



FORMACIÓN MÉDICA

Yo soy Espartaco. Medida del tamaño de efecto basado en la diferencia de medias.

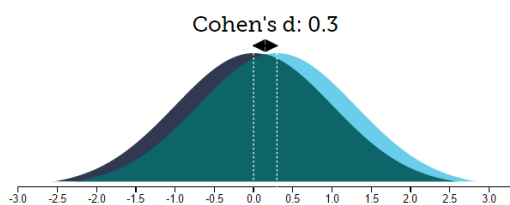
Molina Arias M.

Hospital Infantil Universitario La Paz. Madrid. España.

Resumen

En los estudios que comparan dos intervenciones es útil estudiar el tamaño del efecto que una intervención produce respecto a la otra. La diferencia entre los dos efectos no es útil y es necesario estandarizarla para obtener parámetros más robustos y comparables. Nos centramos en algunos de estos parámetros que nos ayudan a estimar el tamaño del efecto basado en la diferencia de dos medias.

Introducción



En los estudios que comparan dos intervenciones es útil estudiar el tamaño del efecto que una intervención produce respecto a la otra. La diferencia entre los dos efectos no es útil y es necesario estandarizarla para obtener parámetros más robustos y comparables. Nos centramos en algunos de estos parámetros que nos ayudan a estimar el tamaño del efecto basado en la diferencia de dos medias.

Me encontraba yo pensando cómo saber cuándo un efecto es realmente grande y, por aquello de la asociación de ideas, me ha venido a la cabeza alguien grande que, tristemente, nos ha dejado recientemente. Me estoy refiriendo a Kirk Douglas, ese pedazo de actor que siempre recordaré por sus papeles como vikingo, como Van Gogh o como

Espartaco, en la famosa escena de la película en que todos los esclavos, al estilo de nuestro español Fuenteovejuna, se levantan y proclaman ser Espartaco para que no puedan hacerle nada al verdadero (o para que se los fumiguen a todos por igual, mucho más típico del *modus operandi* de los romanos de aquel tiempo).

No me diréis que el tío no era grande. Pero ¿cuánto de grande si lo comparamos con otros? ¿Cómo podemos medirlo? Está claro que no por el número de Oscars, ya que eso solo serviría para medir la miopía prolongada de los llamados académicos del cine, que tardaron lo suyo hasta que le concedieron el premio honorífico por toda su carrera. No es nada fácil encontrar un parámetro que nos defina la grandeza de un personaje como Issur Danielovitch Demsky, que así es como se llamaba el hijo del traperero antes de convertirse en leyenda.

Nosotros lo tenemos más fácil para cuantificar el tamaño del efecto en nuestros estudios, aunque la verdad es que los investigadores suelen estar más interesados en contarnos la significación

estadística que en el tamaño del efecto. Es tan poco habitual calcularlo que, incluso, muchos paquetes estadísticos olvidan contar con rutinas para poder obtenerlo. Nosotros vamos a centrarnos hoy en la forma de medir el tamaño del efecto cuando comparamos dos medias.

Imaginemos que queremos hacer un ensayo para comparar el efecto de un nuevo tratamiento frente al placebo y que vamos a medir el resultado con una variable cuantitativa X. Lo que haremos es calcular la media de efecto entre participantes del grupo experimental o de intervención y la compararemos con la media de los participantes del grupo control. Así, el tamaño del efecto de la intervención respecto al placebo se verá representado por la magnitud de la diferencia entre la media en el grupo experimental y la del grupo control:

$$d = \bar{x}_e - \bar{x}_c$$

Sin embargo, aunque es lo más sencillo de calcular, este valor no nos sirve para hacernos una idea del tamaño del efecto, ya que su magnitud va a depender de varios factores, como la unidad de medida de la variable. Pensemos cómo cambian las diferencias si una media es el doble de la otra según valgan 1 y 2 o 0,001 y 0,002. Para que esta diferencia pueda sernos útil es necesario estandarizarla, así que un señor llamado Gene Glass pensó que podía hacerlo dividiéndola por la desviación estándar del grupo control. Obtuvo así la conocida delta de Glass, que se calcula según la siguiente fórmula:

$$d = \frac{\bar{x}_e - \bar{x}_c}{S_c}$$

Ahora bien, como lo que queremos es hacer una estimación de cuánto valdría el valor de delta en la población, deberemos calcular la desviación estándar utilizando n-1 en el denominador en lugar de n, ya que sabemos que esta cuasivarianza es un mejor estimador del valor poblacional de la desviación:

$$S_c = \sqrt{\frac{\sum_{i=1}^{n_c} (x_{ic} - \bar{x}_c)^2}{n_c - 1}}$$

Pero no os dejéis impresionar por delta, no es más que una puntuación Z (las que se obtienen restando el valor menor su media y dividiéndolo por la desviación estándar): cada unidad del valor de delta equivale a una desviación estándar, por lo que representa la diferencia estandarizada del efecto que se produce entre los dos grupos por efecto de la intervención. Este valor nos permite estimar el porcentaje de superioridad del efecto calculando el área bajo la curva de la normal estándar N (0,1) para un valor de delta (equivale a la desviación estándar) determinado. Por ejemplo, podemos calcular el área que corresponde a un valor de delta = 1,3. Nada más sencillo que utilizar una tabla de valores de la distribución normal estándar o, incluso mejor, la función pnorm () de R, que nos devuelve el valor 0,90. Esto quiere decir que el efecto en el grupo de intervención supera en un 90% el efecto en el grupo control.

