



Onomázein

ISSN: 0717-1285

onomazein@uc.cl

Pontificia Universidad Católica de Chile  
Chile

Quezada, Camilo  
POTENCIA ESTADÍSTICA, SENSIBILIDAD Y TAMAÑO DE EFECTO: ¿UN NUEVO  
CANON PARA LA INVESTIGACIÓN?  
Onomázein, vol. 2, núm. 16, 2007, pp. 159-170  
Pontificia Universidad Católica de Chile  
Santiago, Chile

Disponible en: <http://www.redalyc.org/articulo.oa?id=134516684004>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org

**POTENCIA ESTADÍSTICA, SENSIBILIDAD Y  
TAMAÑO DE EFECTO:  
¿UN NUEVO CANÓN PARA LA INVESTIGACIÓN?**

***STATISTICAL POWER, SENSITIVITY AND SIZE EFFECT:  
A NEW STANDARD FOR RESEARCHERS?***

**Camilo Quezada**

Pontificia Universidad Católica de Chile  
cequezad@uc.cl

To call in the statistician after the experiment is done may be no more than asking him to perform a post mortem examination: he may be able to say what the experiment died of.

(Pedir ayuda a un estadístico una vez que el experimento ya fue realizado puede no ser más que pedirle efectuar una autopsia: posiblemente, lo único que pueda hacer sea decir de qué murió el experimento).

Ronald Fisher

**Resumen**

En este trabajo se presentan algunos conceptos muy importantes en el diseño investigativo, centrándose en tres nociones cada vez más utilizadas en metodología cuantitativa: la *potencia estadística*, la *sensibilidad* y el *tamaño de efecto*. A través de ejemplos de orientación lingüística se explican algunas de las posibles aplicaciones prácticas de estas herramientas en investigaciones pertenecientes al ámbito de las ciencias sociales en general y la lingüística en particular.

**Palabras clave:** potencia estadística, tamaño de efecto, metodología

Fecha de recepción inicial: octubre de 2007

Fecha de aceptación: octubre de 2007

**Abstract**

*This paper discusses some extremely important notions which highly impact research design, focusing on three concepts more and more used in quantitative approaches: statistical power, sensitivity and size effect. By presenting linguistically-oriented examples, explanations are given for the potential uses of these methodological tools both in Social Sciences in general and in Linguistics in particular.*

**Key words:** statistical power, size effect, methodology

Últimamente, son cada vez más frecuentes las investigaciones cuantitativas que incluyen entre sus resultados un índice durante largo tiempo ignorado y omitido: el de la potencia estadística (*statistical power*). Denominado a veces también como “poder”, este índice fue desarrollado conceptualmente durante la primera mitad del siglo XX (cf. Bono y Arnau, 1995, para una muy buena introducción a la historia y los principios de la potencia estadística), pero no fue sino hasta la década de los 60, gracias al trabajo del psicólogo Jacob Cohen, cuando se empezó a sistematizar su aplicación al ámbito de las ciencias sociales en general y la psicología en particular.

La precisión anterior no es gratuita. Tradicionalmente, la psicología y la sociología han sido por lejos las “ciencias” sociales que más han recurrido a la estadística descriptiva y la estadística inferencial para sustentar sus teorías y trabajos. En gran parte ello se debe a que en esas disciplinas la mayoría de las veces se trabaja con unidades claramente delimitadas extraídas de conjuntos particulares a fin de caracterizar conjuntos más generales. De ahí que resulte conceptualmente coherente investigar asumiendo la legitimidad de algunos de los supuestos estadísticos indispensables para trabajar inferencialmente<sup>1</sup>.

Sin embargo, en el ámbito de la lingüística y sus diversas ramificaciones hay varias razones tanto teóricas como conceptuales para afirmar que muchos de los métodos y herramientas de las ciencias exactas no son completamente compatibles con algunos de los problemas que plantea la investigación del lenguaje (estos cuestionamientos pueden ser muchos y muy profundos, pero no viene al caso tratarlos aquí. Para una discusión acerca del impacto que ha tenido el método científico tradicional en las teorías e investigaciones lingüísticas, en especial en lo relativo a la unidad, cf. Quezada, en prensa).

