

Tamaño de la muestra

El cálculo del tamaño de la muestra permite responder a las siguientes preguntas: ¿cuántos individuos es necesario estudiar para poder estimar un parámetro determinado con el grado de confianza deseado? ¿Cuántos sujetos es necesario estudiar para tener las mínimas garantías de poder detectar una determinada diferencia entre los grupos de estudio, en el supuesto de que esta diferencia exista realmente?

El resultado del cálculo debe considerarse orientativo, ya que se basa en asunciones que pueden ser incorrectas. Por ello, es importante realizarlo cuidadosamente, ya que tiene implicaciones sobre la viabilidad del estudio y puede suponer modificaciones en su diseño (modificación de los criterios de selección, prolongación del período de inclusión, etc.).

La inclusión de un número excesivo de sujetos encarece el estudio, tanto desde el punto de vista económico como de los recursos humanos y físicos. Además, puede ser considerado poco ético, ya que un mayor número de pacientes estará sometido innecesariamente a determinadas exploraciones o recibirá la intervención menos eficaz. Por otra parte, 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 erróneamente a la conclusión de que no existen.

En general, puede considerarse que el objetivo de una investigación pertenece a uno de los dos tipos siguientes: *estimación de un parámetro* (p. ej., el porcentaje de pacientes hipertensos que presentan una determinada complicación) o *contraste de hipótesis* (p. ej., evaluar cuál de dos pautas terapéuticas consigue un mayor porcentaje de éxitos).

ESTUDIOS QUE ESTIMAN UN PARÁMETRO

Se trata de estudios en los que los investigadores, partiendo de los valores observados en la muestra estudiada, y utilizando la inferencia estadística, buscan estimar el valor de un parámetro en la población.

Consideraciones previas: intervalo de confianza

Supongamos que se desea estimar el valor medio de la colesterolemia en una población. Se escoge una muestra de 80 sujetos en los que se observa una media de 180 mg/100 ml, con una desviación estándar de 20 mg/100 ml. Si se hubiera estudiado una muestra diferente, seguramente se habrían obtenido cifras distintas aunque los criterios de selección hubiesen sido los mismos, si bien es probable que el valor observado en esta segunda muestra hubiera estado alrededor de 180 mg/100 ml. Por lo tanto, a partir del valor observado en una única muestra no puede conocerse exactamente el verdadero valor en la población de origen, ya que es tan solo una aproximación (estimación puntual). Para conocer entre qué límites es más probable que se sitúe este verdadero valor, debe calcularse el *intervalo de confianza* (IC).

El método de cálculo del IC en las situaciones más frecuentes se desarrolla en los capítulos dedicados al análisis de los datos. La amplitud del IC, es decir, la precisión de la estimación, depende del grado de confianza, la variabilidad del parámetro de interés y el número de sujetos estudiados. Cuanto menor sea la variabilidad del parámetro y mayor el número de sujetos, mayor precisión existirá

en la estimación para un grado de confianza determinado. Cuanta más confianza se desee tener, más amplio será el IC y menor será la precisión obtenida.

Cálculo del tamaño de la muestra

Para calcular el tamaño de la muestra debe conocerse:

- La *variabilidad* del parámetro que se desea estimar. Si no se conoce, puede obtenerse una aproximación a partir de datos propios o de otras investigaciones, o de un estudio piloto. En el caso de las variables cuantitativas se mide por la variancia, y en el de las cualitativas, por el producto $p \cdot (1 - p)$, siendo p el valor esperado de la proporción de interés.
- La *precisión* con que se desea obtener la estimación, es decir, la amplitud del IC. Cuanto más precisa se desee, más estrecho deberá ser este intervalo, y más sujetos deberán ser estudiados. La precisión debe fijarse previamente, en función de la finalidad de la estimación. En algunos casos puede requerirse una gran precisión, mientras que en otros, si solo se necesita conocer aproximadamente entre qué valores se encuentra el parámetro, se requerirá una menor precisión y, en consecuencia, menos sujetos.
- El *nivel de confianza* deseado, que habitualmente se fija en el 95%, correspondiente a un valor α de 0,05. Indica el grado de confianza que se tiene de que el verdadero valor del parámetro en la población se sitúa en el intervalo obtenido. Cuanta más confianza se desee, menor será el valor de α , y más elevado será el número de sujetos necesario.

De estos tres elementos, solo debe conocerse la variabilidad del parámetro, ya que tanto la precisión como el nivel de confianza se fijan de acuerdo con los intereses del investigador.

Estimación de una proporción

La fórmula para el cálculo del número de sujetos para estimar una proporción se presenta en el [cuadro 16.1](#). También puede

Cuadro 16.1 Fórmulas para el cálculo del número de sujetos necesarios para la realización de un estudio cuyo objetivo es la estimación de una media o una proporción

Estimación de una proporción (variable cualitativa):

$$N = \frac{Z\alpha^2 P(1-P)}{i^2}$$

Estimación de una media (variable cuantitativa):

$$N = \frac{Z\alpha^2 s^2}{i^2}$$

N : número de sujetos necesarios.

$Z\alpha$: valor de Z correspondiente al riesgo α fijado (v. [tabla 16.4](#)).

P : valor de la proporción que se supone existe en la población.

s^2 : variancia de la distribución de la variable cuantitativa que se supone que existe en la población.

i : precisión con que se desea estimar el parámetro ($2 \cdot i$ es la amplitud del intervalo de confianza).

utilizarse la tabla A que se encuentra al final del libro.

Ejemplo 16.1. Un estudio se propone estimar el porcentaje de pacientes hipertensos de un centro que están controlados. A partir de datos previos, se estima que debe estar situado alrededor del 40% ($p = 0,40$). Se quiere realizar la estimación con una precisión de $\pm 4\%$ ($i = 0,04$) y una confianza del 95% ($1 - \alpha = 0,95$; $\alpha = 0,05$; $Z\alpha = 1,96$). Aplicando la fórmula correspondiente, se calcula que se precisan 576 sujetos. La tabla A proporciona el mismo resultado.

Esta estimación del número de sujetos realizada *a priori* no es más que una aproximación. Si en el ejemplo el porcentaje real de pacientes controlados es diferente del 40%, puede alterarse el grado de precisión de la estimación.

