



Algunas consideraciones metodológicas relevantes para la investigación aplicada

Borges del Rosal, A.¹ y Sánchez-Bruno, A.
Facultad de Psicología. Universidad de la Laguna.

RESUMEN

Las controversias que se suscitan en el seno de la metodología no llegan al investigador aplicado con la fluidez que sería conveniente. En el presente trabajo se señalan algunos de los problemas más destacados de la investigación que hace uso de la estadística inferencial, así como recomendaciones a seguir en la investigación, desglosada para cada una de las fases de la misma: planificación de la investigación, realización de la misma, análisis de datos y disseminación de resultados.

Palabras clave: Investigación, metodología, inferencia estadística.

ABSTRACT

Controversies that arouse in the heart of methodology do not reach applied researchers with the desirable fluency. In this paper we show some of the most relevant problems of researches that use statistical inference, and some recommendations to follow when investigating, specifying the phase of the research: planning, realization, data analysis and results dissemination.

Keywords: Research, methodology, statistical inference.

¹e-mail: aborges@ull.es



1.- Introducción

Las últimas recomendaciones de la APA (Wilkinson and the Task Force, 1999) ponen sobre el tapete la necesidad de poner especial cuidado en el proceso general de investigación. La comisión, creada para dar respuesta a la polémica suscitada por las críticas a los contrastes de significación estadística (Rozeboom, 1960; Cohen, 1962; Binder, 1963; Falk y Greenbaum, 1995; Kirk, 1996; Abelson, 1997; Harris, 1997; Schmit y Hunter, 1997; entre otros), no se ha limitado a establecer directrices respecto al análisis de datos, sino que ha repasado punto por punto lo que supone el hacer en investigación, recordándonos lo que los libros de texto nos enseñaron en su momento y a lo que, por lo que parece, no se presta toda la atención que merece.

Nuestro equipo de investigación viene trabajando desde el año 93 en la comprobación de la bondad de los contrastes estadísticos de 2 y 3 grupos, analizando mediante simulación el comportamiento de los contrastes paramétricos y no paramétricos para estos casos cuando se alteran los supuestos subyacentes. La elección de estos tests no ha sido aleatoria; se debe a que tanto la prueba *t* de Student como el ANOVA son las más utilizadas para analizar los resultados de las investigaciones en Psicología (Kirk, 1996; Elmore y Woehlke, 1998.). Centrándonos en estudios españoles, Valera y colaboradores (1998) encontraron en su revisión de los años 90-92 de la *Revista de Psicología General y Aplicada* que más de la mitad de los estudios que realizaron contrastes estadísticos lo hacían mediante *t* o ANOVA (un 53,2%). Además, puesto que sabemos que las condiciones de investigación en Psicología aplicada hacen inviable la utilización de grandes muestras, aspecto que también ha quedado recogido en la literatura (Shapiro y Shapiro, 1982), en nuestro trabajo hemos simulado grupos que van desde 5 observaciones hasta 50. Los resultados obtenidos en nuestras investigaciones, unidos a las controversias y recomendaciones que en fechas recientes han aparecido en la literatura (ver, por ejemplo, Borges y Sánchez Bruno, 2002), nos permiten exponer un resumen de los problemas que con mayor frecuencia aparecen en los estudios basados en la inferencia estadística y esbozar algunas recomendaciones a los investigadores aplicados, con el objeto de que les sirvan de guía y les ayuden en las cuatro etapas básicas de la investigación: planificación, realización, análisis de datos e informe de los resultados obtenidos.

2.- Problemas corrientes en los estudios inferenciales

La utilización masiva de la inferencia estadística, en particular el contraste de significación, ha hecho que se produzcan problemas de diversa índole en la interpretación de los resultados. No es nuestra intención profundizar en este tema, pues ya ha sido tratado anteriormente (Borges, 1997; Borges, San Luis, Sánchez-Bruno y Cañadas, 2001) por lo que vamos a comentar sólo algunas de las cuestiones que, a nuestro juicio, resultan más relevantes.

En primer lugar, aún no es demasiado habitual la comprobación de los supuestos subyacentes a los contrastes empleados, lo que puede conducir a que los mencionados supuestos se violen, dando lugar al rechazo de la hipótesis nula a niveles de significación que pueden diferir de forma importante del nivel alfa elegido por el investigador.