Ahora, nada de lo anterior impide, por supuesto, que exista un gran número de investigaciones cuantitativas en el ámbito de la lingüística, sobre todo en los dominios de la lingüística aplicada. Y, por cierto, no

<sup>1</sup> En especial conceptos como la *distribución normal* y el *teorema del límite central*.

tiene por qué ser de otra manera, puesto que el empleo de metodologías cuantitativas no define (ni para bien ni para mal) la calidad o la pertinencia de una investigación. En último término, las herramientas estadísticas son solo eso: herramientas puestas a disposición de los investigadores para enfrentar un problema determinado. Eso sí, al momento de emprender una investigación de corte cuantitativo conviene tener bien presentes cuáles son los supuestos y convenciones implícita y explícitamente asumidos por el paradigma, a fin de no incurrir en faltas de rigor (o de protocolo investigativo).

Precisamente, la potencia estadística es una convención (aunque sea quizás más preciso hablar de canon) que vale la pena tener presente al momento de emprender una investigación cuantitativa, dada su creciente importancia en el mundo investigativo de las ciencias sociales.

## SENSIBILIDAD Y POTENCIA

Definida de manera simple, la sensibilidad de un diseño investigativo reside en su *capacidad* de detectar diferencias o efectos allí donde los haya. Podemos poner esto en términos más concretos si pensamos en una investigación que estudie algún aspecto relacionado con lenguaje, utilizando la tradicional técnica del contraste de promedios. Supongamos que un grupo de lingüistas, luego de estudiar a fondo la literatura relacionada con esquemas de argumentación, llega a la conclusión de que existe una alta correlación entre el uso de los conectores causales y reformulativos y la legibilidad de los textos<sup>2</sup>. En otras palabras, los lectores tienden a encontrar más claros y comprensibles aquellos textos en los que se utilizan mayor cantidad de conectores causales y reformulativos. Deciden entonces emplear estos conocimientos para establecer si estudiantes universitarios de pregrado que cursan carreras distintas utilizan de manera diferente los dos tipos de conectores examinados.

En una primera etapa, los investigadores reúnen muestras de producciones textuales y miden la cantidad de conectores por cada mil palabras, con lo que obtienen un índice normalizado que permite comparar textos de diferente extensión. Sin embargo, al comparar los

<sup>2</sup> No viene al caso discutir aquí acerca de los supuestos teóricos implicados en la formulación del problema. El hecho mismo de que puedan existir distintas definiciones y propuestas para la clasificación de los conectores y de que muchas veces resulte difícil distinguirlos funcionalmente de marcadores discursivos u operadores pragmáticos no hace sino poner en evidencia el tema de la enorme dificultad teórica que puede enfrentar la lingüística cuando tiene que delimitar las unidades con las que quiere trabajar. Lo mismo vale para el irresuelto debate acerca de la *legibilidad* o *densidad textual*.

niveles de legibilidad asignados por los lectores a los textos observados, se dan cuenta de que algunos textos tienen prácticamente los mismos índices de conectores por cada mil palabras aun cuando obtienen evaluaciones de legibilidad marcadamente distintas. Una revisión más detallada de los textos les permite detectar que la mera presencia de conectores no es por sí sola garantía de una mayor claridad, y que un elemento importante es también la *alternancia*. Dicho de otro modo, a mayor cantidad y variedad de conectores empleados, mayor es la claridad percibida.

A la luz de estos datos desarrollan una exhaustiva grilla en la que se reúnen todos (o casi todos) los conectores causales y reformulativos que pueden ser empleados por los escribientes, lo que permite registrar no solo el índice total de ocurrencias de conectores para cada texto, sino que además desglosar esta cifra para saber cuántas veces se emplearon los *distintos* conectores causales y reformulativos utilizados en cada texto. Esta cifra proporciona un índice de alternancia que luego se utiliza para ajustar el índice bruto total, con lo que finalmente se refleja mucho mejor el nivel de legibilidad percibido por los lectores.