El problema de la delta de Glass es que la diferencia de medias depende de la

variabilidad entre los dos grupos, lo que hace que sea sensible a estas diferencias de varianzas. Si las varianzas de los dos grupos son muy diferentes, el valor de delta puede resultar sesgado. Por eso un tal Larry Vernon Hedges quiso contribuir con su propia letra a este particular abecedario y decidió hacer el cálculo de Glass de forma similar, pero utilizando una varianza unificada que no asuma la igualdad de las mismas, según la fórmula siguiente:

$$S_u = \sqrt{\frac{(n_e - 1)S_e^2 + (n_c - 1)S_c^2}{n_e + n_c - 2}}$$

Si sustituimos la varianza del grupo control de la fórmula de la delta de Glass por esta varianza unificada obtendremos la denominada g de Hedges. La ventaja de utilizar esta desviación estándar unificada es que tiene en cuenta las varianzas y los tamaños de los dos grupos, por lo que la g tiene menos riesgo de sesgo que la delta cuando no podemos asumir igualdad de varianzas entre los dos grupos.

De todas formas, tanto la delta como la g tienen un sesgo positivo, lo que quiere decir que tienden a sobreestimar el tamaño del efecto. Para evitar esto, Hedges modificó el cálculo de su parámetro para poder obtener así una g ajustada, según la fórmula siguiente:

$$g_a = g \left(1 - \frac{3}{4gl - 9} \right)$$

donde gl son los grados de libertad, que se calculan como $n_e + n_c$.

Esta corrección es más necesaria con muestras pequeñas (pocos grados de libertad). Es lógico, si miramos la

fórmula, a más grados de libertad, menos necesario será corregir el sesgo.

Hasta ahora hemos tratado de solucionar el problema de calcular un estimador del tamaño del efecto que no esté sesgado por la falta de igualdad de varianzas. El asunto es que, en el rígido y controlado mundo de los ensayos clínicos, lo habitual es que podamos asumir la igualdad de varianzas entre los grupos de las dos ramas del estudio. Podríamos pensar, pues, que si esto se cumple no sería necesario recurrir a los artificios del n-1.

Pues esto mismo pensó Jacob Cohen, así que ideó su propio parámetro, la d de Cohen. Esta d de Cohen es parecida a la g de Hedges, pero todavía más sensible a la desigualdad de varianzas, por lo que solo la usaremos cuando podamos asumir la igualdad de varianzas entre los dos grupos. Su cálculo es idéntico al de la g de Hedges, pero empleando n en lugar de n-1 para obtener la varianza unificada.

Para andar por casa, podemos decir que el tamaño del efecto es pequeño para $d=0,2$, medio para $d=0,5$, grande para $d=0,8$ y muy grande para $d=1,20$. Además, podemos establecer una relación entre d y el coeficiente de correlación de Pearson (r), que también es una medida muy utilizada para estimar el tamaño del efecto.

El coeficiente de correlación nos mide la relación entre una variable independiente binaria (intervención o control) y la variable dependiente numérica (nuestra X). La gran ventaja de esta medida es que es más sencilla de interpretar que los parámetros que hemos visto hasta ahora, que funcionan todos como puntuaciones Z estandarizadas. Ya sabemos que r puede valer de -1 a 1 y el significado de estos valores.

Así, si queréis calcular r a partir de d, no tenéis más que aplicar la siguiente fórmula:

$$r = \frac{d}{\sqrt{d^2 + (1/pq)}}$$

siendo p y q las proporciones de sujetos de los grupos experimental y control ($p=n_e/n$ y $q=n_c/n$). En general, cuanto mayor tamaño de efecto, mayor r y viceversa (aunque hay que tener en cuenta que r también es menor al aumentar la diferencia entre p y q). De todas formas, el factor que más condiciona el valor de r es el valor de d.

Y con esto vamos a terminar por hoy. No creáis que hemos tratado todas las medidas de esta familia. Hay cerca de un centenar de parámetros para estimar el tamaño del efecto, como el coeficiente de determinación, la eta-

cuadrado, la ji-cuadrado, etc., incluso otras que inventó el propio Cohen (no satisfecho solo con la d), como la f-cuadrado o la q de Cohen. Pero esa es otra historia...

Bibliografía

– Ledesma R, Macbeth G, Cortada de Kohan N. Tamaño del efecto: revisión teórica y aplicaciones con el sistema estadístico vista. Rev Latinoam Psicol.2008;40:425-39. ([PDF](#))

– What Works Clearinghouse. Procedures Handbook, Versión 4.0. Disponible en https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf

Correspondencia al autor

Manuel Molina Arias
mma1961@gmail.com
Servicio de Gastroenterología.
Hospital Infantil Universitario La Paz.
Madrid. España.

Aceptado para el blog en febrero de 2020