En segundo lugar, muchas investigaciones presentan serios problemas de heterocedasticidad, con el devastador efecto que tal violación del supuesto paramétrico de igualdad de las varianzas poblacionales tiene en el contraste estadístico (Borges, San Luis y Sánchez Bruno, 1993; Borges, Sánchez-Bruno y Cañadas, 1996; Sánchez-Bruno, Borges, San Luis y Cañadas, 2000; Borges, San Luis y Sánchez-Bruno, 2003). El origen de la heterocedasticidad puede estar en la no aditividad de los efectos de los tratamientos, es decir, en que no sea de aplicación el Modelo Lineal General al problema de estudio concreto, pero más frecuentemente se debe a un muestreo inadecuado o a un mal control de las variables extrañas a la hora de asignar los sujetos a los grupos experimentales.

En este sentido, el empleo de la estadística no paramétrica no resulta más adecuado cuando se incumplen los supuestos de la paramétrica, pudiendo incluso tener efectos aún peores sobre las probabilidades reales de cometer un error de tipo I (Sánchez-Bruno y Borges, 1997; Cañadas, Borges y Sánchez-Bruno, en prensa)

Se ha recomendado también como alternativa a los contrastes de dos o más medias cuando se incumplen los supuestos subyacentes, el uso simultáneo de estimadores robustos (por ejemplo, las medias recortadas) y contrastes que controlan los grados de libertad con procedimientos *bootstrap* (Wilcox, 1997; Wilcox, 1998a; Wilcox, Keselman, y Kowalchuk, 1998)¹ o de contrastes ordinales (Cliff, 1993). No obstante, nuestros propios trabajos (Borges, Sánchez-Bruno, Cañadas y San Luis, 2000; Sánchez-Bruno, Borges y Cañadas, 2000; Sánchez-Bruno, Borges, San Luis y Cañadas, 2000; Sánchez-Bruno y Borges, 2002) no confirman estas expectativas, obteniéndose resultados que, en el mejor de los casos, resultan similares a los que se obtienen con las pruebas paramétricas tradicionales.

Un tercer problema, que se da especialmente en determinados ámbitos de la investigación aplicada, es el uso de grupos de pequeño tamaño, lo que dificulta alcanzar potencias adecuadas en los contrastes, especialmente si el tamaño del efecto esperable es moderado o pequeño (Kazdin y Bass, 1989; Sánchez-Bruno, Borges, San Luis y Cañadas, 1999; Cañadas, Borges, Sánchez-Bruno y San Luis, 2000).

Por último, y como colofón a lo que hemos ido enumerando, es común encontrar inferencias inadecuadas: La “lógica híbrida” (Gigerenzer, 1993) ha dado una apariencia de objetividad en los resultados de las investigaciones a costa de llegar a conclusiones inadmisibles desde la regla de decisión lógica que supone el contraste de hipótesis (Borges, 1997).

Tal como avanzamos anteriormente, presentamos a continuación, sin ánimo de ser exhaustivos, una serie de recomendaciones que esperamos que resulten útiles en las cuatro etapas básicas de la investigación: planificación, realización, análisis de datos e informe de los resultados obtenidos.

¹ Los trabajos de Lix, Keselman y Keselman (1996) y de Keselman y cols. (1998) dan una detallada información sobre el análisis de varianza y procedimientos alternativos.



3.- Recomendaciones

3.1.- En la planificación de la investigación

Todos sabemos que la investigación realizada por los demás miembros de la comunidad científica es la guía que nos abre el camino en nuestro propio hacer, y que cuanto más completa y exhaustiva sea esta revisión teórica más probabilidades tendremos de que el trabajo realizado sea riguroso y que nuestra propia aportación ayude al avance de la Ciencia. La cuidadosa revisión bibliográfica es un aspecto a resaltar. El tipo de información que se puede extraer de ella es muy variado:

- a) Hipótesis de trabajo para la investigación en curso.
- b) Análisis de los estudios previos, diseños utilizados y resultados conseguidos.
- c) Determinación de los tamaños de efecto obtenidos.
- d) Disparidad o coherencia de los resultados obtenidos en los trabajos revisados.

Con respecto al diseño a elegir, caben dos recomendaciones. En primer lugar, afinarlo, con el fin de evitar errores que redunden en disminución de la varianza primaria. En segundo lugar, conviene elegir el diseño en función de los objetivos propuestos, no en razón del mayor *prestigio* presumible de un diseño frente a otro (Wilkinson and the Task Force, 1999).

