Psicothema 2008. Vol. 20, nº 4, pp. 933-938
www.psicothema.com

ISSN 0214 - 9915 CODEN PSOTEG

Diseño y análisis de la potencia: n y los intervalos de confianza de las medias

José Fernando García, Juan Pascual, María Dolores Frías, Dirk Van Krunckelsven* y Sergio Murgui**
Universidad de Valencia, * Katholieke Universiteit Leuven (Belgium) y ** Universidad Católica de Valencia

En este estudio se analiza la validez del criterio del 80% de potencia estadística para que no se solapen las medias de los intervalos de confianza (IC). Varias simulaciones indican que la potencia mínima para que los límites de dos medias no se solapen, cuando el IC está en el 95%, es de 0,80; pero cuando el IC está en el 99%, es 0,86; y cuando el IC está en el 90%, es 0,75. Si hay más de dos medias, la potencia mínima aumenta considerablemente. Siendo todavía mayor este aumento cuando las medias poblacionales no aumentan monotónicamente. Por lo tanto, para garantizar que los límites no se solapen, en la mayoría de las situaciones analizadas es necesario calcular directamente el mínimo número de observaciones, siendo de poca utilidad los criterios convencionales de la potencia mínima de 0,80.

Design and power analysis: n and confidence intervals of means. In this study, we analyzed the validity of the conventional 80% power. The minimal sample size and power needed to guarantee non-overlapping (1 - α)% confidence intervals for population means were calculated. Several simulations indicate that the minimal power for two means (m= 2) to have non-overlapping CIs is .80, for (1 - α) set to 95%. The minimal power becomes .86 for 99% CIs and .75 for 90% CIs. When multiple means are considered, the required minimal power increases considerably. This increase is even higher when the population means do not increase monotonically. Therefore, the often adopted criterion of a minimal power equal to .80 is not always adequate. Hence, to guarantee that the limits of the CIs do not overlap, most situations require a direct calculation of the minimum number of observations that should enter in a study.

La validez estadística, junto con la validez interna, la externa y la de constructo son fundamentales en el diseño de las investigaciones psicológicas (véase Cook y Campbell, 1979). Este tipo de validez es un requisito independiente de los otros tres (validez interna, externa y de constructo) porque es posible que un estudio, que cumpla con los demás requisitos de validez, llegue a conclusiones equivocadas por no garantizar efectivamente la validez estadística. La validez estadística implica que las pruebas: (1) detectarán las relaciones reales entre las variables independientes y las dependientes (no cometiendo errores en la inferencia del tipo II por encima del límite ß establecido), y (2) no detectarán relaciones irreales entre las variables independientes y las dependientes (no cometiendo errores en la inferencia del tipo I por encima del límite α fijado). En este artículo se analiza la validez del criterio de una potencia estadística mínima de 0,80 (ß fijado en 0,20, 1 - ß= 0,80) para no cometer errores en la inferencia del tipo II (véase Cohen, 1992) cuando se calculan los intervalos de confianza de las medias.

Fecha recepción: 30-5-08 • Fecha aceptación: 25-6-08
Correspondencia: José Fernando García
Facultad de Psicología
Universidad de Valencia
46010 Valencia (Spain)
E-mail: fernando.garcia@uv.es

Cohen (1962, 1965, 1988, 1992) advertía de que las conclusiones de un estudio no dependían solamente de que fuera real la relación entre la variable independiente y la dependiente. Aún en el caso de que la relación fuera real, también dependían del tamaño que tuviera la muestra seleccionada cuando se diseñó el estudio. Si una investigación analizaba una relación real entre una variable independiente y otra dependiente, con todos los demás factores constantes, el mero hecho de que se seleccionara una muestra menor o mayor podría decidir, por sí mismo, cuáles serían las conclusiones del estudio. Cuando existe una relación real entre una variable independiente y otra dependiente, el error que puede cometerse en la inferencia estadística es el del tipo II: no detectar la relación existente entre las variables independiente y dependiente del estudio, siguiéndose de la investigación la conclusión equivocada de que tal relación no existe.

La probabilidad de cometer un error del tipo II está monotónicamente relacionada con el tamaño de la muestra: a menor tamaño de la muestra, mayor probabilidad de cometer el error del tipo II. Por lo tanto, al no controlar la probabilidad de cometer un error del tipo II (llamada ß) puede implicar que un estudio diseñado para analizar una relación real entre una variable independiente y otra dependiente pueda tener cualquier probabilidad de no cometer este error del tipo II (a la probabilidad de no cometer este error se llama potencia: 1 - ß). De hecho, la probabilidad de equivocarse puede ser enorme (por ejemplo, si α= 0,05, η^2= 0,14, n= 2 y a= 2, entonces ß= 0,93). La preocupación fundamental de Cohen

(1962) era la de cuántos estudios habrán descartado relaciones reales e importantes por diseñarse con muestras reducidas.

