

Sobre la acotación de efectos y la existencia de relación

Vicente Manzano – 2009

El concepto “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* 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 tamaño del 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. Una vez contemos con esas cotas, podemos interpretar.

Ya nos hemos enfrentado a ello en términos muy conceptuales y cualitativos, considerando la complejidad de cada situación. 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. Lo haremos con los índices que hemos utilizado en clase.

Correlación de Pearson

Se encuentra acotada en el intervalo $(-1,+1)$. Vamos a considerar únicamente su valor absoluto. Para interpretarlo partimos de la porción de variación compartida (r^2). Podemos situar el 10% de variación compartida como un mínimo para afirmar que dos variables tienen algo que ver entre sí. Ello se corresponde con $r = 0,32$. El 90% restante lo repartimos en tres tramos iguales. Luego:

Efecto nulo ($r^2=0,10$)	→ $ r < 0,32$
Efecto pequeño ($r^2=0,40$)	→ $0,32 \leq r < 0,63$
Efecto mediano ($r^2=0,70$)	→ $0,63 \leq r < 0,84$
Efecto grande	→ $ r \geq 0,84$

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.

Chi cuadrado de Pearson

La Chi cuadrado de Pearson o χ^2 se utiliza en situaciones donde queremos comprobar en qué medida una repartición de frecuencias observadas es diferente a la repartición que se esperaba. Se utiliza, entre otras ocasiones, para observar relación entre dos variables nominales a través de una tabla de contingencia. En tal caso, se comparan las frecuencias observadas con las que cabría esperar si ambas variables fueran absolutamente independientes.

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 sólo 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. La experiencia con tablas de contingencia nos ha mostrado que situaciones con una relación muy clara observando las casillas, arroja valores para la *V* difícilmente superiores a 0,5. Haremos, entonces, una repartición más atrevida. Para ello, realizamos una repartición equitativa inicial $(0,25;0,50;0,75)$ y le aplicamos una corrección, elevando esta cantidad a 1,5 para conseguir un resultado que corrija en algo la dificultad de *V* para mostrar relación. De este modo:

Efecto nulo ($V^{1/1.5}=0,25$) $\rightarrow V < 0,125$
Efecto pequeño ($V^{1/1.5}=0,50$) $\rightarrow 0,125 \leq V < 0,35$
Efecto mediano ($V^{1/1.5}=0,75$) $\rightarrow 0,35 \leq V < 0,65$
Efecto grande $\rightarrow V \geq 0,65$

Para calcular la V de Cramer consulta el anexo I

U de Mann-Whitney

Este índice se encarga de tomar la decisión sobre si hay relación entre una variable nominal transversal de dos niveles y una al menos ordinal. Recurre a los rangos y compara la media de rangos de la ordinal en cada uno de los dos grupos de la nominal. Dado que compara la media de rangos, es útil sólo para diferencias de valores promedio o representativos, pero no para medir el resto del comportamiento de las variables.

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 (para acceder a una justificación, consulta el anexo II):

$$I = 2d/n$$

Para interpretar su cuantía podemos acudir a una repartición equitativa:

Efecto nulo: $I < 0,25$
Efecto pequeño: $0,25 \leq I < 0,50$
Efecto mediano: $0,50 \leq I < 0,75$
Efecto grande: $I \geq 0,75$

t de Wilcoxon

Este índice se encarga de tomar la decisión sobre si hay relación entre una variable nominal longitudinal de dos niveles y una al menos ordinal. Recurre a los rangos de las diferencias. 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 (para acceder a una justificación, consulta el anexo VI):

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

Para interpretar su cuantía podemos acudir a la misma repartición equitativa que hemos utilizado para el caso de la U de Mann-Whitney.

H de Kruskal-Wallis

Este índice se utiliza cuando queremos observar la relación entre una variable nominal transversal con más de dos niveles y una al menos ordinal. 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. Para interpretar la V consulta el apartado sobre Chi cuadrado.

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:

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

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 $n_B = n - n_A$ 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):

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

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):

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

La diferencia entre ambos es, por lo tanto:

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

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:

$$\begin{aligned} 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 &= \\ &= 1+10 + 2+9 + 3+8 + 4+7 + 5+6 = \\ &= 11 + 11 + 11 + 11 + 11 + 11 = 5 \cdot 11 = \\ &= 10(10+1)/2 = 55 \end{aligned}$$

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

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

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:

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

(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:

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