La reducción del número de animales de experimentación y el cálculo del tamaño muestral: una mesa con cinco patas.

Villamayor, F
Trial Form Support SA, Barcelona

Recibido 25 septiembre de 2014 / Aceptado 10 noviembre 2014

Resumen: El cálculo del tamaño muestral necesario para la consecución de los objetivos de un experimento está basado en cuatro factores interdependientes: tamaño de efecto, nivel de significación, potencia estadística, y variabilidad de la muestra. El trabajo de planificación previo a la ejecución del estudio es fundamental para obtener el máximo de información con el número mínimo de animales.

Palabras clave: Estadística, tamaño muestral, reducción.

Abstract: Reduction of the number of animals in experiments and sample size calculation: a five-legged table. Sample size calculation to accomplish with the objectives of an experiment is based in four interdependent factors: effect size, significance level, statistical power and sample variability. The planning of the study, prior to execution, is fundamental to obtain the maximum of information from the minimum number of animals.

Keywords: Statistics, sample size, reduction.

Introducción: planificación y tamaño muestral

La segunda ley de la Termodinámica establece que, un sistema aislado, sin un aporte externo de energía, tiende a la uniformidad. En cambio, mediante una aportación de energía externa, es posible crear orden y por tanto generar información [1]. Un experimento no deja de ser en cierto modo un sistema en el que siguen siendo válidas las leyes de la Termodinámica. Por tanto, si no se aporta energía, el resultado final va a ser el caos. La energía deberá aportarla el investigador responsable. Y una parte importante tendrá su fuente en el diseño o planificación del experimento. Es importante insistir en este hecho: Cuanto más esfuerzo se dedique al planificar, mayor será la cantidad de información que se obtenga, mejor será su calidad y, por tanto, los resultados serán más fiables. No es esto ninguna opinión, ni responde a un principio filosófico o una forma de pensar: Es consecuencia directa de una de las leyes de la Física.

El investigador va a planificar su experimento en base a una hipótesis de partida. Cree que ésta puede ser cierta, y dedica sus esfuerzos a reunir datos que la apoyen. Una vez analizados estos datos, decidirá si va a considerar cierta su hipótesis, o no. Pero, dado que se está operando con el método científico, existe una cierta incertidumbre en la decisión que se tome: según la decisión que tome, su resultado, puede ser erróneo.

Es lógico que se plantee si existe alguna manera de controlar esta incertidumbre. El problema es reunir la suficiente evidencia. La materia de que está formada la evidencia son los datos. Y la cantidad de datos está directamente relacionada con el número de unidades experimentales que se utilicen. La unidad experimental es el sujeto experimental mínimo e independiente del resto sobre la que van a medirse los efectos en los que está interesada la investigación. El numero de sujetos experimentales que se utilicen en un estudio es lo que se denomina su "tamaño muestral". Por tanto, cuanto mayor sea el tamaño muestral mayor va a ser la evidencia que permitirá verificar si la hipótesis que se quiere probar es cierta o no. Pero, por muchas razones, el tamaño muestral no puede ser tan grande como se desee. Hay que poner unos límites, principalmente éticos o presupuestarios. El tamaño muestral adecuado debe ser el mínimo que permita llegar a los objetivos del estudio. La utilización de un tamaño muestral adecuado va a permitir controlar la incertidumbre en la decisión de si la hipótesis de trabajo es cierta o no. En el resto de este artículo verá cómo puede lograrse esto y las implicaciones que conlleva.

Material y Métodos: Datos necesarios para el cálculo del tamaño muestral.