En la investigación no es muy sencillo controlar la amenaza del error del tipo II. Si bien es cierto que cuanto mayor sea la muestra menor será la probabilidad de cometer un error del tipo II, hay otros dos factores generales que también intervienen (e.g., García, Frías y Pascual, 1999; Maxwell y Delaney, 1990). (1) Cuanto más estricto sea el límite del error del tipo I, mayor es la probabilidad de cometer un error del tipo II. Por ejemplo, con una muestra del mismo tamaño es más difícil detectar relaciones reales si el error del tipo I se fija en 0,01 que si se fija en 0,05. (2) Cuando el tamaño de la relación entre la variable independiente del estudio y la dependiente es menor, lo que se denomina tamaño del efecto pequeño, es más fácil cometer un error del tipo II. Por ejemplo, con una muestra del mismo tamaño es más difícil encontrar relaciones del 1% (η^2= 0,01) que del 16% (η^2= 0,16). Además si las condiciones del estudio son más de dos (a>2), se añade la dificultad de cuántas medias se comparan y qué patrón siguen estas medias en la población (e.g., Borenstein, Rothstein y Cohen, 1997; Erdfelder, Faul y Buchner, 1996; Maxwell, 2004). Si una sola media poblacional es muy diferente a todas las demás o si todas siguen diferencias monotónicamente escalonadas, son dos casos característicos. En el primer caso, es difícil cometer un error del tipo II comparando la media mayor respecto de cualquiera de las demás, pero también es igualmente difícil no cometer el error al comparar las que estén más igualadas. En el segundo, es difícil cometer el error de tipo II entre las dos más distantes que entre las más cercanas (véase Maxwell, 2004; Borenstein et al., 1997).

Respecto al límite máximo de error del tipo II que es conveniente asumir, se ha establecido tácitamente el convenio del 0,20 (ß= 0,20 y 1 - ß= 0,80); recordando esta potencia mínima del 80% (Cohen, 1965, 1990) al otro límite convencional del 5% para el error del tipo I (véase Cohen, 1994). Este límite fue el mínimo seleccionado por Cohen (1992) para una amplia variedad de pruebas estadísticas. Si bien, la desproporción existente entre los dos límites convenidos para garantizar la validez estadística (α= 0,05 vs. ß= 0,20) ha sido discutida en numerosos trabajos (véase la contraposición de los dos errores estadísticos en Brown, 1983; Cascio y Zedeck, 1983; Cohen, 1965; García, Musitu y Veiga, 2006; Martínez y García, 2007; Nagel y Neef, 1977; Pérez, Navarro y Llobell, 1999; Schneider y Darcy, 1984; Unger, Korade, Lazarov, Terrano, Sisodia y Mirnics, 2005; Wang, Leung, Li y Tan, 2005). Por otra parte, la idea de Cohen (1992, p. 156) de aumentar el límite del error del tipo I para conseguir menos error del tipo II no ha tenido una enorme repercusión porque, en definitiva, los dos errores amenazan la validez estadística (e.g., García et al., 1999).

A raíz de diversas críticas a la prueba de la hipótesis nula (e.g., Anderson, Burnham y Thompson, 2000; Bakan, 1966; Cohen, 1994; Meehl, 1978, 1997; Nickerson, 2000; Schmidt, 1996) un amplio número de investigadores ha propuesto calcular los intervalos de confianza (e.g., Cumming y Finch, 2005; Finch, Thomason y Cumming, 2002; Harlow, Mulaik y Steiger, 1997; Meehl, 2002; Nickerson, 2000; Schmidt, 1996; Steiger y Fouladi, 1997; Wilkinson e Inference, 1999). Argumentan éstos que el intervalo de confianza proporciona información acerca de los límites entre los que los estimadores muestrales variarán en un (1 - α)% de las situaciones. No obstante, también se ha señalado que estos límites serán exactos sólo si se conoce el valor paramétrico, ya que en las otras circunstancias la precisión puede alterarse considerablemente (véase Cumming, Williams y Fidler, 2004; Estes, 1997; Steiger y Fouladi, 1997). Desde este sistema de análisis estadístico se interpreta que los estadísticos muestrales indican diferencias poblacionales (cambios reales en los parámetros sobre los que se pretende realizar la inferencia a partir de la muestra del estudio) cuando los distintos intervalos no se solapen. Evidentemente, si los intervalos muestrales se solapan mientras que los parámetros poblacionales son diferentes, se cometerá un error en la inferencia estadística del tipo II.

