Sobre la acotación de efectos y la existencia de relación

Vicente Manzano-Arrondo, 2009

El “tamaño de efecto” se corresponde con el contenido de un concepto técnico que pertenece a la jerga estadística. Se utiliza en conjunción con la potencia de la prueba o su valor complementario, el error tipo II, de segunda especie o β (probabilidad de mantener la hipótesis nula cuando habría que rechazarla).

Aunque el tamaño de efecto es un concepto ideado para calcular la potencia de una prueba, tiene mucho sentido para acometer otra tarea fundamental: interpretar el valor encontrado en la muestra sin necesidad de acudir a la significación estadística, es decir, facilitar una significación conceptual. En este sentido, el asunto consiste en observar el tamaño del efecto y concluir, antes de seguir con la significación estadística, si ese efecto es indicativo de relación entre las variables o de suficiente diferencia entre los términos que se comparan. Para interpretar el resultado en la muestra no es necesario estandarizarlo, mientras que para calcular la potencia sí. Así que no vamos a hablar de *tamaño de efecto estandarizado* sino sencillamente de *efecto*.

Un ejemplo: la correlación lineal simple de Pearson (*r*) entre las variables “Nivel de acuerdo con la actuación del gobierno” y “Número de años leyendo prensa” es 0,1. Si la muestra es suficientemente grande, rechazaremos la hipótesis nula (que afirma que el valor de la correlación en la población es 0,000) y concluiremos que hay relación. Pero observemos que 0,1 es *muy poco*. El cuadrado de la correlación indica la proporción de variación compartida. En este caso es de 0,01 (¡sólo un 1%!). En otras palabras: ambas variables son claramente independientes, aunque concluyamos erróneamente que están relacionadas.

Para prevenir este comportamiento, habría que interpretar *antes de pasar a la prueba de significación* el valor concreto de *r* o de cualquier índice. A eso nos referimos cuando decimos que hay que detenerse a valorar el efecto. En los cálculos sobre la potencia de una prueba se acude a tamaños de efecto *estandarizados*. Ya he mencionado que no vamos a hacerlo aquí. Nos preocupa contar con versiones acotadas de índices, es decir, índices que tengan un mínimo y un máximo conocidos, para interpretar su cuantía concreta en cada caso. Si el valor está muy cercano al mínimo, diremos que no hay efecto o que el efecto es pequeño. Si el valor está muy cercano al máximo, diremos que el efecto es importante. Así pues, una vez contemos con esas cotas (mínimo y máximo), podremos interpretar.

Lo que vamos a hacer ahora es una pequeña *incorrección* o *injusticia conceptual* que consiste en facilitar unas recetas para catalogar a un efecto de nulo, pequeño, mediano o grande. Esta estrategia está orientada a reducir la ansiedad o la sensación de indefensión en el examen y situaciones similares. Pero no hay que tomárselo al pie de la letra en las situaciones reales, aplicadas o fuera de los muros del aula. Siempre es conveniente detenerse a interpretar el efecto, pero habitualmente no será conveniente ceñirse a unos intervalos cerrados como los que vamos a utilizar aquí.

Todos los índices están acotados en el intervalo (0,1). Luego, todos se interpretarán del mismo modo: cuanto más cercano a 0, menos efecto; cuanto más cercano a 1, más efecto. Para cerrar unos intervalos vamos a basarnos en la propuesta de Cohen (1988) respecto al coeficiente de correlación de Pearson (*r*). Cohen propuso que en torno a r=0,1 representa un efecto pequeño; en torno a 0,3 es mediano; y en torno a 0,5 es grande. Personalmente pienso que son cotas muy generosas. La *r* es un índice muy sensible. Podemos ver un diagrama de dispersión que no dice nada y obtener en cambio un valor

r=0,3. Pero dado que las acotaciones buscan reducir la ansiedad y orientar vagamente, pero no sustituir el sentido común ni el conocimiento de experto sobre la situación, y dado también que esas acotaciones son las más aceptadas, utilizadas y extendidas, pues no vamos a ponernos muy trascendentes con la cuestión y tomamos a Cohen como inspiración no literal. Va a ser no literal porque en lugar de “en torno a”, vamos a considerar esos valores como cotas máximas y afirmar que si B es el valor del índice concreto (en este caso, B=r), entonces las cotas serán:

Efecto nulo	$B \leq 0{,}1$
Efecto pequeño	$0{,}1 < B \leq 0{,}3$
Efecto mediano	$0{,}3 < B \leq 0{,}5$
Efecto grande	$B > 0{,}5$

Correlación de Pearson (y rho de Spearman)

Se encuentra acotada en el intervalo (-1,+1). Por ello vamos a considerar únicamente su valor absoluto. Luego, B = |r|.

