Universidad Miguel Hernández de Elche

Comisión de Ética en la Investigación Experimental

Universidad Miguel Hernández de Elche

http://ceie.umh.es/calculo-del-tamano-muestral/

Cálculo del tamaño muestral

Cálculo del tamaño muestral:

La estimación del número de sujetos requeridos para responder a una pregunta experimental es un paso importante en la planificación de un estudio. Por un lado, una muestra de tamaño excesivo puede resultar en pérdidas de vidas animales y otros recursos, incluyendo tiempo y dinero, ya que una información igualmente válida podría haber sido extraída con un número más reducido de animales. Sin embargo, una subestimación del tamaño de la muestra también representa un desperdicio, ya que un tamaño insuficiente de la muestra tiene una baja probabilidad de detectar una diferencia estadísticamente significativa entre los grupos, incluso si esa diferencia realmente existe. Por lo tanto, un investigador podría concluir erróneamente que los grupos no difieren, cuando en realidad sí que lo hacen.

¿Qué debemos considerar para calcular el tamaño de muestra?

Mientras que la necesidad de llegar a una estimación apropiada de tamaño de la muestra es clara, muchos científicos no están familiarizados con los factores que influyen en la determinación del tamaño de la muestra y con las técnicas para el cálculo del tamaño muestral estimado. Un rápido vistazo a cómo la mayoría de los libros de estadística tratan este tema nos da una idea de por qué muchos investigadores ven los cálculos del tamaño muestral con miedo y confusión.

Si bien los cálculos del tamaño muestral pueden llegar a ser muy complicados, es importante destacar, en primer lugar, que todas estas técnicas producen estimaciones, y, en segundo lugar, que solo hay unos pocos factores principales que influyan en estas estimaciones. Como resultado, es posible obtener estimaciones razonables usando algunas fórmulas relativamente simples.

Al comparar los dos grupos, los principales factores que influyen en el tamaño de la muestra son los siguientes:

1) Cómo de grande debe ser la diferencia para que pueda detectarla

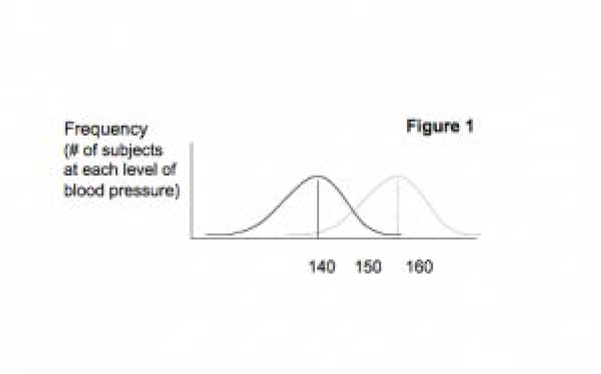
2) Cuánta variabilidad existe en el factor de interés

3) Qué valor “p” tiene previsto utilizar como criterio de significación estadística

4) Cuánta seguridad desea tener de que detectará una diferencia estadísticamente significativa, en el supuesto de que esa diferencia exista

Una mirada intuitiva en un ejemplo sencillo

Supongamos que usted está estudiando individuos con hipertensión renal, y quiere poner a prueba la eficacia de un medicamento que se supone que reduce la presión arterial. Planea comparar la presión arterial sistólica en dos grupos, uno que es tratado con una inyección de placebo, y un segundo grupo que es tratado con el fármaco que se está probando. Si bien no sabemos todavía cuáles son las presiones arteriales que habrá en cada uno de estos grupos, supongamos que si tuviera que probarlo en un número ridículamente elevado de sujetos (por ejemplo 100.000) tratados con placebo o medicamento, su presión arterial sistólica seguiría dos claras distribuciones de frecuencia distintas, como se muestra en la figura 1.



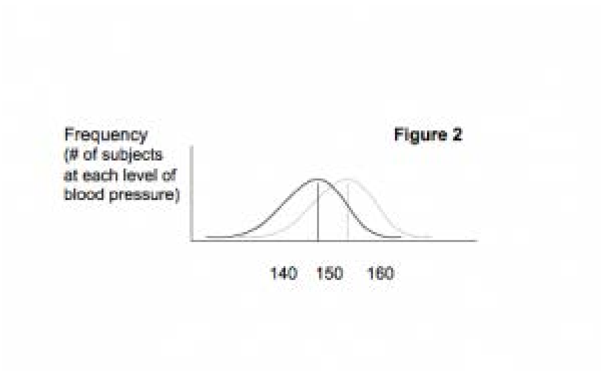
Como usted esperaría, ambos grupos muestran algo de variabilidad en la presión sanguínea, y la distribución de la frecuencia de las presiones observadas es conforme a un curva con forma de campana. Como se muestra aquí, los dos grupos se solapan, pero son claramente diferentes; las presiones sistólicas en el grupo tratado son una media de 20 mm Hg menores que en los controles no tratados.