Por lo tanto, el sistema de calcular los intervalos (cuando la hipótesis nula es falsa) no requiere solamente rechazar la hipótesis nula global, también demostrar que los intervalos de confianza de los estadísticos no se solapan. Por supuesto, el no-solapamiento de los intervalos de confianza y la potencia son el mismo concepto estadístico: a mayor potencia mayor probabilidad de que los intervalos de confianza no se solapen. En consecuencia, la ausencia de solapamiento de los intervalos de confianza de las medias, cuando hay diferencias reales entre los parámetros, depende de los mismos tres factores principales que la potencia: el límite del error del tipo I, el tamaño del efecto y el tamaño de la muestra. (1) Cuanto mayor sea el error del tipo I, mayor es la potencia y menor la distancia que separa a los dos límites de confianza de cada media muestral. Pero también es cierto que los límites de confianza de las medias están más separados simplemente porque se asume mayor error del tipo I. (2) El segundo parámetro es el tamaño del efecto, de manera que cuanto mayor sea esta magnitud en la población, mayor será la potencia y la distancia que separará a las medias muestrales. De forma que aunque los intervalos de confianza sean igual de amplios, al haber mayor separación entre los centros de los mismos (las medias muestrales representarán las diferencias poblacionales) es más fácil que los intervalos no se solapen. (3) El tercer parámetro es el tamaño de la muestra, a mayor número de observaciones mayor potencia estadística y mayor probabilidad de que los intervalos de confianza de las medias no se solapen. Porque el incremento de la muestra reduce la amplitud de los intervalos de confianza, lo mismo que ocurre con el incremento del error del tipo I, pero sin el inconveniente de tener que asumir más error del tipo I.

Con el objetivo de contrastar la validez del convencional criterio de la potencia mínima de 0,80 cuando se calculan los intervalos de confianza, se realizan varias simulaciones que representan un sencillo diseño de una sola variable independiente donde se comparan las m medias muestrales de las m condiciones. Esta situación es habitualmente la de mayor interés para la investigación aplicada (e.g., Brewer, 1972; Cohen, 1962, 1988; Kraemer y Thiemann, 1987). Y si el convencional criterio del cálculo de la potencia mínima de 0,80 fuese válido, se podría aplicar a los estudios donde los investigadores se propongan calcular los intervalos de confianza de las medias para controlar el error del tipo II (e.g., Cumming y Finch, 2005; Finch et al., 2002; Harlow et al., 1997; Meehl, 2002; Nickerson, 2000; Schmidt, 1996; Steiger y Fouladi, 1997; Wilkinson e Inference, 1999).

Parámetros de la simulación

Representamos en todos los casos la situación más idónea para la precisión de los cálculos, (1) se conocen los μ_i valores paramétricos de las m medias del estudio, y (2) el valor paramétrico del tamaño del efecto (escalado mediante η^2) de la fuente de variación de la única variable independiente (X) en el estudio. Seguimos el siguiente proceso para calcular el error típico de estimación. Estimamos la suma de cuadrados de la fuente de variación de la variable independiente:

$$\hat{\eta}^2 SC_{\text{total}} = SC_{\text{x}};$$
$$SC_{\text{total}} = \frac{SC_{\text{x}}}{\hat{\eta}^2}$$

Por otra parte, como sólo hay una variable independiente (X), la

$$SC_{\text{total}} = SC_{\text{x}} + SC_{\text{error}},$$

sustituyendo

$$SC_{\text{error}} = SC_{\text{total}} - SC_{\text{x}} = \frac{SC_{\text{x}}}{\hat{\eta}^2} - SC_{\text{x}} = SC_{\text{x}}\left(\frac{1-\hat{\eta}^2}{\hat{\eta}^2}\right)$$

Por lo tanto,

$$SC_{\text{error}} = SC_{\text{x}} = \left(\frac{1-\hat{\eta}^2}{\hat{\eta}^2}\right).$$

El error típico de estimación será

$$e_{\text{t}} = \sqrt{\frac{MC_{\text{error}}}{n}} = \sqrt{\frac{SC_{\text{x}}\left(\frac{1-\hat{\eta}^2}{\hat{\eta}^2}\right) \Big/ gl_{\text{error}}}{n}}$$

y cada uno de los m límites de confianza (Maxwell y Delaney, 1990, p. 106) se calculan aplicando:

$$L = \overline{Y}_{\text{i}} \pm \sqrt{F_{\alpha,1,gl_{\text{error}}}}\, e_{\text{t}}$$