Dado que la *rho* de Spearman es una r de Pearson aplicada sobre los rangos de las dos variables que se correlacionan, pero sigue operando con los mismos principios, entonces se le aplican también los mismos valores para la interpretación.

t de Student

Hay muchas opciones circulando para calcular un tamaño de efecto asociado a una diferencia entre medias. De todas ellas vamos a utilizar la que traduce t a una r de Pearson, por lo que la interpretación de su efecto sigue el mismo esquema expresado para la r en el apartado anterior.

$$r=\frac{t}{\sqrt{t^2+n-2}}$$

donde n es el número de casos (n=n) para la estrategia longitudinal, o la suma de los tamaños de los dos grupos ($n=n_1+n_2$) , en la estrategia transversal.

Chi cuadrado de Pearson

χ^2 está acotada sólo por debajo. Su valor se mueve desde 0 hasta un límite superior que depende de cada situación. En concreto, depende del tamaño de la muestra y del número de categorías que se estén manejando. Por esta razón es difícil interpretar un efecto observando solo el valor de la χ^2. Para corregirlo, realizamos una recodificación de este índice acudiendo a una estrategia denominada V de Cramer.

La V de Cramer se encuentra acotada en el intervalo (0,1). Es un índice muy conservador, poco sensible a los efectos de relación. Acercarse a 1 es muy difícil. No obstante, no vamos a generar una tabla con valores aun inferiores a los límites establecidos para r, así que partimos de los mismos. Luego, B=V.

Para calcular la V de Cramer consulta el anexo I.

U de Mann-Whitney

Este índice compara la media de rangos de la dependiente u ordinal en cada uno de los dos grupos de la independiente o nominal.

Si *d* es la diferencia (en términos absolutos) entre las dos medias de rangos y *n* es el tamaño de la muestra, entonces un índice acotado en (0,1) es:

$$I = 2d/n$$

Luego, B = I. Para acceder a una justificación, consulta el anexo II.

t de Wilcoxon

Este índice recurre a los rangos de las diferencias para datos apareados. Lo que hace es calcular la diferencia, dentro de cada caso, entre su puntuación en una ocasión y la siguiente. Después ordena las diferencias en términos absolutos y compara la media de rangos de las diferencias negativas o la media de las positivas.

Si *d* es la diferencia (en términos absolutos) entre las dos medias de rangos y *n* es el tamaño de la muestra, entonces un índice acotado en (0,1) es:

$$I = 2d/(n+1)$$

Luego, B = I. Para acceder a una justificación, consulta el anexo VI.

H de Kruskal-Wallis

H también está basada en rangos y sigue una distribución χ^2. Ya sabemos que la chi cuadrado es difícil de interpretar y que acudimos a la *V* de Cramer para ello. Luego, interpretar la *H* de Kruskal-Wallis pasa por traducirla a una *V*, considerando como número de categorías el número de grupos.

F de Friedman

Es en todo similar a la *H* de Kruskal-Wallis salvo por el hecho de que se aplica cuando la nominal es longitudinal en lugar de transversal. El resto le es de aplicación directa, por lo que las operaciones e interpretación mediante la *V* de Cramer son las mismas.

Anexos

Anexo I. Cálculo de la V de Cramer

La V de Cramer es una recodificación de la χ^2 de Pearson. Dado que la χ^2 tiene una cota mínima (0) pero no máxima constante, la V consigue restringirla al intervalo (0,1). El efecto final es el de un índice que dificulta sensiblemente llegar a 1, aunque es viable conseguirlo. El cálculo es el siguiente, donde *n* es el tamaño de la muestra y *k* el número de categorías de la variable que se está manejando. Si se trata de una tabla de contingencia, *k* es el número de categorías de la variable que tiene menos categorías.

$$V=\sqrt{\frac{\chi^2}{n(k-1)}}$$

Anexo II. Cálculo de la recodificación para la U de Mann-Whitney

El valor mínimo para *d* es 0, que se obtiene cuando ambos grupos tienen la misma posición media y, por tanto, no hay relación. La máxima diferencia es la mitad del número de datos, es decir *n*/2 (para saber por qué, consulta el anexo III). Por ejemplo, si el tamaño de la muestra es 60, la máxima diferencia que puede tener lugar entre ambos promedios de rangos es 30. Por lo tanto, para transformar el intervalo (0,*n*/2) en (0,1), hay que dividir *d* entre *n*/2 y:

$$I=\frac{d}{n/2}=\frac{2\mathrm{d}}{n}$$

Anexo III. Por qué la máxima diferencia en *U* es *n*/2

Hemos medido dos variables en un conjunto de *n* casos. Una es nominal, con dos categorías, *A* y *B*. La otra es al menos ordinal.

