COMPARACIÓN DE VARIABLES DE DISTRIBUCIÓN T: UNA APLICACIÓN EN LA DIFERENCIA DE GRUPOS PARA LA VALIDEZ DE CONSTRUCTO

COMPARISON OF T DISTRIBUTION VARIABLES: AN APPLICATION ON THE DIFFERENCE OF GROUPS TO CONSTRUCT VALIDITY

César A. Merino Soto* y Victor Willson**
Universidad de San Martín de Porres, Perú.
Texas A&M University, EE.UU.

Recibido: 05 de abril de 2012 Aceptado: 22 de abril de 2013

RESUMEN

Se presenta una técnica estadística para comparar resultados de la aplicación de pruebas estadísticas basados en distribuciones *t*. Esta técnica se aplica mediante un programa MS Excel, se incluye un indicador de la magnitud del efecto (d de Cohen), el mismo que es parte de las actuales recomendaciones de publicación de resultados estadísticos. Para la demostración de la técnica se usó la información publicada en tres estudios empíricos y se compararon las variables de distribución *t* de Student obtenidas. Los resultados fueron consistentes con la información publicada. Finalmente, se discute las ventajas y las condiciones metodológicas de la apropiada aplicación de la técnica en el contexto de la investigación psicométrica y no psicométrica en psicología y educación.

Palabras clave: *t de Student*, análisis estadístico, validez de constructo, metodología.

ABSTRACT

It is a statistical technique to compare results of the application of statistical tests based on *t* distributions. This technique is applied using a MS Excel program; an indicator of the magnitude of the effect (Cohen d) is included, which is part of the current recommendations for publication of statistical results. For a demonstration of the technique the information published in three empirical studies was used and Student's *t* distribution variables obtained were compared. The results were consistent with the posted information. Finally, the advantages and methodological conditions of the proper application of the technique in the context of the psychometric and not psychometric research in psychology and education is discussed.

Key words: *Student's t*, statistical analysis, construct validity, methodology

Una de las estrategias de análisis en la investigación cuantitativa es comparar los resultados de dos (o más) grupos mediante algún estadístico sumario obtenido de los datos, por ejemplo: medias, varianzas, medianas, proporciones o distribuciones enteras. En la línea de la investigación experimental más básica que es el diseño de comparación de dos grupos (Brown & Melamed, 1990), las técnicas paramétricas o no paramétricas para evaluar la diferencia de medias o de la variabilidad son las más utilizadas.

Tradicionalmente, estas comparaciones recurren a técnicas paramétricas o no paramétricas teniendo en cuenta si los grupos son dependientes o independientes, pero la

* sikayax@yahoo.com.ar

ISSN: 1729-4827 (Impresa)
ISSN: 2233-7666 (Digital)

elección de estas puede tener algunas complicaciones insospechadas por el usuario cuando sobrevalora el nivel de medición sin considerar los problemas conceptuales en ellas (Michell, 1986; Velleman & Wilkinson, 1993; Wright, 1997). Luego de decidir por una comparación de dos grupos mediante una técnica paramétrica, los resultados estadísticos se interpretan mediante la significancia estadística y la magnitud de las diferencias. De todos los procedimientos estadísticos de comparación, uno que es aparentemente menos utilizado es la comparación de las diferencias obtenidas en dos grupos de datos, que provienen a su vez de situaciones experimentales distintos. Analizar estas diferencias de las diferencias es el objetivo de la técnica presentada aquí, por lo tanto los objetivos del presente trabajo son: a) presentar un procedimiento cuantitativo para comparar variables de distribución *t*, b) presentar un programa informático para su implementación, y, c) discutir las condiciones apropiadas para su uso e interpretación.

En la inferencia estadística, el procedimiento básico para proceder a aplicar el modelo estadístico es obtener una forma de estandarizar relaciones que puedan ser comparadas con las condiciones en que ocurre un efecto aleatorio. Esta transformación estandariza los datos empíricos para aplicar alguna prueba estadística inferencial. Por ejemplo, cuando se quieren comparar coeficientes de validez obtenidos mediante correlaciones *r de Pearson*, se puede aplicar una prueba asintótica z para muestras grandes (Chen & Popovich, 2002), o modificaciones de ella para muestras pequeñas (Braden, 1986; Rasmussen, 1988). Para aplicar esta prueba z, primero se transforman las correlaciones en valores *z*, con lo que se logra una distribución de muestreo de *r* menos asimétrica (Chen & Popovich, 2002). Este mismo procedimiento sirve para crear intervalos de confianza alrededor del parámetro correlacional; sin embargo, la distribución *z* no es la única que se puede aplicar en este tipo de transformaciones.