Una vez satisfechos con el sistema de medición, los investigadores proceden a diseñar su estudio. Recordemos que al momento de emprender su aventura investigativa tenían en mente saber si los alumnos universitarios de distintas carreras utilizan o no de manera diferente los conectores causales y reformulativos. De manera que deciden trabajar con tres grupos de estudiantes pertenecientes a las carreras de Ingeniería, Historia y Periodismo (cada grupo consta de 50 estudiantes). Por motivos puramente ficcionales deciden no efectuar una prueba de ANOVA y prefieren utilizar dos pruebas *t*, una para contrastar los promedios de *periodistas* vs. *ingenieros* (Comparación 1, C1) y otra para contrastar los promedios de par *historiadores* vs. *periodistas* (Comparación 2, C2)<sup>3</sup>.

¿Cuál es el rol de la sensibilidad y la potencia en una investigación como esta? Recordemos que la definición propuesta para sensibilidad señalaba que se trata de la capacidad que posee un diseño investigativo para detectar diferencias o efectos allí donde los haya. En términos del estudio aquí imaginado, es muy probable que se encuentren diferencias bastante grandes al efectuar la comparación C1, tal como es muy probable que las diferencias registradas al efectuar la comparación C2 sean bastante bajas. En ambas comparaciones, y siguiendo la convención imperante en la estadística inferencial, se asume que *hipótesis nula* ( $H_0$ ) es aquella según la cual los promedios no son estadísticamente

<sup>3</sup> Las pruebas que emplean la *t* de Student se utilizan para contrastar los promedios de dos grupos. Las pruebas ANOVA se utilizan para comparar los promedios de tres o más grupos.

diferentes (en otras palabras, es aquella situación en la que la intuición inicial de los investigadores no se ve respaldada por los datos), mientras que la *hipótesis alternativa* ( $H_1$ ) es aquella según la cual sí hay diferencias estadísticas significativas entre los promedios de los grupos comparados (o sea, la que le da la razón a los investigadores).

Una vez establecida la importante distinción anterior, es posible proporcionar una definición más técnica de la potencia, que puede ser concebida como *el grado de probabilidad de rechazar estadísticamente la  $H_0$  cuando esta es falsa*, es decir, cuán probable es que los investigadores demuestren estadísticamente que su hipótesis inicial era correcta<sup>4</sup>. En nuestra investigación imaginaria, entonces, la potencia es el grado de probabilidad que nuestros lingüistas tienen de efectivamente detectar estadísticamente diferencias entre los promedios de los grupos estudiados.

## CÓMO SE CALCULA LA POTENCIA

El cálculo de la potencia se efectúa en base a tres cifras: el *n muestral* (la cantidad de elementos de la muestra), el *nivel de error* (denominado también simplemente como  $\alpha$  y generalmente establecido en un nivel de 0,05 o 0,01) y el *tamaño de efecto* (*effect size*). Volviendo a nuestro caso ficticio, contamos con dos de estos datos:

$$\text{Potencia estadística} = \begin{cases} \text{Tamaño muestral: } 50 \text{ (x2)} \\ \text{Nivel de error: } \alpha = 0,05 \text{ (determinado por los investigadores)} \\ \text{Tamaño de efecto: } \text{desconocido} \end{cases}$$

Claramente, el elemento que nos falta para determinar la potencia del estudio es el *tamaño de efecto*. El nombre mismo de este término evoca los orígenes “duros” de las metodologías cuantitativas. Tanto en la psicología como la medicina –y algunas otras disciplinas–, cuando se trabaja experimentalmente se intenta comprobar o poner a prueba la eficacia de tratamientos, medicinas, terapias, etc. El ideal de un experimento puro en psicología o medicina es contar con dos o más grupos de personas elegidas aleatoriamente (grupos que ojalá sean del mismo tamaño). Si los grupos son homogéneos, es decir, si antes del experimento los promedios de cada grupo no muestran diferencias significativas en la dimensión o variable estudiada, se los puede someter a los distintos tratamientos, estableciendo habitualmente un *grupo control* que no recibe terapia alguna o tan solo algún placebo.

---

<sup>4</sup> Para una definición más técnica, véase Cohen, 1992: 156.

Si transcurrido un cierto tiempo se vuelve a medir a los participantes y se constata que presentan discrepancias significativas anteriormente no detectadas, se puede considerar que es razonable atribuir el origen de estas diferencias a la medicina o terapia que hayan recibido los sujetos de cada grupo. Esta diferencia entre los promedios de los grupos recibe el nombre de *tamaño de efecto*, pues entrega una medida de cuán profunda o fructífera fue la intervención, es decir, cuál es la magnitud del efecto del tratamiento.