Como había 100.000 en cada grupo, podemos estar seguros de que los grupos son diferentes. Supongamos ahora que, aunque hubieramos tratado a 100.000, sólo se hubieran obtenido las mediciones de presión de tres de cada grupo, debido a que el aparato de medición de la presión se hubiera roto. En otras palabras, tendríamos una muestra aleatoria de N = 3 de cada grupo, y sus presiones sistólicas serían así:

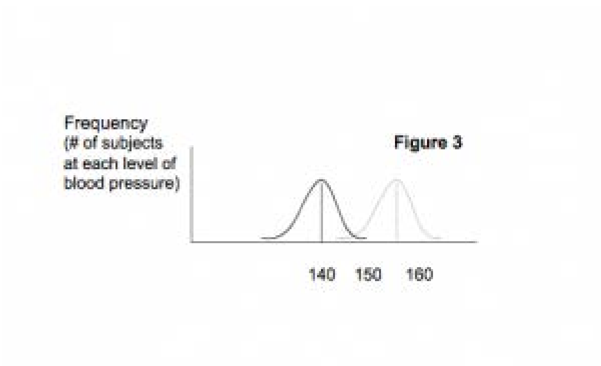
Grupo placebo	**Grupo tratado**
160	155
150	140
140	140

Las presiones son menores en el grupo tratado, pero no podemos estar seguros de que el tratamiento fue exitoso. Hay una clara posibilidad de que la diferencia que vemos sea sólo debido a la casualidad, ya que se tomó una muestra al azar. Entonces la pregunta es: ¿cuántos tendríamos que medir (muestra) en cada grupo para tener la seguridad de que las diferencias observadas no fueron simplemente resultado de la casualidad?

El tamaño de muestra necesario depende de los cuatro factores enumerados anteriormente. Para ilustrar esto intuitivamente, supongamos que las presiones sanguíneas en los sujetos tratados y no tratados se distribuyesen como se muestra en la figura 2 o en la Figura 3.



En la Figura 2 la cantidad de variabilidad es la misma, pero la diferencia entre los grupos es más pequeña. Tiene sentido que se necesite un número mayor de muestras para estar seguro de que las diferencias en la muestra son reales.



En la figura 3 la diferencia de presiones es aproximadamente la misma que en la figura 1, pero hay menos variabilidad en las lecturas de presión dentro de cada grupo. Aquí parece obvio que haría falta una muestra más pequeña para determinar con seguridad la diferencia.

El tamaño de la muestra necesario también depende del “valor p” que utilice. Un “valor de p” menor de 0,05 se utiliza con frecuencia como criterio para decidir si es probable que las diferencias observadas sean debidas al azar. Si p <0,05, significa que la probabilidad de que la diferencia observada se deba al azar es menor del 5%. Si desea utilizar un criterio más rígido (por ejemplo, p <0,01) se necesita una muestra mayor. Finalmente, el tamaño de la muestra tendrá también depende de la “potencia”, que es la probabilidad de observar una diferencia estadísticamente significativa, en el supuesto de que es diferencia realmente exista.

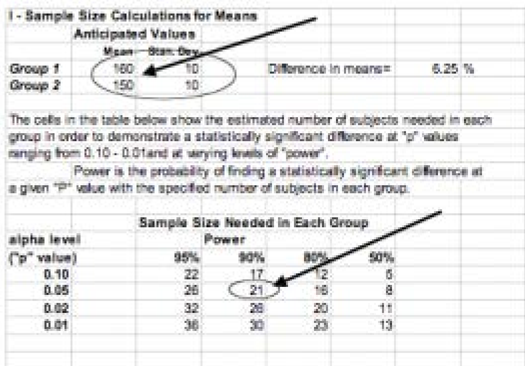
En resumen, con el fin de calcular una estimación del tamaño muestral se necesita hacer alguna estimación de lo diferente que los grupos pueden ser o cómo de grande debe ser la diferencia para que seamos capaces de detectarla, y también se necesita una estimación de la cantidad de variabilidad que habrá dentro de los grupos. Además, en los cálculos también hay que tener en cuenta el “valor de p” que desea utilizar y cuánta “potencia” desea.

Información necesaria para hacer cálculos de tamaño muestral

Puesto que no se ha realizado el experimento aún, no se sabe cómo de diferentes serán los grupos o cuál será la variabilidad (medida por la desviación típica). Pero, en general, se pueden hacer conjeturas razonables. Tal vez de su experiencia (o de los datos anteriormente publicados) se anticipa que los sujetos hipertensos no tratados tienen una presión arterial sistólica media de mm Hg sobre 160 con una desviación estándar de aproximadamente 10 mm Hg. Usted decide que una reducción de la presión arterial sistólica a una media de 150 mm Hg representaría una reducción clínicamente significativa. Dado que nadie ha hecho nunca antes este experimento, no sé cuánta variabilidad habrá en la respuesta, por lo que tendrá que asumir que la desviación estándar para el grupo de prueba es por lo menos tan grande como la de los controles no tratados. A partir de estas estimaciones se puede calcular una estimación del tamaño de la muestra que necesita en cada grupo.