En la [tabla 16.1](#) puede apreciarse cómo diferentes grados de precisión y diferentes

TABLA 16.1 Influencia de la precisión de la estimación y del valor supuesto de la proporción que se desea estimar sobre el número de sujetos necesarios

Influencia de la precisión		
P	i	N
0,40	0,100	92
0,40	0,075	164
0,40	0,050	369
0,40	0,025	1.475
0,40	0,010	9.220
Influencia de la proporción		
P	i	N
0,10	0,05	138
0,20	0,05	246
0,30	0,05	323
0,40	0,05	369
0,50	0,05	384

Todos los cálculos se han realizado con el grado de confianza del 95%.

i: medida de la precisión de la estimación, que corresponde a la mitad del intervalo de confianza; *N*: número de sujetos necesarios para realizar la estimación de *P* con la precisión deseada y el 95% de confianza; *P*: valor supuesto de la proporción que se desea estimar.

valores del porcentaje esperado modifican el número de sujetos necesarios. Así, por ejemplo, para estimar el porcentaje del 40% del ejemplo 16.1 con el mismo nivel de confianza del 95%, los 576 sujetos se convierten en 9.220 cuando se desea una precisión muy alta ($i = 0,01$), o en tan solo 92 si se es menos exigente ($i = 0,1$). Modificando cualquiera de estos valores, puede obtenerse un número de individuos que se aproxime al «deseado» o al disponible. Debe evitarse cualquier manipulación del cálculo, ya que, al reducir el número de sujetos que se van a estudiar, también disminuye el grado de precisión con que el parámetro va a ser estimado y aumenta la amplitud del IC.

En el cálculo del tamaño de la muestra también debe tenerse en cuenta la estrategia de análisis y cómo se presentarán los resultados. Así, si en el ejemplo 16.1 los investigadores presentan el resultado principal en función del sexo, la estratificación hará que la estimación se haya obtenido en un número menor de sujetos en cada estrato, por lo que la precisión será menor de la deseada.

En algunas ocasiones no se conoce el valor aproximado del parámetro que se está buscando. Si no existen datos de la literatura que resulten útiles, o si no puede realizarse una prueba piloto para obtener una primera aproximación a dicho valor, puede adoptarse la postura de la máxima indeterminación. Esta actitud consiste en suponer que el porcentaje que se quiere estimar se sitúa alrededor del 50%, ya que es el valor que requiere una mayor cantidad de individuos para una precisión determinada, de forma que el número obtenido supera cualquier otra combinación de parámetros. Esta actitud también es útil, por ejemplo, cuando se realizan encuestas en las que interesa estimar múltiples parámetros. Su inconveniente es que requiere el estudio de un número de sujetos a veces muy superior al realmente necesario.

Estimación de una media

El cálculo es similar al de la situación anterior (v. cuadro 16.1). También deben fijarse el nivel de confianza y la precisión de la estimación, que se traducen en la amplitud del intervalo alrededor de la media que se desea estimar. En las variables cuantitativas, la medida de la variabilidad viene proporcionada por la variancia de su distribución en la población. La tabla B del final del libro proporciona una fácil solución alternativa al cálculo matemático.

Ejemplo 16.2. Supongamos que se desea estimar la presión arterial diastólica (PAD) de los pacientes diabéticos registrados en un centro de salud. Por estudios previos se conoce que la desviación estándar de la PAD en los sujetos diabéticos es de 25 mmHg ($s = 25 \text{ mmHg}$; $s^2 = 625 \text{ mmHg}^2$). Se quiere realizar la estimación con una confianza del 95% ($1 - \alpha = 0,95$; $\alpha = 0,05$)

y una precisión de 5 mmHg ($i = 5$). Aplicando la fórmula, se puede determinar que son necesarios 96 sujetos. Esta cifra debe aumentarse en función del porcentaje de sujetos en que se prevea que no podrá determinarse la PAD. El mismo resultado se obtiene consultando la tabla B en la línea correspondiente al valor $i/s = 5/25 = 0,200$.

Si no se dispone de una estimación de la variabilidad, puede utilizarse una regla práctica, que consiste en determinar la diferencia entre los valores máximo y mínimo esperables. Dado que se asume que esta variable es de distribución normal, el intervalo $m \pm (2s)$, siendo m la media y s la desviación estándar de la distribución, incluye aproximadamente el 95% de los valores posibles, por lo que, al dividir dicha amplitud de valores entre 4, puede obtenerse una cierta aproximación al valor de la desviación estándar s .

Corrección para poblaciones finitas

En los cálculos anteriores no ha intervenido el tamaño de la población, ya que en ellos se ha asumido que es infinito. Sin embargo, en muchas ocasiones es preciso obtener una muestra de una población de tamaño conocido (finito). En esta situación puede aplicarse la siguiente fórmula, que ajusta el número de individuos necesarios según el tamaño de la población:

$$n_a = \frac{n}{1+n/N}$$

donde n_a es el número de sujetos necesarios, n es el número de sujetos calculado para poblaciones infinitas y N es el tamaño de la población.

Ejemplo 16.3. En el ejemplo 16.1 se ha calculado que son necesarios 576 sujetos para estimar el porcentaje de hipertensos controlados de un centro, realizando determinadas asunciones. Si la población de referencia es de 1.000 hipertensos, aplicando la fórmula anterior podría determinarse que el número necesario se reduce a 365 pacientes hipertensos.

ESTUDIOS QUE COMPARAN DOS GRUPOS

Muchos estudios tienen por objetivo responder a preguntas del siguiente tipo: ¿es más eficaz el fármaco A que el B en el tratamiento de una enfermedad? ¿Existe un aumento del riesgo de malformaciones congénitas en los niños nacidos de madres que han padecido cierto trastorno durante el embarazo?, etc.

Consideraciones previas: pruebas de contraste de hipótesis

Bajo las preguntas que se han planteado en el párrafo anterior subyace una hipótesis, por ejemplo, la de que B es más eficaz que A en el tratamiento de los hipertensos esenciales de edad avanzada. En la realidad, esta hipótesis puede ser verdadera o falsa, y eso es lo que se trata de averiguar. Dado que el investigador no puede estudiar todos los hipertensos esenciales de edad avanzada, la hipótesis debe contrastarse en una muestra de esta población. A partir únicamente de los datos recogidos en un estudio, no puede aceptarse o rechazarse absolutamente la hipótesis que se contrasta. Existe cierto riesgo de error, sea cual sea la conclusión a la que se llegue.