Cabe notar que, si bien este es probablemente un caso ideal en una investigación experimental clínica, no es de ningún modo la única manera de efectuar una investigación y, lo que es más importante, no es la única manera de concebir el tamaño de efecto. De hecho, la elección del término “efecto” implica ya de por sí la creencia en un modelo empírico experimental que busca demostrar la existencia de relaciones causales (cf. Balluerka & Vergara, 2002). Claramente, un modelo experimental que demuestre la existencia de relaciones direccionales de causalidad entre variables es ideal en el caso de la medicina o la psicología experimental, pero es más difícil de implementar o demostrar en el ámbito de las ciencias del lenguaje.

Concebido de manera más general, entonces, cuando por ejemplo se lo utiliza simplemente para efectuar comparaciones, el tamaño de efecto es un indicador que permite hacerse una idea de *cuán distintos* son dos grupos en una o más variables medidas. En nuestro caso, entonces, el tamaño de efecto vendría a ser el *grado de diferencia* entre los promedios de los grupos estudiados en las dos comparaciones planificadas:

**C1:** periodistas vs. ingenieros

**C2:** historiadores vs. periodistas

Claramente, no necesitamos tener una gran base teórica para sospechar que el tamaño de efecto (la diferencia entre los promedios) va a ser mayor en el caso de la primera comparación que en el caso de la segunda. Aquí es donde las cosas comienzan a ponerse interesantes para una investigación. Cuando las diferencias entre dos grupos son grandes, se hacen visibles con muy pocos casos. Es muy probable que los promedios de C1 muestren tendencias muy distintas al comparar 20 o 25 textos en cada grupo, mientras que, por el contrario, es muy probable que al comparar 20 o 25 textos en C2 no se obtengan diferencias tan evidentes, debido a que los grupos aquí comparados son más parecidos entre sí. En otras palabras, el tamaño de efecto es menor.

De esto se sigue entonces que si el tamaño de efecto (las diferencias entre dos grupos) es grande, se necesitará una muestra más

reducida para detectar las disimilitudes (es decir, para rechazar la  $H_0$  según la cual los promedios de ambos grupos no muestran diferencias estadísticamente significativas). Por el contrario, si el tamaño de efecto es bajo, se necesitará una muestra más grande para que las diferencias, si las hay, se vuelvan visibles y se pueda rechazar la  $H_0$ . Por lo tanto, si la potencia es la probabilidad de rechazar la  $H_0$  cuando esta es falsa, si en dos comparaciones utilizamos el mismo número de sujetos en cada grupo pero el tamaño de efecto es mayor en una de las dos comparaciones, claramente la potencia de ambos contrastes será distinta, puesto que dado el mismo número muestral y el mismo nivel de error y distintos tamaños de efecto, habrá más probabilidades de detectar diferencias (rechazar la  $H_0$ ) en un caso que en otro.

## UN POCO MÁS DE PRECISIÓN

En estricto rigor, la potencia estadística es un equilibrio entre lo que se conoce como error Tipo I y error Tipo II. Un error Tipo I es el ya mencionado  $\alpha$ , y consiste en el margen establecido por el investigador para rechazar la  $H_0$  (se comete un error de Tipo I al afirmar que la  $H_0$  es falsa cuando en realidad es verdadera. En nuestro caso, equivaldría a concluir que los promedios de las poblaciones de las cuales se extrajeron las muestras son estadísticamente diferentes cuando en realidad los parámetros no difieren). Este tipo de error es el más grave y peligroso en cualquier investigación cuantitativa, y por lo tanto el que mayor atención recibe y el que más se controla (por convención, el nivel  $\alpha$  establecido en una investigación cuantitativa es casi siempre igual o menor a 0,05). El error Tipo II, en cambio, es el error que se comete cuando se acepta una  $H_0$  que en realidad es falsa. Es decir, se produce cuando se da por cierta la igualdad de promedios pero en realidad los parámetros de las poblaciones son diferentes. Se lo denomina  $\beta$  y se puede cometer por distintas razones, entre ellas un tamaño muestral bajo (recordemos que cuando el tamaño de efecto es bajo, para detectar estadísticamente las diferencias se necesita un tamaño muestral más alto, o de lo contrario es posible que las diferencias existentes entre las poblaciones pasen desapercibidas).