En esta simulación, como se conocen los valores paramétricos de las m medias, sería:

$$L = \mu_{\text{i}} \pm \sqrt{F_{\text{a},1,gl_{\text{error}}}}\, e_{\text{t}}$$

Para los tamaños del efecto fijamos los convencionales (Cohen, 1988) pequeño (d= 0,20), mediano (d= 0,50) y grande (d= 0,80). Aplicando la fórmula

$$\eta^2_{\text{igual } n} = \frac{d^2}{4+d^2},$$

se obtienen los equivalentes: pequeño (η^2= 0,01), mediano (η^2= 0,06) y grande (η^2= 0,14). Y la probabilidad del error de tipo I lo fijamos en el convencional 0,05, el más restringido 0,01 y el menos restringido 0,10. Los cálculos estadísticos se realizaron con el programa SPSS implementando algunas variaciones desde Visual Basic.

Resultados

Dos medias. En la primera columna de la tabla 1 se presentan los resultados de comparar dos medias (μ_1= 5, μ_2= 6) en las tres condiciones del tamaño del efecto (pequeño, mediano y grande) con los tres niveles de error de tipo I (0,05, 0,10 y 0,01). Cuando el error del tipo I es el convencional (α= 0,05) la potencia estadística que garantiza que los límites de los intervalos de confianza de las dos medias no se solapen es aproximadamente el convencional de 0,80. Esto es cierto independientemente de que el tamaño del efecto sea pequeño, mediano o grande (figura 1). Sin embargo, cuando se aumenta el límite del error del tipo I hasta el 10% (α= 0,10), la potencia mínima para que los intervalos no se solapen disminuye a 0,75. Y si, por el contrario, el error de tipo I se limita hasta el 1% (α= 0,01) la potencia mínima necesaria para que no se solapen los intervalos es de 0,86. La potencia permanece invariante ante los cambios del tamaño del efecto; si bien, por otra parte, se observa que covaría con el límite del error de tipo I.

Más de dos medias. Cuando las medias son tres (μ_1= 5, μ_2= 6, μ_3= 7) y el error de tipo I el convencional (α= 0,05), la potencia estadística mínima para que los intervalos no se solapen es de

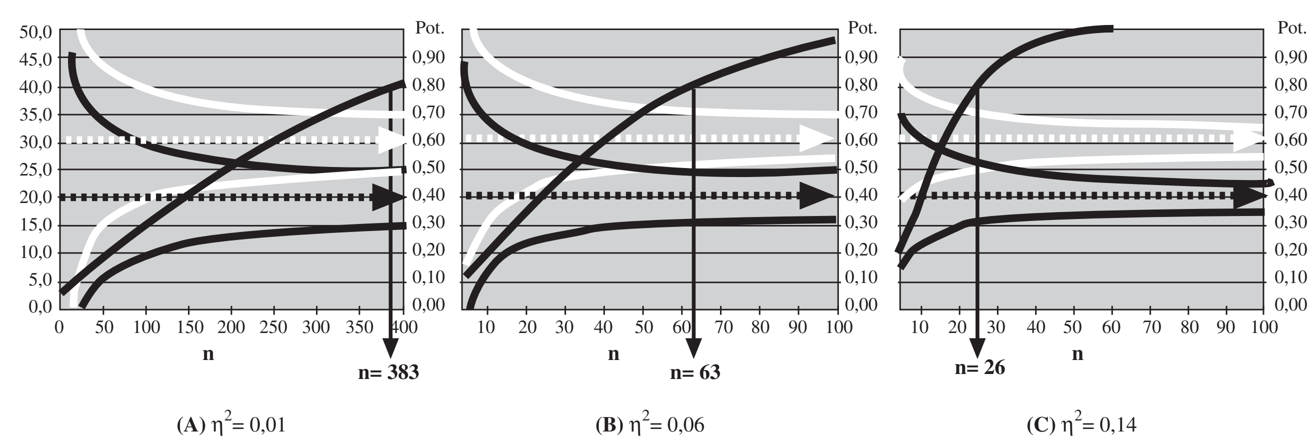


Figura 1. *Incremento en el número de observaciones por condición y cambio en la potencia y en los límites de dos medias (μ_1= 20, μ_2= 30) para el intervalo de confianza del 95% cuando el tamaño del efecto es pequeño (η^2= 0,01), mediano (η^2= 0,06) o grande (η^2= 0,14)*

