

¿EL PODER PARA QUÉ?

Esta famosa pregunta de un estadista colombiano de origen tolimense, no se dijo pensando en estadística, aún cuando existe una conexión histórica entre las palabras: estadista y estadística; de hecho, en sus orígenes, la segunda no era nada más que las matemáticas que necesita la primera para gobernar mejor. Este editorial hace referencia a otro poder menos conocido pero muy importante al diseñar estudios de comparación.

Es conocido el valor de la p que se presenta como resultado de realizar una comparación mediante alguna prueba estadística, sea esta una χ^2 , una t de Student, una prueba de Anova o de Kruskal-Wallis o cualquier otra. Todos sabemos que si el valor de la p es menor a 0.05 podemos afirmar que existe una diferencia estadísticamente significativa entre lo que se compara. De hecho esa $p < 0.05$ es la probabilidad que se tiene de equivocarse al decir que existe una diferencia, aunque no la haya, entre lo que se compara, tratamientos, exposiciones o cualquier otro aspecto que se quiera comparar. Debido a que este error nunca será igual a 0 (cero) siempre se comete este tipo de error, lo importante es que no sea muy grande. Esta p es entonces la probabilidad de cometer un error tipo I o error α , que es afirmar que existe una diferencia donde no la hay. Aunque este tipo de error, en general, se usa con un valor de 0.05, según la pregunta de investigación que se quiera responder, puede estar entre 0.01 y 0.1. No conozco estudios que utilicen una p con valor > 0.1 ; además, este valor se debe establecer desde el inicio de la investigación.

Es importante resaltar aquí, que el hecho de que una diferencia sea estadísticamente significativa no garantiza que esta diferencia sea clínicamente importante, pero ese es otro tema al que no haré referencia.

Así como podemos equivocarnos al decir que existe una diferencia, cuando no la hay, también nos podemos equivocar, y esto es muy frecuente, cuando se dice que no existe esa diferencia, pero sí la hay. Cuando esto sucede, se comete otro error, que se conoce como error tipo II o error β . Este error es, entonces, aquel que se puede cometer cuando al realizar la prueba estadística el valor de p obtenido es mayor que el que se decidió utilizar en el estudio, aunque en la realidad si exista una diferencia. Estos estudios, en general, se conocen como “estudios negativos”, es decir, estudios que no encontraron una diferencia entre lo que se comparó.

Este error no es muy mencionado porque, en general, se prefiere hablar de su complemento $1-\beta$, probabilidad que se conoce como el poder, que en España le dicen potencia.

El error tipo II o β que generalmente se usa es ≤ 0.2 , o sea, que el poder, su complemento, es generalmente ≥ 0.8 . En estudios con humanos, de tipo observacional, no se exige un poder > 0.8 , pero en estudios experimentales de evaluación de intervenciones se prefiere un poder ≥ 0.9 , estos valores son sólo una recomendación.

Ilustremos con un par de ejemplos: un estudio muestra que al comparar dos tipos de intervención para disminuir el dolor postoperatorio se presenta una diferencia de 3 puntos en alguna escala para evaluar dolor y que el valor de $p=0.039$. Es fácil concluir, en este caso, que existe una diferencia estadísticamente significativa a favor de uno de los grupos. Ahora bien, si esta diferencia de 3 puntos es importante clínicamente es otro cuento. Cuando se encuentra una diferencia estadísticamente significativa no es importante pensar en el poder, porque, según el principio de Moisés: “*Si el hacha tumbó el árbol era que tenía suficiente filo*”.

Otro estudio, aunque muestra una diferencia de 4 puntos en la misma escala del anterior, presenta una $p=0.11$, es decir, la diferencia no es estadísticamente significativa. Bajo esta situación es importante preguntarse: ¿Con qué poder se calculó el tamaño de la muestra para esta investigación? Esta pregunta es fundamental, porque, la importancia del poder estriba, no en su valor una vez terminado y analizado el estudio, sino en su utilidad para calcular el tamaño de la muestra.

Los insumos necesarios para calcular el tamaño de la muestra en una investigación son: además de los valores de los errores α y β , o su complemento $1-\beta$, o sea el poder, la diferencia o δ que se quiere demostrar y que debe ser, de preferencia, clínicamente importante. Esta diferencia no siempre es una como tal, porque en muchos estudios se utilizan razones, como en el caso del uso del riesgo relativo. En la actualidad a este δ se prefiere llamarlo tamaño del efecto.

La Tabla 1 muestra esquemáticamente la relación entre error α , error β , poder, δ y el tamaño de la muestra N.

Tabla 1. Diferencias entre α , β , δ y N.

| Error tipo I | Error tipo II | Poder | Delta | Tamaño muestra |
|---------------------|--------------------|--------------------|---------------------|----------------|
| $\alpha \downarrow$ | $\beta \downarrow$ | $1-\beta \uparrow$ | $\delta \downarrow$ | $N \uparrow$ |

La Tabla 1 se interpreta de la siguiente manera:

Si se tienen los demás valores constantes, al disminuir el valor de α el tamaño de muestra, N, aumenta. Bajo la misma condición de constancia de los demás valores si se disminuye β , o sea que si aumenta $1-\beta$, el poder, el valor de N aumenta. También sucede lo mismo si se disminuye el tamaño de δ , es decir, si el tamaño del efecto, clínicamente importante, es cada vez menor.

Como se dijo al comienzo, la importancia del poder radica en su uso para el cálculo del tamaño de la muestra, al diseñar una investigación. Es decir, con un poder ≥ 0.8 tenemos esa fuerza para decir que, si no se encuentra una diferencia estadísticamente significativa, es porque la diferencia del tamaño del efecto es menor de la que se usó para realizar el cálculo del tamaño de la muestra y que, por lo tanto, no sería clínicamente importante. La anterior afirmación contesta la pregunta del título de este editorial.

Esta pequeña introducción puede ser complementada con miles de artículos publicados sobre el tema en la mayoría de revistas científicas; además, existen algunos libros muy útiles para una revisión más o menos amplia del tema, entre los que merecen mencionarse:

Kraemer HC, Thiemann S. How Many Subjects? Statistical Power Analysis in Research. Newbury Park: Sage Publications; 1987.

Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd. Edition. New Jersey: Routledge Academic; 1988.

Luis Carlos Orozco Vargas, MD MSc. Epidemiología
 Profesor Asociado Escuela de Enfermería
 Universidad Industrial de Santander
 Correo electrónico: lcorovar@gmail.com