Error α y error β

Desde el punto de vista estadístico, el investigador se enfrenta al problema del contraste de una hipótesis mediante la asunción de que no existen diferencias en la eficacia de los dos fármacos A y B (*hipótesis nula*). A partir de los resultados observados en la muestra, el investigador utilizará las pruebas de significación estadística para evaluar si existe la suficiente evidencia que le permita rechazar esta hipótesis nula y, consecuentemente, aceptar la hipótesis alternativa de que sí existen diferencias entre ambas terapéuticas.

El *error tipo I*, conocido también como *error α* (tabla 16.2), se comete cuando el investigador rechaza la hipótesis nula, siendo esta verdadera en la población. Es equivalente a encontrar un resultado falso positivo, ya que el investigador concluye que hay una diferencia, cuando en realidad no existe.

TABLA 16.2 Tipos de error aleatorio en una prueba estadística de contraste de hipótesis

		Realidad (población)	
		Existe diferencia o asociación (Ho falsa)	No existe diferencia o asociación (Ho cierta)
Resultado de la prueba (muestra)	Diferencia o asociación significativa (rechazo de Ho)	No error	Error tipo I α
	Diferencia o asociación no significativa (no rechazo de Ho)	Error tipo II β	No error

Ho: hipótesis nula.

El *error tipo II*, o *error β* (v. tabla 16.2), se comete en la situación contraria, cuando el investigador no rechaza la hipótesis nula, siendo esta falsa en la población. Es equivalente a un resultado falso negativo, ya que el investigador concluye que ha sido incapaz de encontrar una diferencia que existe en la realidad. Si β representa la probabilidad de un resultado falso negativo, su complementario $1 - \beta$, conocido como *poder* o *potencia estadística*, representa la probabilidad de observar en la muestra una determinada diferencia o un efecto, si existen en la población.

Para calcular el tamaño de la muestra, se fijan *a priori* los riesgos que se está dispuesto a asumir de cometer estos errores. Lo ideal sería poder reducirlos a cero, pero ello no es posible. Una de las formas para disminuirlos consiste en aumentar el número de sujetos estudiados.

Hipótesis unilaterales e hipótesis bilaterales

Desde el punto de vista estadístico, otro aspecto que afecta al cálculo del tamaño de la muestra es el de la formulación precisa de la hipótesis que se contrasta. Si se pretende determinar si dos fármacos A y B difieren en su eficacia, sin decir cuál de ellos ha de ser mejor o peor, se habla de *hipótesis bilateral*, o de dos colas (fig. 16.1). En cambio, si la hipótesis alternativa es que el fármaco A es más eficaz que el B, se habla de *hipótesis*

unilateral, o de una cola, ya que únicamente interesa evaluar una de las direcciones de la comparación.

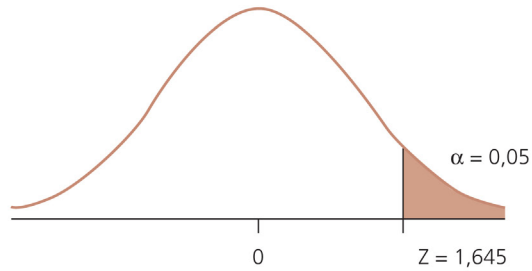
Desde un punto de vista estricto, son preferibles las hipótesis bilaterales, ya que son más conservadoras y requieren un mayor número de participantes.

Magnitud de la diferencia o de la asociación

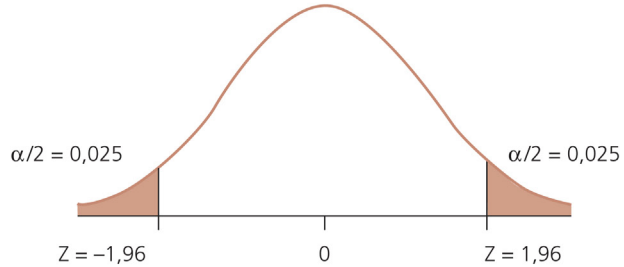
La magnitud de la diferencia o de la asociación tiene una gran influencia sobre el tamaño de la muestra. Si la diferencia es muy amplia (p. ej., si la eficacia del fármaco A es el doble que la del B), se podrá detectar con más facilidad que si es de escasa magnitud, y requerirá un número menor de sujetos. En la tabla 16.3 puede apreciarse cómo una diferencia de 40 mg/100 ml en la media de colesterol de dos muestras puede ser significativa, aunque se hayan estudiado tan solo 20 sujetos, mientras que una pequeña diferencia de 3 mg/100 ml no alcanza suficiente significación aunque se hayan estudiado 200 individuos en cada grupo.

Para calcular el tamaño de la muestra, debe definirse la mínima magnitud de la diferencia o de la asociación que se desea detectar y que se considera de relevancia clínica. Esta diferencia debe fijarse en términos realistas, dentro del rango de valores esperables. A menudo, es tentador utilizar diferencias más amplias, ya que se obtiene un tamaño de la

Prueba unilateral*
 $H_0: A \leq B$
 $H_a: A > B$



Prueba bilateral**
 $H_0: A = B$
 $H_a: A \neq B$



*Especifica la dirección de la diferencia.

**Interesa determinar si A y B son diferentes, sin especificar la dirección de la diferencia. Dado que se asume que la distribución es normal, el valor de α se distribuye simétricamente en las dos colas de la distribución.

H_0 : hipótesis nula.

H_a : hipótesis alternativa.

FIGURA 16.1 Valores de la distribución normal tipificada (Z) correspondientes al valor α del 5% en las pruebas estadísticas unilaterales y bilaterales.