En la situación que se aborda, la tarea es comparar variables *t* obtenidas en el marco de pruebas de hipótesis. Por ejemplo, en la situación de coeficientes de regresión, se prueba que $b = 0$ aplicando una prueba tipo *t* (Chen & Popovich, 2002; Willson, 1983), así como al comparar correlaciones en muestras dependientes (Chen & Popovich, 2002) o varianzas para muestras relacionadas (Pitman, 1939). Y en la comparación de medias provenientes de grupos independientes o dependientes, se elige la popular *t de Student*.

Una de las estrategias de validez de constructo es la comparación de grupos (Cronbach & Meehl, 1955; Thorndike, 1989), y para ello frecuentemente se usa la prueba *t de Student* para comparar medias. Asumiendo que se desea probar la validez de la medida *X* bajo esta estrategia diferencial, si en el grupo *n* se obtiene una diferencia estadísticamente significativa que compara las medias de dos grupos independientes *A* y *B*, entonces es posible confiar en términos probabilísticos que las diferencias no ocurrieron por error de muestreo (Cohen, 2001). Pero, si ocurre que en otra muestra *m* se hizo la misma comparación entre las medidas independientes *A* y *B*, y se aceptó la hipótesis nula, ¿Se podría concluir directamente que el método *X* fue más válido (diferenció mejor) en la muestra *n* y no en la muestra *m*? Esta pregunta no puede ser respondida por el procedimiento *t* de Student, ya que se examinaron las diferencias (entre *A* y *B*) en dos análisis que no los relacionan empíricamente. El investigador solo puede hacer un juicio

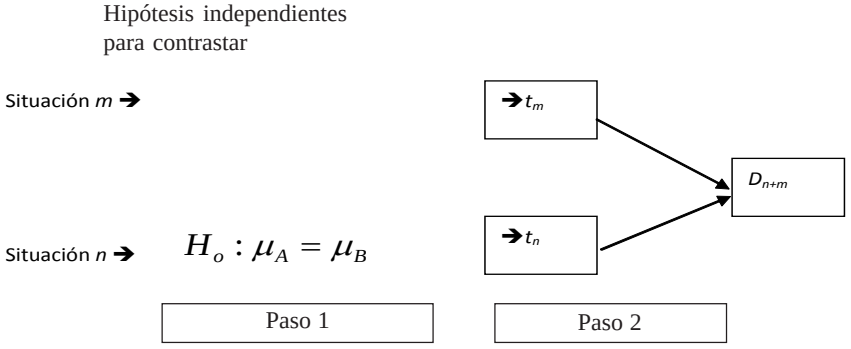

Figura 1. Comparación de grupos A y B dentro de dos situaciones m y n: secuencia de análisis

* sikayax@yahoo.com.ar

ISSN: 1729-4827 (Impresa)
ISSN: 2233-7666 (Digital)

basado en la lógica pero no probando formalmente la diferencia de las diferencias halladas. La Figura 1 ilustra esta situación.

Comparación de variables t

Willson (1983) comunicó una técnica que compara dos variables de distribución *t* al reanalizar los resultados de Levy (1983), quien presentó una versión adaptada para niños autistas sobre la Prueba de Vocabulario de Figuras de Peabody-PPVT (Dunn, 1965). Pero la técnica estadística originalmente fue expuesta por Ghosh (1975). Levy (1983) demostró que los niños autistas que respondieron la versión adaptada del *PPVT* ganaron puntos cuando los mismos respondieron a la versión original del *PPVT*, mientras que los niños normales no variaron sus puntajes. El desempeño de los niños autistas y normales en las dos versiones fueron independientemente comparados usando la *t de Student* para muestras relacionadas, y Levy determinó el respaldo a su versión por las diferencias estadísticamente significativas en los niños autistas. Wilson cuestionó su metodología estadística ya que respondió a una pregunta que comparaba simultáneamente diferencias, con un enfoque que comparaba independientemente las diferencias entre grupos.

Para aplicar este procedimiento se siguen los siguientes pasos:

1. Obtener resultados de prueba *t de Student* desde la comparación de grupos independientes, cada par de grupos separadamente. Esto puede referirse como para comparación 1 (variable t_1) y la comparación 2 (variable t_2).
2. Estimar la diferencia D, entre en las variables t_1 y t_2, obtenidas en el paso 1.
3. Estimar *t* crítico para usarlo en el contraste para *D*, usando los grados de libertad *n* y *m* asociados a las variables t_1 y t_2, respectivamente. Elegir el nivel alfa que usualmente es 0.05 o 0.01.
4. Estimar D_{n+m}, que es el *D* crítico o valor nulo con el cual comparar el valor *D* del paso 2. Esta se obtiene con los siguientes ecuaciones

$$D_{n+m} = t_\alpha \sqrt{2}\left[1 + \frac{1}{n}R_{1t} + \frac{1}{n^2}R_{2t} + \frac{1}{n^3}R_{3t}\right] \quad (1)$$