Convencionalmente, se asume que con un  $\alpha$  de 0,05 y un  $\beta$  de 0,20 se logra un equilibrio conveniente entre estos dos tipos de error (cf. Cohen, 1992). Ahora, como en términos estadísticos la potencia es igual a  $1-\beta$ , el nivel ideal de potencia debiera ser siempre igual o superior a  $1 - 0,2$ , es decir, 0,8. Este es, en una escala que va de 0 a 1, el nivel mínimo de potencia requerido para una investigación cuantitativa según Cohen.

Ahora, una vez establecido este nivel mínimo aceptable, quedan varios problemas todavía por resolver. El principal es el siguiente: ¿cómo podemos saber si nuestro tamaño de efecto es grande, mediano o pequeño? La convención (o canon) vuelve otra vez a la carga. Cohen (1992) estableció una serie de rangos para determinar si el valor de un tamaño de efecto es importante, mediano o menor. Gracias a sus tablas podemos “encasillar” el valor de un tamaño de efecto determinado en tal o cual magnitud. Además, lo que es mucho más importante, estableció criterios precisos y diferenciados para cada una de las principales pruebas estadísticas empleadas en las ciencias sociales (entre ellas las pruebas *t*, las pruebas de ANOVA y las pruebas de correlación y regresión).

De manera tal que en las investigaciones cuantitativas se pueden utilizar pautas claras (pero convencionales y en cierta medida arbitrarias, con todo lo que ello pueda implicar) para trabajar con el concepto de potencia.

## APLICACIONES PRÁCTICAS

La estrecha interrelación de los elementos empleados en el cálculo de la potencia permite efectuar estimaciones sumamente interesantes para cualquier investigación cuantitativa, estimaciones que pueden ser efectuadas *a priori* (antes de realizar la investigación) o *post hoc* (cuando ya se ha efectuado la investigación). Una primera y extremadamente útil aplicación de la potencia estadística tiene que ver con la posibilidad de determinar, *a priori*, el tamaño muestral requerido para que la investigación tenga una potencia aceptable. En otras palabras, podemos calcular (más bien pedirle a algún software<sup>5</sup> que lo haga por nosotros) cuál es el número de sujetos, textos u oraciones que necesitamos para nuestra investigación, puesto que el valor de  $\alpha$  lo determinamos nosotros mismos, la potencia deseada va a ser siempre 0,8 o más, y el tamaño de efecto lo podemos ingresar recurriendo a las tablas de Cohen.

Al aplicar estos criterios en el marco de una hipótesis unidireccional en la que se asume un tamaño de efecto *grande* en la diferencia de promedios entre ingenieros y periodistas, se obtiene lo siguiente:

---

5 Todos los datos numéricos entregados en este trabajo para el cálculo a priori de tamaño muestra y potencia fueron obtenidos con el software estadístico G\*Power3, de distribución gratuita y desarrollado por Franz Faul, de la Universitat Kiel de Alemania, y disponible en <http://www.psycho.uni-duesseldorf.de/abeilungen/aap/gpower3/>

Potencia estadística mínima = 0,8  
Tamaño muestral requerido: **21** (x2)  
Nivel de error:  $\alpha = 0,05$   
Tamaño de efecto: **grande** ( $d = 0,8$ )

Con los mismos datos pero con un tamaño de efecto *mediano* se obtiene lo siguiente:

Potencia estadística mínima = 0,8  
Tamaño muestral requerido: **51** (x2)  
Nivel de error:  $\alpha = 0,05$   
Tamaño de efecto: **mediano** ( $d = 0,5$ )

Por último, al utilizar un tamaño de efecto *pequeño* (el que esperaríamos encontrar al comparar periodistas e historiadores), los resultados son los siguientes:

Potencia estadística mínima = 0,8  
Tamaño muestral requerido: **310** (x2)  
Nivel de error:  $\alpha = 0,05$   
Tamaño de efecto: **pequeño** ( $d = 0,2$ )

Como puede verse, se pasa de un  $n$  total de 42 a uno de 102, y luego a uno de 620. Claramente, no es lo mismo reunir y analizar 42 textos que 620. De modo tal que la potencia estadística puede ser fundamental para determinar el tamaño muestral requerido y por lo tanto definir la viabilidad práctica de la investigación.