En lo referente a la elección de la muestra, es preciso realizar los procedimientos aleatorios adecuados tanto en la selección de los sujetos como en su asignación a los grupos cuando éste sea el caso. Cuando la asignación aleatoria no sea posible, habrá que minimizar la influencia de variables extrañas, sin olvidar que tanto abandonos como casos perdidos forman parte de esas variables.

También en relación con la elección de la muestra, hay que determinar el tamaño muestral en función del tamaño del efecto esperable y de la potencia deseable. Esta cuestión, puesta en evidencia sucesivas veces (Cohen, 1962; Sedlmeier y Gigerenzer, 1989; Kazdin y Bass, 1989; Clark-Carter, 1997; Valera, Sánchez-Meca y Marín, 2000), no parece obtener la atención debida por parte de los investigadores.

Mención aparte requiere la explicitación de las variables incluidas en el estudio y cómo han de ser medidas. Esto incumbe también a la integridad de la terapia cuando se estudian los efectos de intervenciones (Gresham, 1997; Amstrong, Ehrardt, Cool y Poling, 1997; Wickstrom, Jones, LaFleur y Witt, 1998).

Conviene tener en cuenta que todos los aspectos que se recojan en la investigación deben tener una justificación y un enlace con las hipótesis que se van a contrastar. Suele ser corriente, no obstante, que los investigadores noveles amplíen los límites de la investigación² “por si pudiera venir bien”. Esta práctica, además de no aportar beneficios claros, puede

² Por ejemplo, incluyendo más ítems en un cuestionario, que no tienen una clara razón de ser.



oscurecer la investigación e incrementar el componente de error. Parafraseando a Cohen (1990), en este caso “más es menos”.

Por último, recordar que, aunque pueda parecer que un estudio piloto alarga y encarece la investigación, no sólo no es así, sino que las ventajas del mismo superan con mucho los posibles inconvenientes. En el estudio piloto se ponen a prueba los instrumentos a utilizar, salen a la luz determinados problemas que pueden afectar a la investigación e incluso arruinarla, además de que en muchos casos sirven como entrenamiento, tanto para el investigador como para el resto del equipo.

3.2. En la realización de la investigación

Dejando a un lado que cada procedimiento de investigación requiere que se tengan en cuenta cuestiones metodológicas específicas (por ejemplo, la metodología de una encuesta difiere de la de un experimento de laboratorio), algunos aspectos son comunes en cualquier investigación.

En primer lugar, es preciso describir la disminución de la muestra por efecto de abandonos, casos perdidos, etc. (Wilkinson and the Task Force, 1999). La pérdida de sujetos en la investigación es siempre un problema, pero puede ser más o menos grave dependiendo de cuántos sujetos no continúen, y, caso de poderlo saber, de las razones para no seguir en la investigación.

En segundo lugar, es recomendable que el investigador describa la forma de tratar los sesgos posibles o los claramente detectados (Wilkinson and the Task Force, 1999).

Una tercera recomendación para esta fase de la investigación es la estandarización cuidadosa del procedimiento de recogida de datos, de manera que se eviten sesgos por una implementación inadecuada.

3.3.- Durante el análisis de datos

Las sugerencias, en esta fase de la investigación, son numerosas y responden a diversas controversias que han ido sucediéndose en el desarrollo de nuestra disciplina.

- a) Exploración cuidadosa de los datos, con especial hincapié en el uso de análisis gráficos (Lix, Keselman y Keselman, 1996; Wilkinson and the Task Force, 1999)
- b) En instrumentos psicométricos, calcular la fiabilidad para la muestra estudiada y su correspondiente intervalo de confianza. (Thompson, 1994; Thompson y Snyder, 1998; Fan y Thompson, 2001).
- c) Comprobación previa de los supuestos antes de cualquier enfoque inferencial (Keselman y cols., 1998)
- d) Elección del nivel de significación por criterios teóricos, no sólo por la costumbre (Gigerenzer, 1993)