Ordenamos los *n* datos de la variable al menos ordinal, desde el de menor valor hasta el mayor. Ahora contamos las posiciones en las que se encuentran los n_A datos del conjunto *A*. Y hacemos lo mismo con los $n_B = n\text{-}n_A$ del conjunto *B*. Las posiciones que ocupan son sus rangos.

En un caso de máxima relación entre ambas variables, uno de los dos subgrupos se encontrará amontonado en las primeras posiciones y el otro en las últimas, de tal forma que todos los datos de uno de los subgrupos son superiores a todos los datos del otro. Para seguir, supongamos que el que ocupa las posiciones inferiores es el *A*, aunque el resultado es idéntico si suponemos que es el subgrupo *B* el que se encuentra en la zona de valores inferiores.

La suma de rangos de *A* será la suma de los n_A primeros números, es decir (ver anexo IV):

$$Suma\ de\ A=\frac{n_A(n_A+1)}{2} \quad Media\ de\ A=\frac{n_A(n_A+1)}{2n_A}=\frac{n_A+1}{2}$$

La suma de rangos de *B* será la suma de los n_B números que van desde n_A+1 hasta *n*, es decir (ver anexo V):

$$Suma\ de\ B=n_A\, n_B+\frac{n_B(n_B+1)}{2} \quad Media\ de\ B=\frac{n_A\, n_B}{n_B}+\frac{n_B(n_B+1)}{2n_B}=\frac{2n_A+n_B+1}{2}=\frac{n_A+n+1}{2}$$

La diferencia entre ambos es, por lo tanto:

$$d = Media\ de\ B - Media\ de\ A = \frac{n_A + n + 1}{2} - \frac{n_A + 1}{2} = \frac{n}{2}$$

Anexo IV. Suma de los primeros *n* rangos

Para calcular cuál es la suma de los primeros *n* números naturales, piensa en que sumas el primero (1) y el último (*n*) y obtienes *n*+1. Ahora suma el segundo (2) y el penúltimo (*n*-1). Observa que obtienes lo mismo, n+1. Sigues así hasta que llegues al centro, lo que implica obtener la misma suma en *n*/2 ocasiones. Luego, la suma total será *n*/2 veces *n*+1, es decir *n*(*n*+1)/2.

Ejemplo:

La suma de los primeros 10 números es:

1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 =
= 1+10 + 2+9 + 3+8 + 4+7 + 5+6 =
= 11 + 11 + 11 + 11 + 11 + 11 = 5·11 =
= 10 (10+1)/2 = 55

Anexo V. Suma de los rangos que van desde *a* hasta *b*

La suma de *n* números naturales correlativos que van desde *a* hasta *b* (por lo que *b*=*a*+*n*-1, o bien *n*=*b*-*a*+1) puede calcularse desde la expresión de la suma de los primeros *n* números.

Imagina el listado de números desde 1 hasta *n*. Súmales a todos *a*-1. Entonces el listado irá desde *a* hasta *a*+*n*-1 = *b*. Como hemos sumado *n* veces la cantidad *a*-1, entonces, la suma será

$$Suma\ desde\ a\ hasta\ b = a+n-1 = \sum_{i=a}^{a+n-1} i = n(a-1) + \frac{n(n+1)}{2} = \frac{(b+a)(b-a+1)}{2}$$

Anexo VI. Por qué la máxima diferencia en *t* es (*n*+1)/2

La diferencia mínima será, como siempre, 0. La máxima ocurrirá cuando los valores de una de las dos ocasiones sean siempre superiores a los de la otra ocasión. Como consecuencia: todas las diferencias tendrán el mismo signo. Así que la suma de rangos de las diferencias con ese signo será la suma de los *n* rangos, mientras que la suma del otro signo será 0. Luego, la diferencia entre las medias de ambos será la media total de rangos:

$$Suma\ de\ n\ rangos = \frac{n(n+1)}{2} \qquad Media = \frac{Suma}{n} = \frac{n+1}{2}$$

(Para una justificación sobre la suma de n rangos, consulta el anexo IV).

Así pues, como los rangos de las diferencias pueden ir desde 0 hasta (n+1)/2, para que se muevan en el intervalo de 0 a 1 será necesario dividir la diferencia de rangos entre (n+1)/2, es decir:

$$I = \frac{d}{\frac{n+1}{2}} = \frac{2d}{n+1}$$

Referencias

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. New Jersey: Lawrence Erlbaum.