Un detalle importante no mencionado hasta aquí es el que tiene que ver con la “elección” del tamaño de efecto. En psicología se emplea con bastante frecuencia una técnica llamada “metaanálisis”. Esta técnica consiste en revisar la literatura investigativa relacionada con un determinado problema para tomar nota de los resultados obtenidos por otros y luego analizarlos. De esta manera, en lugar de basarse en el mero arbitrio personal, es posible estimar el tamaño de efecto a partir de datos concretos obtenidos previamente por otros investigadores. Claramente, esto solo se puede hacer cuando efectivamente existe una cierta cantidad de trabajos empíricos relacionados con el problema investigativo escogido. O sea, cuando existen investigaciones cuantitativas previas en el área y el tema escogidos. De más está decir que en psicología y sociología esto es mucho más frecuente que en la lingüística en general, por lo que en el ámbito de las ciencias del lenguaje muchas veces esta estimación del tamaño de efecto deberá ser efectuada con algún estudio

piloto o simplemente a partir de decisiones basadas en la bibliografía o los conocimientos teóricos de los investigadores.<sup>6</sup>

Otra interesante aplicación práctica tiene que ver con la posibilidad de determinar la potencia y el nivel  $\alpha$  implicados en un estudio. Se trata también de una aplicación *a priori*, y básicamente consiste en contestar la siguiente pregunta: si tengo un  $n$  muestral determinado y sé (o creo) que el tamaño de efecto es tal y tal, ¿cuál sería la potencia y el nivel de error que obtendría al efectuar la investigación en esas condiciones? En nuestra hipotética investigación las preguntas serían estas:

**P1:** si tengo dos grupos de 50 sujetos cada uno y el tamaño de efecto es **grande**, ¿qué potencia y qué nivel  $\alpha$  obtendría al realizar el estudio utilizando una prueba  $t$  para comparar promedios?

**P2:** si tengo dos grupos de 50 sujetos cada uno y el tamaño de efecto es **pequeño**, ¿qué potencia y qué nivel  $\alpha$  obtendría al realizar el estudio utilizando una prueba  $t$  para comparar promedios?

En el primer caso, al ingresar los datos a G\*Power3 se obtiene una potencia de 0,95 y un nivel  $\alpha$  de 0,01. En otras palabras, al efectuar una investigación en estas condiciones las probabilidades de detectar diferencias significativas con un margen de error igual o menor a 0,01 son muy altas (0,95 en una escala de 0 a 1). En el segundo caso, se obtiene un nivel  $\alpha$  de 0,13 (muy superior al  $\alpha$  máximo de 0,05 convencionalmente establecido en la estadística inferencial), con una potencia de 0,45, lo que quiere decir que la probabilidad de demostrar estadísticamente la hipótesis previa de los investigadores es extremadamente baja (0,45 en una escala de 0 a 1).

Puede verse fácilmente que una de las dos posibles investigaciones va a ser mucho más fructífera que la otra, pues en el caso de C1 se llegará efectivamente a resultados estadísticamente significativos, mientras que en la otra no (en el entendido, por supuesto, de que los tamaños de efecto adoptados no estén demasiado alejados de la realidad de los grupos que conforman la muestra). Así, mediante el uso de la potencia estadística se puede saber si vale o no la pena efectuar una investigación antes de embarcarse en ella. En efecto, si el equipo investigador no tiene acceso material a más informantes o no cuenta

<sup>6</sup> Esto de ningún modo invalida el uso de la potencia, pues la estimación “intuitiva” del tamaño de efecto se puede utilizar simplemente para determinar el tamaño de la muestra en la etapa del diseño de la investigación, o para determinar la viabilidad de la investigación. Posteriormente, una vez concluido el estudio, se podrán obtener el tamaño de efecto y la potencia reales de la investigación.

con el tiempo y los recursos requeridos para ampliar la muestra, un análisis *a priori* de la potencia permitirá conocer las probabilidades de que la investigación logre detectar la significación estadística de las diferencias o los efectos que busca establecer.