- e) Utilizar los estadísticos adecuados a las preguntas teóricas planteadas. Por ejemplo, no realizar un test ómnibus cuando se quiere concretar un contraste determinado (Keselman y cols., 1998).
- f) Ser muy precavidos en el uso de comparaciones múltiples (Cañadas, 1999). *Nunca* usar contrastes *a priori* si no se tiene justificación teórica para ello y, desde luego, *jamás* tras un ANVAR ómnibus, como en ocasiones hemos podido detectar en la literatura española (y extranjera), incluso en revistas de reconocido prestigio.
- g) A ser posible, utilizar los estadísticos más simples (Wilkinson and the Task Force, 1999)
- h) Acompañar las pruebas de significación con intervalos de confianza (Cohen, 1990; Frick, 1996) y contrastar la significación estadística con el tamaño del efecto obtenido. Esto permitirá superar la eterna dicotomía entre resultados *significativos estadísticamente* y resultados *relevantes desde una perspectiva clínica y/o científica* (Barlow y Hersen, 1988; Kazdin, 1982), llegando así a la *significación práctica* (Kirk, 1996).
- i) No usar estadísticos multivariados si no se va a extraer toda la información adecuada (Keselman y cols., 1998).
- j) No llevar a cabo análisis estándar por ser los más utilizados en la literatura. Por ejemplo, aunque las pruebas *t-Student*, ANVAR o correlación y regresión siguen siendo las preferidas, su comportamiento cuando se incumplen sus supuestos hace recomendable en ese caso el uso de otros procedimientos (Keselman y cols., 1998; Wilcox, 1998b.)

3.4.- El informe de resultados

La diseminación de resultados es otro de los aspectos de vital importancia en una investigación, por lo que vamos a señalar algunas recomendaciones para esta fase final.

En primer lugar, es preciso informar de las características de la población de referencia³ (sujetos, estímulos o estudios) de la que se ha extraído la muestra (Wilkinson and the Task Force, 1999) y no únicamente de ésta, como suele ser más la regla que la excepción en los artículos publicados.

En línea con el mayor interés suscitado por alternativas al contraste de hipótesis, se debe informar de los tamaños del efecto e intervalos de confianza, prestando mayor atención a la *significación sustantiva* que a la mera *significación estadística* (Keselman y cols., 1998)

Es necesario interpretar los efectos encontrados teniendo en cuenta el diseño empleado, su credibilidad y su generalizabilidad, conectándolo con la literatura conocida sobre el particular. Recuérdese que hay investigaciones más humildes y más ambiciosas, y

³ Aunque a veces el problema no sea de simple solución. Recomendamos al respecto el interesante trabajo de Frick, R.W. (1998) "Interpreting statistical testing: Process and propensity, not population and random sampling. *Behavior Research Methods, Instruments and Computers*, 30, 527-535"



que todas tienen cabida si están bien realizadas, dentro de las limitaciones que imponga el procedimiento de investigación seguido.

Hay que presentar conclusiones apropiadas y explícitas, señalando las limitaciones del trabajo para superarlas en estudios futuros sobre el tema en cuestión. Es conveniente eliminar comentarios poco operativos y sin contenido como, por ejemplo, “es preciso investigar más sobre el tema” (Wilkinson and the Task Force, 1999).

Por último, es preciso recordar que los resultados de un estudio individual no tienen ningún sentido si no se pueden incluir en trabajos integradores⁴, como es el caso del meta-análisis (véase, por ejemplo, los trabajos del equipo de Sánchez Meca sobre meta-análisis de terapia: Sánchez Meca, Olivares y Rosa, 1998; Sánchez Meca, Rosa y Olivares, 1999). Schmidt (1996) aboga por una nueva forma de hacer, fundamentando los hallazgos en función de los resultados obtenidos en meta-análisis.

4.- Referencias

- Abelson, R.P. (1997) On the surprising longevity of flogged horses: Why there is a case for the significance test. *Psychological Science*, 8, 12-15.
- Amstrong, K.J., Ehrardt, K.E., Cool, R.T. y Pooling, A. (1997). Social validity and treatment integrity data: Reporting in articles published in *The Journal of Developmental and Physical Disabilities*, 1992-1995. *Journal of Developmental and Physical Disabilities*, 9, 359-367.
- Barlow, D.H. y Hersen, M. (1988) *Diseños experimentales de caso único*. Barcelona: Martínez Roca (original: 1984, Pergamon Press)
- Binder, A. (1963) Further considerations on testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychological Review*, 70, 107-115.
- Borges, A. (1997) Algunos problemas frecuentes en la interpretación de los contrastes de hipótesis estadísticas en psicología. *Iberpsicología*, 2:3:7 <http://fs-morente.filos.ucm.es/publicaciones/iberpsicologia/iberpsicologia.htm>
- Borges, A. y Sánchez Bruno, J. A. (2002) La simulación al servicio de los contrastes estadísticos: resumen de métodos y estado de la cuestión en grupos pequeños e independientes. *Metodología de las Ciencias del Comportamiento*, 4, 255-271
- Borges, A., San Luis, C. y Sánchez-Bruno, A. (1993, julio). *Contraste de la hipótesis nula para la diferencia de muestras: Alternativa frente al problema de Berhrens-Fisher*. Póster presentado al III Simposio de Metodología de las Ciencias Sociales y del Comportamiento. Santiago de Compostela.