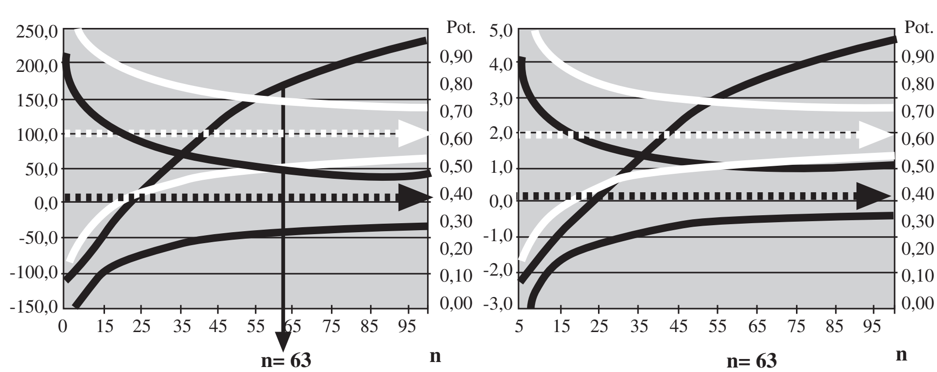
0,999, ya muy alejada de la convencional 0,80. Si se aumenta el límite del error del tipo I hasta el 10% (α= 0,10), la potencia desciende poco, 0,996; pero si se aumenta hasta el 1% (α= 0,01) la potencia mínima necesaria para que los límites no se solapen es de 0,99999. Por lo tanto, el límite mínimo de la potencia de 0,80 no predice en ninguno de los tres niveles de error del tipo I que los intervalos no se solapen, y además, los valores de la potencia alcanzan prácticamente el valor asintótico de la función, considerando un número razonable de cifras decimales. En el caso de que el número de medias se aumente a más de tres, los resultados de la simulación indican lo mismo que en el caso anterior, pero de manera más clara: la potencia mínima se incrementa de tal manera que con cuatro medias y con 10 cifras decimales no se puede distinguir ningún cambio en el límite de error de tipo II. Por el contrario, se observa que el número de observaciones sí que varía considerablemente conforme lo hace el límite de error de tipo I.

Tabla 1

Potencia mínima (1-ß) y número mínimo de observaciones (*n* y *N*) para que los intervalos de confianza no se solapen variando el tamaño del efecto (η^2), el número de medias (*m*) y el límite del error del tipo I (α)

η^2		Número de medias 2	3	4	5	6
α= 0,05						
0,01	*n* [*N*]	383 [766]	1016 [3048]	1904 [7616]	3044 [15220]	4439 [26634]
	1-ß	0,7923281682	0,9993796014	0,9999999996	1,0000000000	1,0000000000
0,06	*n* [*N*]	63 [126]	163 [489]	303 [1212]	483 [2415]	704 [4224]
	1-ß	0,7971627535	0,9993991190	0,9999999996	1,0000000000	1,0000000000
0,14	*n* [*N*]	26 [52]	65 [195]	120 [480]	191 [955]	277 [1662]
	1-ß	0,7989390508	0,9993874269	0,9999999996	1,0000000000	1,0000000000
α= 0,10						
0,01	*n* [*N*]	270 [540]	716 [2148]	1341 [5364]	2145 [10725]	3127 [18762]
	1-ß	0,7528526116	0,9961343579	0,9999998117	1,0000000000	1,0000000000
0,06	*n* [*N*]	45 [90]	115 [345]	214 [856]	341 [1705]	496 [2976]
	1-ß	0,7602138651	0,9961818355	0,9999998165	1,0000000000	1,0000000000
0,14	*n* [*N*]	19 [38]	46 [138]	85 [340]	135 [675]	196 [1176]
	1-ß	0,7673591667	0,9960842286	0,9999998168	1,0000000000	1,0000000000
α= 0,01						
0,01	*n* [*N*]	660 [1320]	1754 [5262]	3287 [13148]	5257 [26285]	7665 [45990]
	1-ß	0,8573059570	0,9999933241	1,0000000000	1,0000000000	1,0000000000
0,06	*n* [*N*]	107 [214]	280 [840]	522 [2088]	834 [4170]	1215 [7290]
	1-ß	0,8585675209	0,9999934408	1,0000000000	1,0000000000	1,0000000000
0,14	*n* [*N*]	44 [88]	111 [333]	206 [824]	328 [1640]	478 [2868]
	1-ß	0,8628688389	0,9999931045	1,0000000000	1,0000000000	1,0000000000

Nota: *n*, número de observaciones por condición; *N*, número total de observaciones. Las medias poblacionales en esta simulación son: μ_1= 5, μ_2= 6, μ_3= 7, μ_4= 8, μ_5= 9 y μ_6= 10