No existe un único método para el cálculo del tamaño muestral. Una ojeada a Internet nos revela que existe como mínimo una fórmula distinta para cada diseño experimental y para cada objetivo que pueda plantearse. Por ejemplo, en el Centre for Clinical Research and Biostatistics (CCRB) de la Chinese University of Hong Kong publican una útil página web (http://www.cct.cuhk.edu.hk/stat/) para el cálculo en línea del tamaño muestral para diversos diseños y objetivos experimentales. No cubre todos los casos, pero podemos usarlo para ilustrar lo que sigue, y su funcionamiento es simple (existen otros recursos, algunos gratuitos (ver Anexo), así como programas comerciales para el cálculo del tamaño muestral, que cubren los supuestos no contemplados por el recurso del CCRB). En primer lugar hay que escoger el tipo de experimento que se propone realizar: comparar medias, proporciones, supervivencias, etc. Y luego hay que escoger el diseño: Una sola muestra, dos muestras independientes, dos muestras dependientes, etc. Y finalmente hay que escoger la hipótesis de partida: Igualdad entre tratamientos, prueba de no-inferioridad, prueba de superioridad, prueba de equivalencia...

En este punto se abre un formulario en el que hay que introducir los datos o parámetros necesarios para el cálculo del tamaño muestral necesario para el experimento. Existe un formulario para cada uno de los posibles caminos en el árbol de decisión previo: Objetivo-Diseño-Hipótesis. No obstante, los conceptos comunes subyacentes son comunes.

Tomemos por ejemplo el caso de la comparación de dos medias en un diseño en paralelo para la prueba de igualdad. Para efectuar el cálculo

* e-mail: fvillam@icloud.com

es necesario definir las dos hipótesis que van a ser contrastadas: La denominada "hipótesis nula" (H_0) va a ser que no existen diferencias entre las medias de los dos grupos experimentales, es decir $_2$- $_1$=0. Frente a ella se define la "hipótesis alternativa" (H_1) que es la que quiere probarse: Existe una cierta diferencia entre las medias de ambos grupos experimentales, es decir $_2$- $_1$ 0. Los materiales o parámetros necesarios para el cálculo son [2]:

Nivel de significación ( )

El nivel de significación es la probabilidad de un error estadístico de tipo I. Es el que se cometería si se aceptase como cierta la hipótesis alternativa cuando en realidad es falsa (probabilidad de un resultado *falso positivo*). Habitualmente se adopta como referencia =0,05 (es decir, que exista un 5% de probabilidades de que en un experimento se observen diferencias entre medias que no se corresponden con la realidad).

Potencia estadística (1-)

Es el complementario de , el error estadístico de tipo II, que se comete cuando se decide rechazar la hipótesis alternativa cuando en realidad es cierta (probabilidad de un resultado *falso negativo*). La potencia estadística es por tanto la probabilidad de aceptar la hipótesis alternativa cuando es realmente cierta. Habitualmente se trabaja con una potencia estadística que toma valores entre un 80% y un 90%.

Variabilidad muestral (^{2})

Se expresa mediante la varianza de la muestra (^{2}).

Tamaño de efecto ($_1$- $_2$).

Es la diferencia mínima que se desea que el experimento pueda detectar entre las medias de los dos grupos experimentales, con un nivel de significación y una potencia 1- . Debe ser una diferencia que aporte significado a la investigación, y dentro de los límites de plausibilidad que marque el modelo experimental que se vaya a utilizar.

Resultados

Vamos a ilustrar los resultados del cálculo del tamaño muestral con el ejemplo descrito en la Tabla 1 y la Tabla 2. Son datos reales de un estudio piloto (nunca publicado) realizado en ratas para medir el cambio en la presión arterial media tras la administración de un cierto tratamiento. La utilidad de este estudio fue la estimación de la variabilidad muestral en las condiciones del experimento, la cual se desconocía. La diferencia entre las variabilidades de los dos grupos experimentales no fue estadísticamente significativa, así que puede aceptarse que la desviación típica común es igual a la total =8,11, y su varianza ^{2}=65,77. Asimismo, dado que el tratamiento consistió en la administración de una sustancia de referencia, de la cual se conocía su efecto, se determinó que un efecto de interés sería obtener al menos una diferencia tan grande como la observada en el estudio piloto, $_1$- $_2$≈8 mmHg.