TABLA 16.3 Influencia del número de sujetos estudiados y la magnitud de la diferencia sobre la significación estadística

Ejemplo 1. Una diferencia de gran magnitud es estadísticamente significativa incluso con un reducido número de sujetos				
	n	Colesterolemia (media \pm DE)	t	p
Muestra 1	20	260 \pm 20		
Muestra 2	20	220 \pm 20	6,32	<0,001
Ejemplo 2. Una diferencia de escasa magnitud no es estadísticamente significativa, aunque se haya estudiado un elevado número de sujetos				
	n	Colesterolemia (media \pm DE)	t	p
Muestra 1	200	218 \pm 20		
Muestra 2	200	215 \pm 20	1,5	0,14

DE: desviación estándar; n: número de sujetos de la muestra; p: probabilidad de cometer un error α ; t: valor de la t de Student-Fisher de comparación de dos medidas independientes. Habitualmente, valores menores de 0,05 se aceptan como evidencia suficiente para decidir que la diferencia es estadísticamente significativa.

muestra más reducido. Esta actitud debe evitarse, ya que implica que diferencias menores a la establecida, a pesar de que pueden ser de importancia clínica, no alcanzarán el grado de significación estadística con el número de sujetos estudiados. Si, por ejemplo, se utiliza una diferencia del 50% en la eficacia de los dos fármacos en el cálculo del tamaño de la muestra, se necesitará un número reducido de sujetos, pero si se observa una diferencia del 40%, que puede ser muy importante desde el punto de vista clínico, no será estadísticamente significativa, ya que el investigador ha decidido considerar como importantes solo las diferencias superiores al 50%.

En función del enfoque del estudio, esta magnitud puede fijarse como una diferencia entre grupos, o bien, en términos relativos, como un cociente entre riesgos.

Cálculo del tamaño de la muestra

Para calcular el tamaño de la muestra deben utilizarse los siguientes elementos principales (cuadro 16.2):

- Definir la hipótesis que se va a contrastar, precisando si es unilateral o bilateral.
- Establecer el riesgo de cometer un error de tipo I (α) que se está dispuesto a aceptar. Habitualmente suele aceptarse un riesgo α del 5%, y preferiblemente con hipótesis bilaterales, ya que son más conservadoras.
- Establecer, así mismo, el riesgo que se acepta de cometer un error tipo II (β). Habitualmente se sitúa entre el 5 y el 20%. La elección variará en función de las consecuencias que pueda tener cometer dicho error. A menudo es más fácil enfrentar esta decisión a partir del concepto de poder o potencia estadística ($1 - \beta$), que es la capacidad del estudio para detectar una determinada diferencia. Aceptar el riesgo de cometer un error β del 20% significa que, si la diferencia que se busca existe en la realidad, el estudio tiene un 80% de probabilidades (potencia) de detectarla.
- Definir la mínima magnitud de la diferencia, el efecto o la asociación que se quiera ser capaz de detectar. Debe basarse en

Cuadro 16.2 Elementos que intervienen en el cálculo del número de sujetos necesarios para la realización de un contraste de hipótesis

1. Hipótesis unilateral o bilateral.
2. Riesgo aceptado de cometer un error α .
3. Riesgo aceptado de cometer un error β (potencia: $1 - \beta$).
4. Magnitud mínima de la diferencia o asociación que se considera importante detectar.
5. Variabilidad de la variable de respuesta en el grupo de referencia.
6. Otros factores:
 - Prueba estadística que se utilizará en el análisis.
 - Tipos especiales de diseño del estudio.
 - Comparación de más de dos grupos.
 - Grupos de tamaño diferente.
 - Equivalencia de las intervenciones.
 - Estratificación por las características basales.
 - Múltiples variables de respuesta.
 - Porcentaje de pérdidas y abandonos esperado durante el seguimiento.

datos de estudios previos o de la literatura que definan el rango de los valores esperables, y en la mínima magnitud que se considere de relevancia clínica.

- Es necesario, también, disponer de alguna medida de la variabilidad de la variable de respuesta en la población o el grupo de referencia.

De estos cinco elementos, solo debe conocerse el último, ya que los cuatro primeros son fijados por el investigador. A continuación, se aplica la fórmula correspondiente, que depende de la prueba estadística que vaya a ser utilizada en el análisis. En el cuadro 16.3 se presentan las fórmulas de uso más habitual, correspondientes a la comparación de dos variables dicotómicas (dos proporciones) o de dos variables cuantitativas (dos medias). Las tablas 16.4 y 16.5 recogen los valores de la variable normal tipificada, correspondientes a los riesgos α y β utilizados con mayor frecuencia en dicho cálculo, respectivamente.

Cuadro 16.3 Fórmulas para el cálculo del número de sujetos necesarios por grupo en un estudio cuyo objetivo es la comparación de dos muestras del mismo tamaño

Comparación de dos proporciones (variable cualitativa):

$$N = \frac{[Z\alpha\sqrt{2P\cdot(1-P)} + Z\beta\sqrt{P1\cdot(1-P1)+P2\cdot(1-P2)}]^2}{(P1-P2)^2}$$

Comparación de dos medias (variable cuantitativa):

$$N = \frac{2 \cdot (Z\alpha + Z\beta)^2 \cdot s^2}{d^2}$$

N : número de sujetos necesarios en cada uno de los grupos.

$Z\alpha$: valor de Z correspondiente al riesgo α fijado (v. tabla 16.4).

$Z\beta$: valor de Z correspondiente al riesgo β fijado (v. tabla 16.5).

$P1$: valor de la proporción que se supone existe en el grupo de referencia.

$P2$: valor de la proporción que se supone existe en el grupo de estudio.

$P2 - P1$: valor mínimo de la diferencia que se desea detectar (variable cualitativa).

P : media ponderada de las proporciones $P1$ y $P2$.

s^2 : variancia de la distribución de la variable cuantitativa que se supone que existe en el grupo de referencia.

d^2 : valor mínimo de la diferencia que se desea detectar (variable cuantitativa).

Tabla 16.4 Valores de $Z\alpha$ utilizados con mayor frecuencia en el cálculo del tamaño de la muestra

α	$Z\alpha$	
	Prueba unilateral	Prueba bilateral
0,200	0,842	1,282
0,150	1,036	1,440
0,100	1,282	1,645
0,050	1,645	1,960
0,025	1,960	2,240
0,010	2,326	2,576

Tabla 16.5 Valores de $Z\beta$ utilizados con mayor frecuencia en el cálculo del tamaño de la muestra

β	Potencia ($1 - \beta$)	$Z\beta$
0,01	0,99	2,326
0,05	0,95	1,645
0,10	0,90	1,282
0,15	0,85	1,036
0,20	0,80	0,842
0,25	0,75	0,674
0,30	0,70	0,524
0,35	0,65	0,385
0,40	0,60	0,253
0,45	0,55	0,126
0,50	0,50	0,000

Las fórmulas pueden resultar algo complicadas, por lo que se recomienda utilizar tablas como las presentadas al final del libro (tablas C y D) o aplicaciones informáticas que resuelven el cálculo.

