesis nula. Por ejemplo, si se está probando la eficacia de un nuevo fármaco para tratar una enfermedad para la que no existe tratamiento eficaz, no conviene asumir un valor α muy pequeño, puesto que la ausencia de un tratamiento médico actual eficaz puede hacer que se obtengan resultados falsamente positivos. Las consecuencias de tomar la decisión de asumir la existencia de diferencias, es decir, de rechazar la H_0 , deben condicionar la elección del valor de α (p. ej., se valora la eficacia de un nuevo producto para el tratamiento de la úlcera duodenal, y este producto es muy similar al que se toma como referencia en cuanto a efectos secundarios, coste, administración, etc. Por lo tanto, se precisa un grado de evidencia más elevado, es decir, un valor α más bajo).

A pesar de la amplia aceptación en las publicaciones científicas de esta metodología desde sus orígenes, existe un gran debate respecto a su validez metodológica y, sobre todo, respecto a su aplicación habitual, controversia que va calando, aunque de forma muy lenta, en la comunidad investigadora frente a la inercia de su aplicación como meras recetas para la toma de decisiones¹¹.



Tamaño muestral

Para determinar el tamaño muestral de un estudio debemos considerar diferentes situaciones⁵⁻⁷:

- Estudios para determinar parámetros (estudio descriptivo). Se pretende la aproximación a valores poblacionales (proporciones, medias) a partir de una muestra.
- Estudios para contraste de hipótesis (estudio analítico). Se pretende comparar si las medias o las proporciones de los grupos de estudio son diferentes (hipótesis alternativa e hipótesis del estudio).

Estudios para determinar parámetros (estudio descriptivo)

Con estos estudios, transversales o series de casos, se pretende hacer inferencias a valores poblacionales (proporciones, medias) a partir de una muestra. El objetivo del estudio es descriptivo.



Estimar una proporción

Si deseamos estimar una proporción, debemos saber:

- El grado de confianza o seguridad ($1-\alpha$). El grado de confianza prefijado da lugar a un coeficiente (Z_α) cuyo valor para una seguridad del 95% es de 1,96 y para una seguridad del 99% es de 2,58.
- La precisión o error muestral que se desea (d).
- Una idea del valor aproximado del parámetro (p) que queremos medir (en este caso, una proporción). Esta idea se puede obtener revisando la bibliografía y/o por estudios piloto previos. En caso de no tener esta información, utilizaremos el valor $p = 0,5$ (50%), valor que determina el tamaño muestral mayor.

Ejemplo: ¿A cuántas personas tendríamos que estudiar para conocer la prevalencia de úlcera duodenal?

Seguridad = 95%; precisión = 3%; proporción esperada = asumamos que puede ser próxima al 5%; si no tuviésemos ninguna idea de esta proporción utilizaríamos el valor $p = 0,5$ (50%), que maximiza el tamaño muestral:

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

donde

- $Z_\alpha^2 = 1.962$ (ya que la seguridad es del 95%)
- $p =$ proporción esperada (en este caso, 5% = 0,05)
- $q = 1 - p$ (en este caso, $1 - 0,05 = 0,95$)
- $d =$ precisión (en este caso deseamos un 3%)

$$n = \frac{1,96^2 \times 0,05 \times 0,95}{0,03^2}$$

Si la población es finita, es decir, conocemos el total de la población y deseásemos saber cuántos del total tendremos que estudiar, la respuesta sería:

$$n = \frac{N \times Z_\alpha^2 \times p \times q}{d^2 \times (N-1) + Z_\alpha^2 \times p \times q}$$

donde

- $N =$ total de la población
- $Z_\alpha^2 = 1,96^2$ (si la seguridad es del 95%)
- $p =$ proporción esperada (en este caso, 5% = 0,05)
- $q = 1 - p$ (en este caso, $1 - 0,05 = 0,95$)
- $d =$ precisión (en este caso deseamos un 3%)

¿A cuántas personas tendría que estudiar de una población de 15.000 habitantes para conocer la prevalencia de úlcera duodenal?

Seguridad = 95%; precisión = 3%; proporción esperada = asumamos que puede ser próxima al 5%; si no tuviese ninguna idea de esta proporción utilizaríamos el valor $p = 0,5$ (50%), que maximiza el tamaño muestral.

$$n = \frac{15.000 \times 1,96^2 \times 0,05 \times 0,95}{0,03^2 (15.000 - 1) + 1,96^2 \times 0,05 \times 0,95} = 200$$

Estimar una media

Si deseamos estimar una media debemos conocer:

- El grado de confianza o seguridad ($1 - \alpha$). El grado de confianza prefijado da lugar a un coeficiente (Z_α). Para una seguridad del 95% = 1,96; para una seguridad del 99% = 2,58.
- La precisión con que se desea estimar el parámetro ($2 \times d$ es la amplitud del intervalo de confianza).
- Una idea de la varianza S^2 de la distribución de la variable cuantitativa que se supone existe en la población.