Con estos datos se planteó cuál debería ser el tamaño muestral de un estudio en que se comparase el control con otro tratamiento, y se quisiese detectar una tamaño de efecto igual a 8, asumiendo una varianza muestral igual a 65,77, con un nivel de significación igual a 5%, y una potencia estadística del 80%. Si se entran estos datos en el formulario se obtiene que el tamaño muestral necesario es N=17 animales por grupo experimental.

Tabla 1. *Resultados de un estudio piloto en que se compara el cambio en la presión arterial media (PAM) a los 10 minutos tras la administración del tratamiento, respecto al valor basal, en ratas.*

	PAM, cambio en 10 min (mmHg)		
Grupo Exp.	Media	N	Desv. típ.
Vehículo	3,54	3	8,28
Test	11,60	6	7,20
Total	8,92	9	8,11

Tabla 2. *Resultados de la prueba de la t de Student para comparar el cambio en la PAM en los dos grupos experimentales.*

t		-1,515
gl		7
Sig. (bilateral)		,174
Diferencia de medias		-8,063
Error típ. de la diferencia		5,322
95% Intervalo de confianza para la diferencia	Inferior	-20,647
	Superior	4,521

Discusión

El ejemplo presentado en los resultados es bastante trivial, pero permite empezar la discusión sobre como puede reducirse el número de animales de experimentación. En primer lugar, debe quedar clara la relación que existe entre estos cuatro factores (tamaño de efecto, nivel de significación, potencia estadística y variabilidad) y el tamaño muestral. Una vez hechos los cálculos y decidido qué hacer y cómo, y con cuántos, es como si se hubiese construido una mesa con cinco patas capaz de mantenerse perfectamente estable. La posterior alteración la altura de cualquiera de las cinco patas hará que la mesa cojee. La potencia estadística y la variabilidad de la muestra están directamente relacionados con el tamaño muestral necesario. Por el contrario, el nivel de significación y el tamaño del efecto están inversamente relacionados con el tamaño muestral. Fijados tres de los cuatro parámetros, el cuarto determina el tamaño muestral y entonces la mesa es estable. Alterar cualquiera de los parámetros sin ajustar el resto da como resultado un experimento mal planificado, una mesa coja.

Existe una decisión importante que debe tomarse al inicio de la planificación, y es la referente a la hipótesis de trabajo. Es importante porque ya se ha visto que según cuál sea la hipótesis de trabajo, la fórmula de cálculo del tamaño muestral puede variar y por tanto también el resultado del cálculo. Grosso modo, existen dos tipos de hipótesis alternativas: las denominadas bilaterales, y las unilaterales [3]. Una hipótesis alternativa bilateral, en cuanto a la diferencia entre dos medias se refiere, especifica que se espera que esta diferencia sea distinta de cero, sin importar que sea positiva, o negativa. De ahí la denominación de bilateral. En el experimento del ejemplo se ha optado por una primera aproximación que sería la de realizar una prueba bilateral. Esto puede ser razonable cuando se desconoce el posible efecto del tratamiento. Pero cuando se tiene una idea cierta de cuál podría ser ese efecto o de cuál sería el sentido de las diferencias entre grupos experimentales que tendría sentido biológico, es muy aconsejable plantearse realizar una prueba unilateral. En efecto, se sabe que el tratamiento incrementa el cambio de la PAM respecto al grupo Control, y por tanto cuando se investigue un nuevo producto va a resultar seguramente interesante saber si mejora el resultado del

tratamiento estándar con que se ha realizado el estudio piloto. Por tanto, el efecto que se quiere demostrar que existe es que la diferencia entre las medias del grupo tratado y el grupo control sea superior a 8 mmHg. La hipótesis alternativa será que $_2$- $_1$>0, y la hipótesis nula será que $_2$- $_1$ 0. Una prueba unilateral es más potente que la correspondiente bilateral. Puede comprobarse que, si se efectúa el cálculo, el resultado es N=13 animales por grupo experimental, para detectar un efecto superior 8 mmHg, con una variabilidad ^{2}=65,77, un nivel de significación del 5% y una potencia estadística del 80% (para quien quiera reproducir el cálculo, deberá doblarse el nivel de significación unilateral deseado al introducir los parámetros en el formulario).