⁴ Téngase en cuenta que una investigación, aunque sea realizada con una muestra numerosa, es sólo un caso particular.



- Borges, A., San Luis, C. y Sánchez-Bruno (2003). Dominance Statistics: A simulation study on the d statistic. *Quality and Quantity*, 37, 303-316.
- Borges, A., Sánchez-Bruno y Cañadas, I. (1996). El contraste de las diferencias de medias con grupos pequeños, con escalas ordinales y en ausencia de normalidad. *Psicológica*, 17, 455-466.
- Borges, A., Sánchez Bruno, J.A., Cañadas, I. y San Luis, C. (2000, julio) *Robust statistics in the test of Statistical hypotheses: an alternative when parametric assumptions are violated*. Póster presentado al 22nd Biennial Conference of the Society for Multivariate Analysis in the Behavioural Sciences. Londres.
- Borges, A., San Luis, C., Sánchez-Bruno, A. y Cañadas, I. (2001) El juicio contra la hipótesis nula. Muchos testigos y una sentencia virtuosa. *Psicothema*, 13 (1) 173-178
- Cañadas, I. (1999). *Análisis de varianza*. Tenerife: Resma.
- Cañadas, I., Borges, A., Sánchez-Bruno, A. y San Luis, C. (2000). Estudio de la potencia de los contrastes de medias con dos y tres grupos, con tamaño de efecto pequeño y en condiciones de no normalidad y homo-heterocedasticidad. *Psicothema* 12, 114-116.
- Cañadas, I., Borges, A. y Sánchez-Bruno, A. (en prensa). La t de Student y sus alternativas, ante la violación de los supuestos. *Psicothema*.
- Clark-Carter, D. (1997) The account taken of statistical power in research published in the *British Journal of Psychology*. *British Journal of Psychology*, 88, 71-83.
- Cliff, N. (1993) Dominance Statistics: Ordinal Analyses to answer ordinal questions *Psychological Bulletin*, 114, 494 – 509
- Cohen, J. (1962) The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, 65, 145-153.
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45, 1304-1312.
- Elmore, P.B. y Woehlke, P.L. (1998) Statistical methods employed in *American Educational Research Journal*, *Educational Researcher* and *Review of Educational Research* from 1978 to 1987. *Educational Researcher*, 17, 19-20.
- Falk, R. y Greenbaum, C.W. (1995). Significance tests die hard. *Theory and Psychology*, 5, 75-98.
- Fan, X. y Thompson, B. (2001). Confidence Intervals about score reliability coefficients, please: an EPM guidelines editorial. *Educational and Psychological Measurement*, 61(4), 517-531.



- Frick, R.W. (1996). The appropriate use of null hypothesis testing. *Psychological Methods*, 1, 379-390.
- Gigerenzer, G. (1993) The Superego, the Ego and the Id in statistical reasoning. En G. Keren y Ch. Lewis (Eds.) *A handbook for data analysis in the Behavioral Sciences. Methodological issues*. Hillsdale, NJ: LEA.
- Greshman, F.M. (1997) Treatment integrity in single-subject research. En R.D. Franklin, D.B. Allison y B.S. Gorman (Eds) *Design and analysis of single-case research*. Mahwah, N.J.: LEA.
- Harris, R.J. (1997) Significance tests have their place. *Psychological Science*, 8, 8-11.
- Kazdin, A.E. (1982) Single-case experimental design in clinical research and practice. En A.E. Kazdin y A.H. Tuma (Eds.) *New directions for methodology of social and Behavioral Sciences. Single-case research designs*. Nº13. San Francisco: Josey Bass.
- Kazdin, A.E. y Bass, D. (1989). Power to detect differences between alternative treatments in comparative psychotherapy outcome research. *Journal of Consulting and Clinical Psychology*, 57, 138-147.
- Keselman, H.J., Huberty, C.J., Olejnik, S., Cribbie, R.A., Donahue, B., Kowalchuk, R.K., Lowman, L.L., Petoskey, M.D., Keselman, J.C. y Levin, J.R. (1998) Statistical practices of educational research: An analysis of their Anova, Manova, and Ancova analysis. *Review of Educational Research*, 68, 350-386
- Kirk, R.E. (1996) Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56, 746-759.
- Lix, L.M., Keselman, J.C. y Keselman, H.J. (1996) Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. *Review of Educational Research*, 66, 579-619.
- Rozeboom, W.W. (1960) The fallacy of the null hypothesis significance test. *Psychological Bulletin*, 57, 416-428.
- Sedlmeier, P. y Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, 105, 309-316.
- Sánchez-Bruno, A. y Borges, A. (1997). Violación del supuesto de normalidad en contrastes estadísticos para grupos pequeños. Comunicación presentada al V Simposio de Metodología de las Ciencias del Comportamiento. Sevilla.
- Sánchez-Bruno, J.A. y Borges, A. (2002, julio) The bootstrap procedure in the trimmed means tests. *23rd Meeting of the Society of Multivariate Analysis in the Behavioural Sciences (SMABS)*. Tilburg (Holanda)