Comparación de dos proporciones

La comparación de dos proporciones es, probablemente, la prueba de contraste de hipótesis más utilizada en investigación clínica. Se trata de estudios que comparan dos grupos, y para ello utilizan una variable de respuesta cualitativa dicotómica.

Ejemplo 16.4. Un estudio tiene por objetivo determinar si un nuevo tratamiento T consigue un mayor porcentaje de éxitos en las sobreinfecciones respiratorias que el tratamiento estándar E. Lo primero que debe conocerse es el porcentaje de curaciones en pacientes de características similares a los que van a ser estudiados, obtenido con el tratamiento estándar E. Supongamos que esta cifra se sitúa alrededor del 40% ($P1 = 0,4$).

El siguiente paso es determinar la diferencia mínima que se quiere detectar, es decir, responder a la pregunta: ¿a partir de qué porcentaje de éxitos con el nuevo tra-

tamiento se considerará que este es mejor que E y, por lo tanto, se estará dispuesto a modificar la pauta terapéutica habitual? Es decir, si el porcentaje de individuos curados con T es del 41%, ¿puede considerarse que esta diferencia del 1% es un resultado lo suficientemente importante para modificar la pauta terapéutica? ¿O se exigirá un mínimo, por ejemplo, del 50% de éxitos? La respuesta a estas preguntas depende de muchos factores, tales como la seguridad del fármaco, la facilidad de administración o el coste, entre otros. Supongamos que los investigadores consideran que, si se cura el 50% de pacientes con T ($P_2 = 0,5$), se aceptará como la elección terapéutica.

A continuación, solo falta determinar los niveles de riesgo de cometer algún tipo de error aleatorio que se está dispuesto a asumir. Supongamos que se acepta el nivel de riesgo α habitual del 5% con una hipótesis bilateral y un riesgo β del 20% (potencia: $1 - \beta = 0,80$).

Aplicando la fórmula del [cuadro 16.3](#), puede calcularse que son necesarios 387 sujetos por grupo de estudio. Con la tabla C del final del libro se obtiene la misma cifra.

Este número indica los sujetos que deben finalizar el estudio para tener un 80% de probabilidades de detectar una diferencia igual o superior a la fijada, con un nivel de error α del 5%. Por lo tanto, hay que incrementarlo en función del número de pérdidas de seguimiento y de abandonos que se prevea que ocurrirán durante el estudio.

El tamaño de la muestra requerido es mayor cuanto más pequeña es la diferencia que se desea detectar y menos elevados son los riesgos α y β que se aceptan. Es importante destacar que una misma diferencia del 5% entre dos porcentajes requiere el máximo número de pacientes cuando estos se sitúan alrededor del 50%. Ello se debe a que, si el porcentaje del grupo de referencia es del 5%, una diferencia del 5% implica que el grupo de estudio debe pasar a un 10%, es decir, dobla la eficacia. En cambio, si el porcentaje del grupo de referencia es del 50%, una diferencia del 5% implica solamente un incremento relativo del 10%, es decir, pasar del 50 al 55%.

Estimación de un riesgo relativo

Si el objetivo del estudio es determinar la magnitud de la asociación en términos relativos, la medida que se utilizará en los estudios prospectivos (ensayos clínicos y estudios de cohortes) es el riesgo relativo (RR). Lo que debe fijarse en este caso es la mínima magnitud del RR que se quiere ser capaz de detectar. En esta situación puede utilizarse la misma fórmula que en el caso de la comparación de dos proporciones, teniendo en cuenta que $RR = P_2/P_1$. La magnitud de la asociación que se quiere detectar corresponde al mínimo RR que se considerará de importancia clínica, donde P_1 es el riesgo de desarrollar la enfermedad en el grupo no expuesto y P_2 el riesgo de desarrollarla en el grupo expuesto ($P_2 = P_1 \cdot RR$).

La tabla D del final del libro presenta el resultado del cálculo del número de sujetos necesarios para los estudios que estiman un RR en las situaciones más habituales.

Ejemplo 16.5. Se pretende realizar un estudio de cohortes con el objetivo de estimar el riesgo de desarrollar un infarto agudo de miocardio asociado a la hipercolesterolemia. El riesgo esperado en la cohorte no expuesta (P_1) es del 10% a los 10 años (incidencia acumulada). El RR que se considera relevante detectar es de 2. El valor del riesgo de cometer un error α será el convencional del 5% (hipótesis bilateral) y el de cometer un error β , del 20% (potencia: $1 - \beta = 0,80$). A partir de estos datos, puede calcularse el valor de P_2 :

$$P_2 = P_1 \cdot RR = 0,10 \cdot 2 = 0,20 \text{ (o } 20\%)$$

Aplicando la fórmula del [cuadro 16.3](#) o la tabla D del final del libro, puede determinarse que son necesarios 199 sujetos por grupo. Este número representa el tamaño mínimo tanto del grupo de expuestos como del de no expuestos. Por lo tanto, si desea estudiar una cohorte general, o si se deben identificar los sujetos expuestos a partir de una muestra de la población general, se deberá determinar el número de sujetos que habrá que examinar para identificar el número suficiente de expuestos. Si la frecuencia de la exposición en la población general (f) fuera de un 20%, sería necesario examinar a $N/f = 199/0,2 = 995$ sujetos para encontrar a 199 expuestos.

Estimación de una odds ratio

En los estudios de casos y controles, la magnitud de la asociación se estima mediante la *odds ratio* (OR). Se trata de una situación similar a la de la estimación de un RR, en la que debe conocerse la proporción de exposición esperada en los controles (P1) y fijar la magnitud de la OR que se desea detectar. Análogamente a la situación anterior, la proporción esperada de exposición entre los casos es:

$$P2 = P1 \cdot OR / [1 + P1 \cdot (OR - 1)]$$

La tabla E del final del libro presenta el resultado del cálculo del número de sujetos necesarios para los estudios que estiman una OR en las situaciones más habituales.