El parámetro de la variabilidad muchas veces es el más difícil de estimar. Debe realizarse el esfuerzo de encontrar el dato en trabajos semejantes publicados con anterioridad. Si ello no es posible, entonces, deberá plantearse la realización de un estudio piloto, con pocos animales, como el que hemos descrito en el ejemplo, con la finalidad de obtener esta estimación de la varianza.

Y una vez conocida la variabilidad muestral, existe otra manera de reducir el tamaño muestral necesario, que es precisamente reducir dicha variabilidad. Ello puede conseguirse de diversas maneras. Si el dato que se va a recopilar depende, por ejemplo, de una intervención sobre el animal, ésta deberá estandarizarse al máximo para evitar que variaciones en la condición inicial del animal provoquen una mayor variabilidad en la respuesta. También puede ser útil plantearse la utilización de una cepa que tenga una menor variabilidad intrínseca (por ejemplo, cepas consanguíneas) [4]. En este aspecto, la creatividad y la pericia del investigador, deben hacerse notar para mejorar la calidad de su reactivo biológico.

Conclusiones

El investigador debe plantearse qué quiere encontrar, cómo quiere hacerlo, qué riesgos de equivocarse puede asumir, y conocer e intentar reducir la variabilidad del material biológico con que va a trabajar. De esta manera podrá calcular el número de animales óptimo para lograr sus objetivos de investigación.

Invertir en la planificación y el diseño experimental es una garantía de que la información que se vaya a obtener será suficiente, y de calidad. Esta inversión es la que hace surgir el orden e invierte la tendencia termodinámica hacia el caos.

Apéndice

Recursos en línea gratuitos que pueden utilizarse para el cálculo de tamaño muestral.

No existe un recurso que cubra todas las necesidades. Muchos se complementan entre sí. Y es una buena práctica efectuar los cálculos en más de uno, para contrastar resultados.

-Sample Size Estimation: http://www.cct.cuhk.edu.hk/stat/Means.htm

Es el utilizado para trabajar el ejemplo que se ha utilizado en este artículo. Es muy sencillo, y cubre muchas posibilidades de diseño e hipótesis distintas. Bien documentado.

-G*Power: Statistical Power Analyses for Windows and Mac [5]: http://www.gpower.hhu.de

Un programa que hay que descargar e instalar en el ordenador. Muy completo, y bien documentado, en parte, ya que el documento de ayuda está sin terminar.

-PS: Power and Sample Size Calculation: http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize

Un programa muy sencillo, que tiene la ventaja que los resultados generan un texto en inglés que puede ser usado directamente como justificación del tamaño muestral.

-Java applets for power and sample size [6] http://homepage.stat.uiowa.edu/~rlenth/Power/

Una colección de programas en Java que pueden ser ejecutados desde el navegador de Internet o bien descargados y ejecutados directamente desde el ordenador. Muy completo, y bien documentado.

Bibliografía

1. Morowitz, H. J. (1978). Entropía para biólogos. Ed. Hermann Blume. Madrid
2. Hopkins, W. G. (2006). Estimating Sample Size for Magnitude-Based Inferences. Sportscience 10:63-70
3. One-tail vs. two-tail P values. En GraphPad Statistics Guide http://www.graphpad.com/guides/prism/6/statistics/index.htm?one-tail_vs__two-tail_p_values.htm. Consultado el 9 de noviembre de 2014.
4. Festing, M. F. W. et al (2002). The Design of Animal Experiments: Reducing the use of animals in research through better experimental design. SAGE Publications Ltd. Londres.
5. Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39:175-191.
6. Lenth, R. V. (2006-9). Java Applets for Power and Sample Size [Computer software]. Consultado el día 9 de noviembre de 2014, desde http://www.stat.uiowa.edu/~rlenth/Power.