(A) μ_1= 5, μ_2= 100 **(B)** μ_1= 0,25, μ_2= 2,00

Figura 2. *Ejemplo de la independencia de la escala de medida de las medias (α= 0,05 y η^2= 0,06)*

Escala de medida y desproporción entre las medias. Como se aprecia en la figura 2, las simulaciones realizadas son independientes de la métrica de las medias: son aplicables a cualquier situación porque la escala de η^2 es independiente de la escala de medida de la variable dependiente. Pero los resultados de los análisis sí que cambian de manera muy importante cuando las medias poblacionales no aumentan en la misma proporción. En la tabla 2 se simula cómo afecta la desproporción de la media mayor a la potencia y al número de observaciones necesarias para que los límites de los intervalos no se solapen, en el caso de que el tamaño del efecto sea mediano. Los resultados son similares cuando el tamaño del efecto es pequeño o grande.

Discusión

La principal conclusión de este trabajo es que la potencia mínima recomendada por Cohen (1965) de 0,80 es generalmente un criterio insuficiente para los trabajos que calculen los intervalos de confianza de las medias e intenten no cometer un error del tipo II. Paradójicamente, en el único caso que se ha comprobado cierta lógica empírica a esta regla convencional de 0,80 es, precisamente, cuando el error del tipo I se fija en el otro límite convencional de 0,05 (véase Cohen, 1994). En el resto de las situaciones probadas la precisión con que la potencia de 0,80 permite predecir cuándo los límites de los intervalos no se solaparán es muy poca. Para no cometer un error de tipo II utilizando los intervalos de confianza es difícil determinar unas reglas sencillas que se ajusten a todos los casos estudiados fijando un límite mínimo en la potencia, parece que el sistema que ofrece más garantías actualmente consiste en estimar directamente el número de observaciones mínimas tal y como se ha realizado en este trabajo.

El sistema tradicional que contempla sólo la perspectiva de la hipótesis nula es mucho más potente, pero porque asume muchas menos restricciones. Basta con rechazar la hipótesis nula global de igualdad que se prueba con el estadístico *F*, para que se pueda afirmar que se han constatado las diferencias, y no se ha cometido el error del tipo II. Sin embargo, olvidarse que después habrá que determinar qué pares de medias son estadísticamente significativas puede llevar a cometer errores del tipo II en las pruebas específicas posteriores (e.g., Borenstein et al., 1997; Maxwell, 2004). Realmente, ¿cuántas investigaciones concluyen a partir de las pruebas de *F*, cuando tienen más de dos condiciones? Por otra parte, el cálculo de los intervalos de confianza entraña más dificultades, porque no sólo es necesario rechazar la hipótesis nula y acertar en la tendencia (el patrón de orden) que llevarán las medias en la población, también es preciso estimar la magnitud del efecto (Cumming y Finch, 2005; Nickerson, 2000; Schmidt, 1996; Steiger y Fouladi, 1997; Wilkinson e Inference, 1999). Esta mejora en la precisión también implica el mayor costo que supone limitar el error del tipo II (e.g., Arvey, Salas y Maxwell, 1992).

Para que no ocurra con el sistema de los intervalos de confianza lo mismo que con las pruebas de la hipótesis nula (Anderson et al., 2000; Bakan, 1966; Cohen, 1994; Meehl, 1978, 1997; Nickerson, 2000; Schmidt, 1996) habrá que contemplar las situaciones donde las diferencias sean reales y exista un riesgo efectivo de cometer errores del tipo II. Aunque el tamaño del efecto sea tan pequeño como lo son habitualmente los efectos psicológicos (Cohen, 1988; Frick, 1996). Si ya se ha recomendado mucha cautela para

Tabla 2

Resultado del incremento de la media mayor (μ_3= [7-11]) en el tamaño de la muestra mínimo (*n* y *N*) y la potencia (1-ß) para que los intervalos contiguos no se solapen cuando m= 3, μ_1= 5, μ_2= 6, el *IC*= 95% y η^2= 0,06