Ejemplo 16.6. Un estudio de casos y controles pretende evaluar el riesgo de padecer un infarto agudo de miocardio asociado al uso de anticonceptivos orales en mujeres jóvenes. Se estima que la proporción esperada de uso de anticonceptivos orales en la población candidata a ser incluida en el grupo control es del 30% ($P1 = 0,30$). El incremento mínimo del riesgo que se quiere detectar es de 3 ($OR = 3$). El riesgo α es de 0,05 (hipótesis bilateral) y el riesgo β es de 0,10 (potencia: $1 - \beta = 0,90$). Por lo tanto:

$$P2 = (0,3 \cdot 3) / [1 + 0,3 \cdot (3 - 1)] = 0,56$$

Aplicando la fórmula del [cuadro 16.3](#), puede determinarse que son necesarios 74 sujetos por grupo. La tabla E del final del libro proporciona un resultado similar.

Comparación de dos medias

Para calcular el tamaño de la muestra debe conocerse el valor de la variancia de la distribución de la variable de respuesta en el grupo de referencia, decidir la mínima diferencia que se considera de relevancia clínica y si interesa poder detectar y fijar los niveles de error α y β que se aceptan.

Ejemplo 16.7. Un estudio se propone comparar los valores de fructosamina de dos grupos de diabéticos sometidos a dos pautas terapéuticas diferentes, T1 y T2, con la finalidad de determinar cuál de ellas consigue un mayor control de la enfermedad. Se utiliza habitualmente la pauta T1, con

la que se consigue unos valores medios de fructosamina de 3 mmol/l, con una desviación estándar de 0,625 mmol/l. Por lo tanto, el valor de la variancia s^2 es de $(0,625)^2 = 0,39$. A continuación, se establece que la diferencia mínima entre ambos grupos que se considera de relevancia clínica es de 0,5 mmol/l ($d = 0,5$). El riesgo de error α que se está dispuesto a asumir es de 0,05. Dado que se pretende evaluar qué pauta es más eficaz, se trata de una hipótesis bilateral. Interesa tener una capacidad del 90% para detectar la diferencia, si existe realmente ($1 - \beta = 0,90$).

Para utilizar la tabla F del final del libro debe calcularse el valor d/s , que corresponde al cociente entre la diferencia que se quiere detectar y la desviación estándar de la variable de respuesta en el grupo de referencia. En este caso, corresponde a $0,5/0,625 = 0,8$. Buscando este valor en la primera columna de las tablas, y leyendo la fila correspondiente al valor de a de 0,05 en una hipótesis bilateral, en el punto de intersección con la potencia de 0,90 se obtiene un tamaño de 33 sujetos por grupo; es decir, en total, debe incluirse en el estudio un mínimo de 66 personas. El mismo resultado se obtiene utilizando la fórmula del [cuadro 16.3](#).

Comparación de grupos cuando la variable de respuesta es ordinal

Cuando la variable de respuesta se mide en una escala ordinal, o cuando en el análisis deben utilizarse pruebas estadísticas no paramétricas, el cálculo es más complejo. Por criterios de operatividad, suele calcularse el tamaño como si fuera a utilizarse una prueba paramétrica. La estimación será tanto mejor cuanto menos se aleje la distribución de la variable de la ley normal.

En el anexo 8 se presenta el cálculo del tamaño de la muestra en otras situaciones.

Equivalencia de dos intervenciones

En ocasiones se diseña un estudio para determinar si una nueva intervención, que ofrece alguna ventaja sobre la de comparación, como ser menos costosa o más segura, es tan eficaz como ella (estudios de equivalencia) o no es terapéuticamente inferior (estudios de no inferioridad). En esta situación, la determinación del tamaño de la muestra

requiere tener en cuenta algunas consideraciones especiales. El problema radica en que una prueba de significación estadística no permite demostrar la equivalencia de dos intervenciones. La ausencia de una diferencia estadísticamente significativa entre ellas no debe interpretarse como prueba de que sean terapéuticamente equivalentes.

La demostración de la equivalencia requiere descartar pequeñas diferencias en la respuesta observada. Para considerar que dos intervenciones son equivalentes, la diferencia entre ellas debe ser tan pequeña que pueda considerarse clínicamente irrelevante. Por ello, debe definirse cuál es la máxima diferencia que puede existir entre dos intervenciones para considerarlas equivalentes. Este margen de equivalencia, conocido como *delta* (Δ), debe ser menor que la mínima diferencia clínicamente importante, que es la que se utilizaría para determinar el tamaño de la muestra en un estudio de superioridad. En los estudios de equivalencia, una forma de calcular el tamaño de la muestra necesario es tomar el valor del margen de equivalencia como la diferencia que se desea ser capaz de detectar. Sin embargo, este procedimiento infraestima el número de sujetos necesarios.

Tal como se comenta en el capítulo 5, la interpretación de los resultados de los estudios de equivalencia se basa en el análisis del IC de la diferencia observada, de manera que se considera que existe equivalencia si todo el IC se sitúa dentro de los márgenes de equivalencia (o si su límite inferior está por encima de dicho margen en los estudios de no-inferioridad). Para obtener un IC lo bastante estrecho es necesario estudiar un número de sujetos sustancialmente mayor que en los estudios de superioridad. Por lo tanto, en el cálculo del tamaño de la muestra en este tipo de estudios debe tenerse en cuenta no solo el margen de equivalencia establecido, sino también el valor esperado de la respuesta (Tamayo-Sarver et al., 2005), tal como se muestra en la fórmula del **cuadro 16.4**. La tabla G del final del libro presenta el número de sujetos necesarios para estudios de equivalencia en las situaciones más habituales cuando la variable de respuesta es dicotómica.

Cuadro 16.4 Fórmula para el cálculo del tamaño de la muestra en un estudio de equivalencia

$$N = \frac{2 \cdot P \cdot (1-P) \cdot (Z\alpha + Z\beta)^2}{d^2}$$

Equivalencia entre dos medias (variable cuantitativa):

$$N = \frac{2 \cdot s^2 \cdot (Z\alpha + Z\beta)^2}{(d - d^*)^2}$$

N: número de sujetos necesarios en cada uno de los grupos.