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

Ejemplo: Si deseamos conocer la media de la glucemia basal de una población, con una seguridad del 95 % y una precisión de ± 3 mg/dl y tenemos información por un estudio piloto o revisión bibliográfica que la varianza es de 250 mg/dl:

$$n = \frac{1,96^2 \times 250}{3^2} = 106,7$$

Si la población es finita, como previamente se señaló, es decir conocemos el total de la población y desearíamos saber cuántos del total tendríamos que estudiar, la respuesta sería:

$$n = \frac{N \times Z_\alpha^2 \times S^2}{d^2 \times (N-1) + Z_\alpha^2 \times S^2}$$

Estudios para contraste de hipótesis

Estos estudios pretenden comparar si las medias o las proporciones de las muestras son diferentes. Habitualmente, el investigador pretende comparar 2 tratamientos (diseño experimental u observacional). Para el cálculo del tamaño muestral se precisa conocer:

- La magnitud de la diferencia a detectar que tenga interés clínicamente relevante. Se pueden comparar 2 proporciones o 2 medias.
- Tener una idea aproximada de los parámetros de la variable que se estudia (bibliografía y/o estudios previos).
- Grado de significación *a priori* o p (probabilidad de cometer un error α).
- Poder estadístico ($1 - \beta$) (probabilidad de cometer un error β).
- Definir si la hipótesis va a ser unilateral o bilateral.

- Bilateral: cualquiera de los dos parámetros a comparar (medias o proporciones) puede ser mayor o menor que el otro. No se establece dirección ($A \neq B$).

- Unilateral: Cuando se considera que uno de los parámetros debe ser mayor que el otro, indicando, por tanto, una dirección de las diferencias ($A > B$).

El tamaño muestral y, por lo tanto, la formulación dependerá del:

- Tipo de la variable estudiada.
- Tipo de estudio. Los diseños apareados requieren menor tamaño muestral y los estudios observacionales exigen mayor tamaño muestral, dado que es imprescindible construir modelos multivariantes.
- Período de seguimiento.
- Porcentaje de no respuestas o de pérdidas en el seguimiento. En todos los estudios es preciso estimar las posibles pérdidas de pacientes por razones diversas (pérdida de información, abandono, ausencia de respuesta), por lo que se debe incrementar el tamaño muestral respecto a estas pérdidas.
- Número de pruebas de significación (múltiples contrastes) en un mismo estudio o análisis.
- Número de análisis por estratos (análisis de subgrupos).

Es evidente que el cálculo del tamaño muestral tiene importantes connotaciones éticas económicas y estadísticas. Así, un cálculo desmesuradamente grande en un diseño experimental puede situar en situación de riesgo a los últimos pacientes del estudio si antes se hubiera podido determinar la eficacia de, por ejemplo, la medida terapéutica evaluada; económicas por cuanto un cálculo ajustado permitirá una adecuada contención de los costes del estudio y estadísticas por las razones obvias ya detalladas.

Bibliografía



● Importante ●● Muy importante

- Contandriopoulos AP, Champagne F, Potvin L, Denis JL, Boyle P. Preparar un proyecto de investigación. Barcelona: SG Editores, 1991.
- Fernández Pérez C, Moreno Sierra J, Silmi Moyano A, Resel Estevez L. Investigación clínica en urología y metodología científica. Arch Esp Urol 2003;56:573-6.
- Fernández Pérez C, Barreales Tolosa L, Coll Torres E. La estadística como una herramienta de apoyo y no como una herramienta de decisión. Arch Esp Urol 2003; 56:595-604.
- Martín Andrés A, Luna del Castillo JD. Bioestadística para las ciencias de la salud. 4.ª ed. Madrid: NORMA, 1993.
- Hulley SB, Cummings SR. Diseño de la investigación clínica. Un enfoque epidemiológico. Barcelona: Doyma, 1993.
- Cook TD, Campbell DT. Quasi-Experimentation. Design & Analysis Issues for Field Settings. Boston: Houghton Mifflin Company, 1979.
- Kleinbaum DG, Kupper LI, Morgenstern H. Epidemiologic Research. Principles and quantitative methods. Belmont, California: Lifetime Learning Publications 1982;54:186-7.
- Dawson-Saunders B, Trapp RG. Bioestadística Médica. 2.ª ed. México: Editorial el Manual Moderno, 1996.
- Milton JS, Tsokos JO. Estadística para biología y ciencias de la salud. Madrid: Interamericana McGraw Hill, 2001.
- Rosner B. Fundamentals of biostatistics. Boston: Duxbury Press, 2000.
- Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. Br Med J 1986;292:746-50.