Medias y patrón de efectos (α)					**Distancias entre las medias**					**1 vs. 2**			**2 vs. 3**		
μ_3	μ	α_1	α_2	α_3	$d_{máx}$	$d_{3\text{-}2}$	$d^{\%}_{2\text{-}1}$	$d^{\%}_{3\text{-}2}$	*IIC*	*n*	1 - ß	*N*	*n*	1 - ß	*N*
7,00	6,00	-1,00	0,00	1,00	2,00	1,00	50,0	50,0	1,000	163	0,999	489	163	0,999	489
7,25	6,08	-1,08	-0,08	1,17	2,25	1,25	44,4	55,6	0,800	206	1,000	618	133	0,997	399
7,50	6,17	-1,17	-0,17	1,33	2,50	1,50	40,0	60,0	0,667	256	1,000	768	115	0,991	345
7,75	6,25	-1,25	-0,25	1,50	2,75	1,75	36,4	63,6	0,571	313	1,000	939	104	0,983	312
8,00	6,33	-1,33	-0,33	1,67	3,00	2,00	33,3	66,7	0,500	377	1,000	1131	96	0,974	288
8,25	6,42	-1,42	-0,42	1,83	3,25	2,25	30,8	69,2	0,444	447	1,000	1341	90	0,965	270
8,50	6,50	-1,50	-0,50	2,00	3,50	2,50	28,6	71,4	0,400	524	1,000	1572	86	0,958	258
8,75	6,58	-1,58	-0,58	2,17	3,75	2,75	26,7	73,3	0,364	607	1,000	1821	82	0,948	246
9,00	6,67	-1,67	-0,67	2,33	4,00	3,00	25,0	75,0	0,333	698	1,000	2094	80	0,943	240
9,25	6,75	-1,75	-0,75	2,50	4,25	3,25	23,5	76,5	0,308	795	1,000	2385	77	0,934	231
9,50	6,83	-1,83	-0,83	2,67	4,50	3,50	22,2	77,8	0,286	898	1,000	2694	75	0,928	225
9,75	6,92	-1,92	-0,92	2,83	4,75	3,75	21,1	78,9	0,267	1009	1,000	3027	74	0,924	222
10,00	7,00	-2,00	-1,00	3,00	5,00	4,00	20,0	80,0	0,250	1126	1,000	3378	73	0,920	219
10,25	7,08	-2,08	-1,08	3,17	5,25	4,25	19,0	81,0	0,235	1249	1,000	3747	71	0,913	213
10,50	7,17	-2,17	-1,17	3,33	5,50	4,50	18,2	81,8	0,222	1380	1,000	4140	70	0,909	210
10,75	7,25	-2,25	-1,25	3,50	5,75	4,75	17,4	82,6	0,211	1517	1,000	4551	69	0,904	207
11,00	7,33	-2,33	-1,33	3,67	6,00	5,00	16,7	83,3	0,200	1661	1,000	4983	69	0,904	207

Nota: μ, media total; μ_i, media de la condición *i*; α_i, efecto en cada condición *i* (μ - μ_i); $d_{máx}$, distancia entre la media menor y mayor; $d_{i\text{-}j}$, distancia absoluta entre la medias *i* y *j* ($d_{i\text{-}j}$= μ_i - μ_j); $d^{\%}_{i\text{-}j}$, distancia relativa entre la medias *i* y *j* ($d^{\%}_{i\text{-}j}$= $d_{i\text{-}j}$ / $d_{máx}$); *IIC*, índice de incremento constante (cociente entre la distancia menor y la mayor); *n*, número de observaciones por condición; *N*, número total de observaciones del estudio. $d_{1\text{-}2}$ siempre es 1 examinar mediante cálculos de la potencia basados en la prueba *F* posteriores diferencias entre pares de medias (Borenstein et al., 1997), aún más se tienen que hacer respecto de este otro sistema. Las limitaciones de este trabajo son importantes porque se han simulado las situaciones más idóneas (Cumming et al., 2004; Estes, 1997; Steiger y Fouladi, 1997). Esta consideración aún hace más acuciante que se desarrollen nuevos sistemas para controlar eficazmente el problemático error del tipo II en los diseños psicológicos (Maxwell, 2004). O de momento, otra alternativa prudente puede consistir simplemente en ajustar las pretensiones predictoras de los modelos psicológicos a las posibilidades actuales de la psicología científica (véase Frick, 1996).

Referencias

Anderson, D.R., Burnham, K.P., y Thompson, W.L. (2000). Null hypothesis testing: Problems, prevalence and an alternative. *Journal of Wildlife Management, 64*, 912-923.

Arvey, R.D., Salas, E., y Maxwell, S.E. (1992). The relative power of training evaluation designs under different cost configurations. *Journal of Applied Psychology, 77*, 155-160.

Bakan, D. (1966). Test of significance in psychological research. *Psychological Bulletin, 66*, 423-437.

Borenstein, M., Rothstein, H., y Cohen, J. (1997). *Power and precision*. Teaneck, NJ: Biostat.

Brewer, J.K. (1972). Power of statistical tests in american-educational-research-journal. *American Educational Research Journal, 9*, 391-401.

Brown, G.W. (1983). Errors, Type I and II. *American Journal of Disorders in Childhood, 137*, 586-591.