Z α : valor de *Z* correspondiente al riesgo α fijado (v. **tabla 16.4**).

Z β : valor de *Z* correspondiente al riesgo β fijado (v. **tabla 16.5**).

P: proporción que se espera en el grupo control (en caso de una variable cualitativa)

*S*²: variancia de la distribución de la variable de respuesta (en caso de una variable cuantitativa)

d: diferencia máxima entre las respuestas a ambos tratamientos que se tolerará para concluir que son equivalentes (margen de equivalencia)

*d**: diferencia esperada entre las respuestas a ambos tratamientos

Ejemplo 16.8. Un estudio se propone evaluar la eficacia de una pauta monodosis para el tratamiento de las infecciones urinarias frente a la pauta habitual de 7 días, cuya eficacia se sitúa alrededor del 90% ($P = 0,90$). La monodosis es menos costosa y facilita el cumplimiento del paciente, por lo que interesa determinar si la monodosis y la pauta de 7-10 días pueden considerarse equivalentes. Los investigadores definen que considerarán que lo son si la diferencia de eficacia entre ellas no supera el 10% (margen de equivalencia $\Delta = 0,10$). Por otro lado, los investigadores esperan encontrar que la monodosis es ligeramente inferior, en un 5%, a la pauta de 7 días ($d^* = 0,05$). Se aceptan los niveles de error α del 0,05 (hipótesis bilateral) y β del 20% (potencia: $1 - \beta = 0,80$). Aplicando la fórmula del **cuadro 16.4** para una variable dicotómica, puede determinarse que son necesarios 565 sujetos por grupo. La tabla G del final del libro proporciona el mismo resultado.

El mismo resultado se obtendría si el estudio se plantease con una hipótesis de no-inferioridad, en el que los investigadores definirían que considerarían que la monodosis es mejor alternativa terapéutica si no es más de un 10% inferior a la pauta de 7 días.

Este procedimiento suele sobreestimar el número de sujetos necesarios. Para obtener un cálculo más ajustado, puede utilizarse una fórmula basada en el cálculo del IC, tal como se comenta a continuación.

CÁLCULO DEL TAMAÑO DE LA MUESTRA BASADO EN INTERVALOS DE CONFIANZA

Algunos autores piensan que si los resultados de un estudio se analizan, se interpretan y se presentan mediante IC, el cálculo del tamaño de la muestra debería basarse también en IC.

Al interpretar los resultados de un estudio se recomienda tener en cuenta la diferencia mínima que se considera relevante desde el punto de vista clínico (v. cap. 28). Si el límite inferior del IC se encuentra por encima del valor de esta diferencia mínima, se considera que el resultado del estudio es no solamente estadísticamente significativo, sino también concluyente a favor de la existencia de una diferencia clínicamente relevante. De modo similar, si el valor superior del IC está por debajo del valor de la diferencia mínima, el resultado es concluyente a favor de la ausencia de una diferencia clínicamente relevante.

En algunas ocasiones existe desacuerdo entre el método del cálculo del tamaño de la muestra basado en el contraste estadístico de hipótesis y la interpretación de los resultados basada en los IC. Este desacuerdo es evidente cuando el investigador ha calculado el tamaño de la muestra necesario para detectar, con una determinada potencia estadística, una diferencia mínima entre dos tratamientos, pero el análisis del IC del resultado muestra que valores inferiores a dicha diferencia son compatibles con los datos observados. En esta situación, el investigador no puede concluir con certeza cuál es la mejor alternativa terapéutica.

Ejemplo 16.9. En un estudio que comparaba la cirugía convencional (P1) con la radical (P2) en el tratamiento del cáncer gástrico, el cálculo del tamaño de la muestra se basó en la asunción de una diferencia en la supervivencia a los 5 años del 14% (P1 = 20%; P2 = 34%) (Fayers et al., 2000). Tomando como base estas tasas de supervivencia, se requerían 400 pacientes por grupo si se deseaba una potencia estadística del 90%. Al inicio del estudio, un grupo de 26 cirujanos rellenaron un cuestionario en el que manifestaban que una mejora en la tasa de supervivencia del 9,5% podía considerarse clínicamente relevante. Si los resultados del estudio hubieran sido los de las asunciones usadas para el cálculo del tamaño de la muestra, y 400 pacientes hubieran finalizado el estudio en cada grupo, se obtendría una diferencia del 14% con un IC 95% cuyos límites serían 5,4 y 22,6%. Este resultado, aunque es estadísticamente significativo, no descarta que la diferencia en la supervivencia pueda ser inferior al 9,5%, por lo que no sería concluyente.

Este problema se puede evitar basando el cálculo del tamaño de la muestra en los IC mediante la siguiente fórmula:

$$n = \frac{8z_{\alpha/2}^2 p(1-p)}{W_{CI}^2}$$

donde n es el número de individuos por grupo, Z_{α} es el valor de Z correspondiente al riesgo α fijado, P es la proporción del parámetro estimado en la población ($P = [P1 + P2]/2$) y W_{IC} es la amplitud del IC.

Ejemplo 16.10. Consideramos el estudio planteado en el ejemplo anterior (Fayers et al., 2000). Aplicando la fórmula:

$$n = \frac{8 \cdot 1,96^2 \cdot 0,27 \cdot 0,73}{W_{IC}^2}$$

Si pensamos que la diferencia más realista entre grupos es del 14% (P1 = 20%, P2 = 34%) y deseamos que el límite inferior del IC del resultado esté por encima de la mínima diferencia clínicamente relevante (9,5%), la diferencia entre ambos valores es del 5,5%, por lo que la amplitud del IC será del 11% ($W_{IC} = 2 \times 0,055 = 0,11$). Aplicando la fórmula anterior, se obtiene que se requieren 748 pacientes por grupo.

Debido a que W_{IC} será habitualmente un valor pequeño, el tamaño de la muestra será superior al calculado con el método tradicional, el resultado será más preciso y con la suficiente potencia estadística. Además, este abordaje tiene la ventaja de facilitar la comunicación entre el clínico y el estadístico. Para el clínico es más difícil entender conceptos como el de la potencia estadística y el contraste de hipótesis que la interpretación de un IC.