- Sánchez Bruno, A., Borges del Rosal, A., San Luis, C. y Cañadas, I. (1999, noviembre). *Algunos problemas frecuentes de análisis de datos en la investigación en Psicología del deporte: potencia de los contrastes en grupos pequeños*. Póster presentado al I Congreso Internacional de Psicología Aplicada al Deporte. Madrid.
- Sánchez Bruno, A., Borges del Rosal, A. y Cañadas, I. (2000). El contraste de medias recortadas ante la violación de los supuestos paramétricos. *Psicothema, 12*, 506-508.
- Sánchez Bruno, A., Borges, A., San Luis, C. y Cañadas, I. (2000, julio). *Approximation to the Effect Size based on the Statistic Δ , in the Comparisons of Samples with Normal and Non-normal, and Homo- and Heterocedastic Distributions*. Póster presentado en la 22nd Biennial Conference of the Society for Multivariate Analysis in the Behavioural Sciences. Londres.
- Sánchez Meca, J., Olivares, J. y Rosa, A.I. (1998). El problema de la adición al tabaco: meta-análisis de las intervenciones conductuales en España. *Psicothema, 10*, 535-549.
- Sánchez Meca, J., Rosa, A.I. y Olivares, J. (1999). Las técnicas cognitivo-conductuales en problemas clínicos y de salud: meta-análisis de la literatura española. *Psicothema, 11*, 641-654.
- Schmidt, F.L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods, 1*, 115-129.
- Schmidt, F.L. y Hunter, J. (1997) Eight common but false objections to the discontinuation of significance testing in the analysis of research data. En L. Harlow, S. Mulaik y J. Steiger (Eds.) *What if there were no significance tests?* Mahwah, N.J.: LEA.
- Shapiro, D.A. y Shapiro, D. (1982) Meta-analysis of comparative therapy outcome studies: A replication and refinement. *Psychological Bulletin, 92*, 581-604.
- Thompson, B. (1994) Guidelines for authors. *Educational and Psychological Measurement, 54*, 837-847.
- Thompson, B. y Snyder, P.A. (1998). Statistical significance and reliability analysis in recent JCD research articles. *Journal of Counseling and Development, 76*, 436-441.
- Valera, A., Sánchez-Meca, J. y Marín, F. (2000) Contraste de hipótesis e investigación psicológica española: Análisis y propuestas. *Psicothema, 12*, 549-552.
- Valera, A., Sánchez-Meca, J., Marín, F. y Velandrino, A.P. (1998) Potencia estadística de la *Revista de Psicología General y Aplicada* (1990-1992). *Revista de Psicología General y Aplicada, 51*, 233-246.
- Wickstrom, K.J., Jones, K.M., LaFleur, L.H. y Witt, J.C. (1998) An analysis of treatment integrity in school-based behavioral consultation. *School of Psychology Quarterly, 13*, 141-154.



- Wilcox, R.R. (1997) A bootstrap modification of the Alexander-Govern Anova method, plus comments on comparing trimmed means. *Educational and Psychological Measurement*, 57, 655-665.
- Wilcox, R.R. (1998a) The goals and strategies of robust methods. *British Journal of Mathematical and Statistical Psychology*, 51, 1-39.
- Wilcox, R.R. (1998b) How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist*, 53, 300-314.
- Wilcox, R.R., Keselman, H.J. y Kowalchuk, R.K. (1998) Can tests for treatment group equality be improved?: The bootstrap and trimmed means conjecture. *British Journal of Mathematical and Statistical Psychology*, 51, 123-134.
- Wilkinson, L. and Task Force on Statistical Inference APA Board of Scientific Affairs (1999) Statistical Methods in Psychology journals: Guidelines and explanation. *American Psychologist*, 54, 594-604.