1. Cálculos del tamaño muestral para una diferencia de medias

Los cálculos reales pueden ser un poco engorrosos, y la mayoría de la gente ni siquiera quiere ver ecuaciones. Por lo tanto, puede usar la **hoja de cálculo (ANEXO II)** incluida en la página web del CEEA, que hace todos los cálculos automáticamente. Todo lo que tiene que hacer es introducir las medias estimadas y las desviaciones estándar para cada grupo. En el programa ejemplo se ha asumido que el grupo control (grupo 1) tiene una media de 160 y una desviación estándar de 10. Queríamos saber cuántos sujetos iba a necesitar en cada grupo para detectar una diferencia significativa de 10 mm Hg. Así que, hemos considerado apropiada una media de 150 para el grupo 2 y asumido que la desviación estándar para este grupo sería la misma que para el grupo 1.


I - Sample Size Calculations for Means

	Anticipated Values			
	Mean	Stan. Dev.		
Group 1	160	10	Difference in means=	6.25 %
Group 2	150	10		

The cells in the table below show the estimated number of subjects needed in each group in order to demonstrate a statistically significant difference at "p" values ranging from 0.10 - 0.01and at varying levels of "power".

Power is the probability of finding a statistically significant difference at a given "P" value with the specified number of subjects in each group.

Sample Size Needed in Each Group

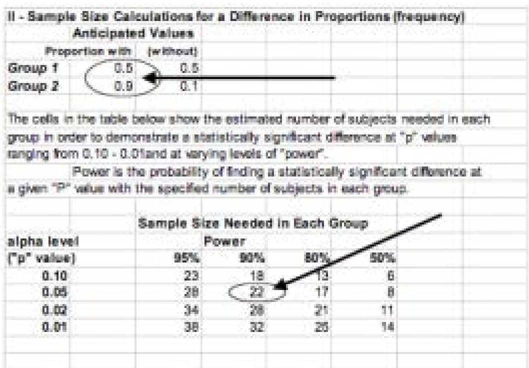
alpha level ("p" value)	Power 95%	90%	80%	50%
0.10	22	17	12	5
0.05	26	21	16	8
0.02	32	26	20	11
0.01	36	30	23	13

La hoja de cálculo realmente genera una tabla que muestra estimaciones de tamaños muestrales para diferentes valores de “p” y diferentes niveles de potencia. Muchas personas arbitrariamente utilizan p = 0,05 y un nivel de potencia del 80%. Con estos parámetros se necesitarían cerca de 16 pacientes en cada grupo. Si usted quisiera un 90% de potencia, se necesitarían cerca de 21 pacientes en cada grupo.

El formato de esta hoja de cálculo hace que sea fácil jugar a “qué pasaría si”. Si quisiera tener una idea de cuántos sujetos podría necesitar si el tratamiento redujera la presión 20 mm Hg, sólo necesitaría cambia la media para el grupo 2 a 140, y todos los cálculos se volverían a hacer automáticamente.

2. Cálculo del tamaño muestral para una diferencia en las proporciones

La página siguiente de la misma hoja de cálculo genera cálculos de tamaño muestral para la comparación de diferencias en la frecuencia de un evento. Supongamos, por ejemplo, que un tratamiento dado fue exitoso el 50% del tiempo y que quería probar un nuevo tratamiento con la esperanza de que tendría éxito el 90% del tiempo. Todo lo que tiene que hacer es conectar estos (como fracciones definidas entre 0 y 1) en la hoja de cálculo, y los tamaños de muestra estimados se calculan automáticamente como se muestra aquí:


II - Sample Size Calculations for a Difference in Proportions (frequency)

Anticipated Values

	Proportion with	(without)
Group 1	0.5	0.5
Group 2	0.9	0.1

The cells in the table below show the estimated number of subjects needed in each group in order to demonstrate a statistically significant difference at "p" values ranging from 0.10 - 0.01and at varying levels of "power".

Power is the probability of finding a statistically significant difference at a given "P" value with the specified number of subjects in each group.

Sample Size Needed in Each Group

alpha level ("p" value)	Power 95%	90%	80%	50%
0.10	23	18	13	6
0.05	28	22	17	8
0.02	34	28	21	11
0.01	38	32	26	14

La ilustración de la hoja de cálculo muestra que para tener un 90% de probabilidad de mostrar una diferencia estadísticamente significativa (con $P < 0,05$) en estas proporciones propuestas (50 y 90%), necesitaría cerca de 22 pacientes en cada grupo.

Fuente: Boston University Research Committees