Tal como hemos comentado en el apartado anterior, este enfoque es especialmente adecuado en el caso de los estudios de equivalencia y de no inferioridad.

Ejemplo 16.11. Retomemos el ejemplo 16.8 sobre un estudio hipotético que se propone evaluar la eficacia de una pauta monodosis para el tratamiento de las infecciones urinarias frente a la pauta habitual de 7 días, cuya eficacia se sitúa alrededor del 90% ($P = 0,90$). Los investigadores habían establecido el margen de equivalencia delta (o de no inferioridad) en el 10%. Dado que esperaban encontrar una diferencia del 5%, su IC debería tener una amplitud máxima que permitiera descartar el valor de dicho margen (un 5% por encima y por debajo del valor observado, por lo que $W_{IC} = 2 \times 0,05 = 0,10$). Aplicando la fórmula, se determina que son necesarios 336 sujetos por grupo, estimación inferior a los 565 calculados por el procedimiento utilizado en el ejemplo 16.8.

De hecho, si se hubieran estudiado 336 sujetos por grupo y se hubiera obtenido una diferencia del 5% entre los grupos, los límites del IC 95% de esta diferencia serían 0,00013 y 0,09987 (es decir, 0,01 y 9,99%), lo que permite concluir la no inferioridad de la monodosis (en caso de obtenerse estos resultados), ya que todo el IC queda por debajo del margen del 10% fijado.

Si se hubieran estudiado 565 sujetos por grupo, como se había calculado anteriormente, y se hubiera encontrado una diferencia del 5%, los límites del IC 95% en este caso serían 1,2 y 8,8%. Este resultado permitiría llegar a la misma conclusión, pero habiendo estudiado un mayor número de sujetos.

CORRECCIÓN POR LAS NO RESPUESTAS, LAS PÉRDIDAS Y LOS ABANDONOS

El número de sujetos calculado debe ampliarse en función del porcentaje de no respuestas, pérdidas y abandonos durante el seguimiento y de las desviaciones del protocolo que puedan esperarse, de forma que se asegure que finalizará el estudio el número de pacientes requerido. Una fórmula para hacerlo es la siguiente:

$$N_a = N[1/(1 - R)]$$

donde N representa el número de sujetos teórico; N_a , el número de sujetos ajustado; y R , la proporción esperada de pérdidas.

Ejemplo 16.12.. Para realizar un estudio se ha calculado que son necesarios 300 sujetos ($N = 300$) y se espera un 20% de pérdidas ($R = 0,20$). El número de pacientes que deberían iniciar el estudio sería $N_a = 300 [1/(1 - 0,2)] = 375$ sujetos.

La utilización de esta fórmula asegura que el estudio mantenga la potencia estadística deseada, pero no evita que se puedan producir sesgos si las no respuestas, las pérdidas o los abandonos no se han producido aleatoriamente, es decir, si los sujetos que finalizan el estudio no son representativos de los que lo inician (lo que suele ser habitual).

ESTRATEGIAS PARA MINIMIZAR EL NÚMERO DE SUJETOS NECESARIO

Estas estrategias se basan en conseguir una población homogénea, disminuir la variabilidad de las medidas y aumentar la frecuencia de aparición del fenómeno de interés, por lo que deben aplicarse siempre que sea posible para aumentar la potencia del estudio (cuadro 16.5).

Algunas de estas estrategias enumeradas pueden implicar una modificación de las hipótesis de trabajo. Por ello, es importante realizar, ya en fases tempranas del diseño, una estimación aproximada del número de sujetos necesario, con el fin de evaluar la factibilidad del estudio y discutir si es necesario replantear o adaptar la hipótesis de trabajo o el estudio en general.

Cuadro 16.5 Estrategias para disminuir el número de sujetos necesarios en un estudio

1. Al seleccionar la variable de respuesta, usar:
 - Variables cuantitativas que aumentan la potencia de las pruebas estadísticas.
 - Variables precisas.
 - Variables de presentación frecuente.
2. Al medir las variables, usar:
 - Promedios de medidas repetidas.
 - Técnicas de medición precisas.
 - Criterios y técnicas estandarizados.
3. Seleccionar la población de estudio:
 - Homogéneas (criterios de selección restrictivos).
 - Poblaciones de alto riesgo, en las que es más frecuente observar la respuesta.
4. Al diseñar el estudio, valorar si es posible:
 - Asignar aleatoriamente los sujetos a los grupos.
 - Usar como medida de la respuesta las diferencias individuales entre el inicio y el final del estudio.
 - Usar diseños especiales (p. ej., diseños cruzados).
 - Estratificar por factores pronósticos.

BIBLIOGRAFÍA DE LOS EJEMPLOS

Fayers PM, Cuschieri A, Fielding J, Craven J, Uscinska B, Freedman LS. Sample size calculation for clinical trials: the impact of clinician beliefs. *Br J Cancer* 2000;82:213-9.

BIBLIOGRAFÍA

Bland JM. The tyranny of power: is there a better way to calculate sample size? *BMJ* 2009;339:b3985.

Campbell MJ, Julious SA, Altman DG. Estimating sample size for binary, ordered categorical, and continuous outcomes in two group comparison. *BMJ* 1995;311:1145-8.

Kerry SM, Bland JM. Sample size in cluster randomisation. *BMJ* 1998;316:549.

Kosinski AS, Chen Y, Lyles RH. Sample size calculations for evaluating a diagnostic test when the gold standard is missing at random. *Stat Med* 2010;29:1572-9.

Machin D, Campbell MJ, Tan SB, Tan SH. Sample size tables for clinical studies. Chichester: Wiley; 2009.

Obuchowski NA. Sample size calculations in studies of test accuracy. *Stat Methods Med Res* 1998;7:371-92.

Schlesselman JJ. Case-control studies: design, conduct, analysis. Nueva York: Oxford University Press; 1982.

Tamayo-Sarver JH, Albert JM, Tamayo-Sarver M, Cydulka RK. Advanced statistics: how to determine whether your intervention is different, at least as effective as, or equivalent: a basic introduction. *Acad Emerg Med* 2005;12:536-42.

Wittes J. Sample size calculations for randomized controlled trials. *Epidemiol Rev* 2002;24:39-53.