Cascio, W.F., y Zedeck, S. (1983). Open a new window in rational research planning: Adjust alpha to maximize statistical power. *Personnel Psychology, 36*, 517-526.

Cohen, J. (1962). Statistical power of abnormal-social psychological-research: A review. *Journal of Abnormal Psychology, 65*, 145-153.

Cohen, J. (1965). Some statistical issues in psychological research. En B.B. Wolman (Ed.): *Handbook of clinical psychology* (pp. 95-121). New York: McGraw-Hill.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J.: Erlbaum.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45*, 1304-1312.

Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychological Bulletin, 112*, 155-159.

Cohen, J. (1994). The earth is round (p-less-than .05). *American Psychologist, 49*(12), 997-1003.

Cook, T.D., y Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.

Cumming, G., y Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60*, 170-180.

Cumming, G., Williams, J., y Fidler, F. (2004). Replication and researchers' understanding of confidence intervals and standard error bars. *Understanding Statistics, 3*, 299-311.

Erdfelder, E., Faul, F., y Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods Instruments & Computers, 28*, 1-11.

Estes, W.K. (1997). On the communication of information by displays of standard errors and confidence intervals. *Psychonomic Bulletin & Review, 4*, 330-341.

Finch, S., Thomason, N., y Cumming, G. (2002). Past and future American Psychological Association guidelines for statistical practice. *Theory & Psychology, 12*, 825-853.

Frick, R.W. (1996). The appropriate use of null hypothesis testing. *Psychological Methods, 1*, 379-390.

García, J.F., Frías, M.D., y Pascual, J. (1999). *Los diseños de la investigación experimental: comprobación de las hipótesis* [Experimental research designs: Verification of hypotheses]. Valencia, Spain: Cristóbal Serrano Villalba.

García, J.F., Musitu, G., y Veiga, F. (2006). Autoconcepto en adultos de España y Portugal [Self-concept in adults from Spain and Portugal]. *Psicothema, 18*, 551-556.

Harlow, L.L., Mulaik, S.A., y Steiger, J.H. (1997). *What if there were no significance tests?* Mahwah, N.J.: Erlbaum.

Kraemer, H.C., y Thiemann, S. (1987). *How many subjects? Statistical power analysis in research*. Newbury Park, CA: Sage.

Martínez, I., y García, J.F. (2007). Impact of parenting styles on adolescents' self-esteem and internalization of values in Spain. *The Spanish Journal of Psychology, 10*, 338-348.

Maxwell, S.E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences and remedies. *Psychological Methods, 9*, 147-163.

Maxwell, S.E., y Delaney, H.D. (1990). *Designing experiments and analyzing data: A model comparison perspective*. Belmont, CA: Wadsworth.

Meehl, P.E. (1978). Theoretical risks and tabular asterisks: Karl, Ronald, and slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46*, 806-834.

Meehl, P.E. (1997). The problem is epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. En L.L. Harlow, S.A. Mulaik y L.H. Steiger (Eds.): *What if there were no significance tests?* (pp. 393-425). Mahwah, NJ: Erlbaum.

Meehl, P.E. (2002). Cliometric metatheory: II. Criteria scientists use in theory appraisal and why it is rational to do so. *Psychological Reports, 91*, 339-404.

Nagel, S.S., y Neef, M. (1977). Determining an optimum level of statistical significance. *Evaluation Studies Review Annual, 2*, 146-158.

Nickerson, R.S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5*, 241-301.

Pérez, J.F.G., Navarro, D.F., y Llobell, J.P. (1999). Statistical power of Solomon design. *Psicothema, 11*, 431-436.

Schmidt, F.L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods, 1*, 115-129.

Schneider, A.L., y Darcy, R.E. (1984). Policy implications of using significance tests in evaluation research. *Evaluation Review, 8*, 573-582.

Steiger, J.H., y Fouladi, R.T. (1997). Noncentrality interval estimation and the evaluation of statistical methods. En L.L. Harlow, S.A. Mulaik y L.H. Steiger (Eds.): *What if there were no significance tests?* (pp. 221-257). Mahwah, NJ: Erlbaum.

Unger, T., Korade, Z., Lazarov, O., Terrano, D., Sisodia, S.S., y Mirnics, K. (2005). True and false discovery in DNA microarray experiments: Transcriptome changes in the hippocampus of presenilin 1 mutant mice. *Methods, 37*, 261-273.

Wang, Y.G., Leung, D.H., Y., Li, M., y Tan, S.B. (2005). Bayesian designs with frequentist and Bayesian error rate considerations. *Statistical Methods in Medical Research, 14*, 445-456.

Wilkinson, L., e Inference, T.F.S. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604.