Por último, al efectuar análisis *post hoc*, es posible obtener los valores reales de tamaño de efecto y por lo tanto calcular la potencia observada del estudio. Estos son datos cada vez más requeridos, pues complementan de manera importante la significación estadística aportada por el nivel  $\alpha$  obtenido en un estudio. Es más, para algunos investigadores la significación estadística del nivel  $\alpha$  no implica la existencia de un efecto real si no está acompañada de un nivel adecuado de potencia. De ahí que sea tan importante incluir esta noción en el diseño de las investigaciones cuantitativas.

## PARA RESUMIR

Lo expuesto hasta ahora representa tan solo una fracción de todas las teorías y fórmulas implicadas en el cálculo de la potencia. Se trata de un resumen bastante simplificado expuesto simplemente para introducir el concepto de potencia estadística. Hay más cálculos *a priori* posibles, y hay más pruebas estadísticas susceptibles de ser abordadas desde la perspectiva de la potencia. Afortunadamente, en la actualidad se cuenta ya con una buena cantidad de bibliografía (tanto en español como en inglés) que aborda el tema de la potencia y el tamaño de efecto, y los paquetes de análisis estadísticos más importantes (como SPSS) ya incluyen también la posibilidad de entregar los índices de potencia observada para una buena parte de las principales pruebas estadísticas empleadas en las ciencias sociales.

Como se vio hasta aquí, entonces, las posibles aplicaciones prácticas de la potencia son varias:

- Se puede determinar un tamaño muestral adecuado antes de efectuar una investigación.
- Se puede determinar la viabilidad o inviabilidad de una investigación dadas ciertas limitantes (habitualmente relacionadas con el tamaño muestral).
- Se controla el riesgo de efectuar errores Tipo II, que por lo general no son tenidos en cuenta en las investigaciones cuantitativas.
- Se puede obtener un índice de potencia real observada para una investigación. Cuando este nivel es lo suficientemente alto, la investigación gana en rigor y en posibilidades de publicación y aceptación.

Estas aplicaciones por sí solas ya deberían ser lo suficientemente atractivas como para comenzar a tomar en serio el tema. Por lo demás, permiten resolver de manera sencilla algunos problemas bastante complicados, si bien conviene volver a insistir en algo reiterado ya varias veces a lo largo de este trabajo: la potencia es esencialmente una convención. Se trata de un constructo teórico de origen matemático que permite abordar ciertas cuestiones estadísticas a partir de grados de mayor o menor probabilidad, y como tal descansa sobre ciertos supuestos y es en más de alguna medida arbitrario.

Sin embargo, el concepto de sensibilidad (la mayor o menor probabilidad de rechazar la  $H_0$ ) no es una convención sino un hecho implícito en cualquier investigación cuantitativa: cada vez que se efectúa una investigación existe el riesgo de obtener una hipótesis nula. Desde esta perspectiva, entonces, e independientemente de si se adoptan o no los parámetros de Cohen, la mera consideración del problema puede llevar a una muy saludable reflexión sobre cuáles son los factores que pueden ayudar a los investigadores a rechazar la hipótesis nula. Se llegará por tanto a concluir que aumentar el tamaño muestral es una buena manera de aumentar la sensibilidad de una investigación, pero no la única. Dependiendo del problema escogido y de las variables medidas, puede haber (y de hecho las hay) otras maneras de aumentar la sensibilidad y la potencia. Interesantemente, estas otras maneras no están convencionalizadas ni son arbitrarias, pues tienen que ver con aspectos relativos al diseño de cada investigación particular. Pero esto, claro está, ya es otro tema.

## BIBLIOGRAFÍA CITADA

- BALLUERKA, N. & M. I. Vergara, 2002: *Diseños de investigación experimental en psicología*, New Jersey: Prentice Hall.
- BONO, R. & J. Arnau, 1995: "Consideraciones generales en torno a los estudios de potencia", en *Anales de Psicología*, 11, 193-202.
- COHEN, J., 1992: "A power primer", en *Psychological Bulletin*, 112, 155-159.
- QUEZADA, C., en prensa. "Sobre métodos y textos", en PARDO, L. (ed.), *El Discurso de la pobreza en América Latina. Estudio de la Red Latinoamericana de Análisis Crítico del Discurso*, Santiago de Chile: Editorial Frasis.