.*n* = grados de libertad para la variabnle de distribución *t*
.*m* = grados de libertad para la variable m de distribución *t*

Donde,

$$\lambda = \frac{n}{m} \quad (2)$$

$$R_{1t} = \left(\frac{1+\lambda}{16}\right)\left(t_\alpha^2 + 5\right) \quad (3)$$

$$R_{2t} = \left(\frac{1+\lambda^2}{1536}\right)\left(37t_\alpha^4 + 200\,t_\alpha^2 + 171\right) - \left(\frac{\lambda}{256}\right)\left(9t_\alpha^4 - 24t_\alpha^2 + 7\right) \quad (4)$$

$$R_{3t} = \left(\frac{1+\lambda^3}{8192}\right)\left(81t_\alpha^6 + 349t_\alpha^4 - 293t_\alpha^2 - 1153\right) - \left(\frac{\lambda(1+\lambda)}{24576}\right)\left(231t_\alpha^6 + 773t_\alpha^4 - 499t_\alpha^2 - 2871\right) \quad (5)$$

* sikayax@yahoo.com.ar

ISSN: 1729-4827 (Impresa)
ISSN: 2233-7666 (Digital)

Aplicaciones

La investigación psicométrica y no psicométrica puede ser enriquecida con este enfoque, pero la disposición inicial para percibir la aplicación de esta técnica es la condición esencial. Para ejemplificar mejor el uso de la técnica, se usa la información normativa publicada por Merino, Cohen y Díaz (2003), respecto a una medida de autorreporte de crianza para niños. Merino et al. (2003) reportaron evidencias de validez de constructo para el Inventario de Percepción Parental - IPP (Hazzard, Christiensen & Margolin, 1983) en niños peruanos, la misma que contiene dos escalas evaluativas: Crianza Positivo y Crianza Negativo, y dos objetivos: Padre y Madre, independientes. Se examinaron las diferencias entre niños de diferente género y grado escolar, todos provenientes de un colegio privado. Una de las hipótesis examinadas fue sobre las diferencias entre la percepción de la crianza positiva en el padre y la madre, ambas reportadas por niños y niñas independientemente. Enfocándose en el trato positivo de ambos padres (Escala Madre Positiva y Padre Positivo), la pregunta es si el trato de las madres y de los padres varía en relación al género del niño. Los niños pueden reportar que las madres pueden ser más punitivas hacia ellos y esto podría estar cuantificado en la significancia estadística hallada luego de usar una *t de Student* para muestras independientes; esta hipótesis también se puede explorar para las niñas y luego explorar también en la percepción del trato positivo de los padres. Pero ambos contrastes son realizados independientemente y así no responden directamente a la hipótesis básica sobre las diferencias de sexo en el trato recibido. Esta situación planteada es similar a la planteada por Levy (1983), aunque este autor usó una prueba *t* para muestras dependientes; pero nuestro ejemplo y la situación de Levy requieren una comparación de las diferencias derivadas de la *t de Student*.

Usando los datos publicados por Merino et al. (2003), en la escala Padre Positivo, los varones reportaron similar percepción que las mujeres en el trato positivo recibido por el padre ($t[274] = 0.00$, $p = 1.00$, *d Cohen* = 0.00), y en la escala Madre Positiva ocurrió aparentemente la misma tendencia, ($t[274] = 1.49$, $p = 0.13$, *d Cohen* = -0.18). Aplicando el enfoque propuesto en este artículo, se obtuvo el *D* calculado (-1.49) y el *D* de contraste, $D_{n+m}(548) = 3.134$, $p < 0.01$, $d = -0.13$. La significancia hallada sugiere que las diferencias en el trato ocurren en la madre y no el padre, siendo las primeras más positivas hacia las niñas que hacia los niños; estas diferencias son, sin embargo pequeñas y posiblemente de escaso valor práctico. Esta misma conclusión se obtuvo en el trabajo de Merino et al. (2003), pero tales autores poseían los datos completos.

Un segundo ejemplo se explica desde una investigación que evaluó la estructura factorial de una Escala de Autoconcepto de Habilidades Académicas (Merino & Díaz, 2003), originalmente elaborado por Dayton (1968). En el estudio de adaptación, Merino y Díaz (2003) reportan las diferencias entre varones y mujeres en cada una de las 7 subescalas, y estas comparaciones hallaron tres diferencias más allá del error de muestreo en las subescalas Historia ($t[88] = -2.36$, $p < 0.05$, $d = -0.50$), Música ($t[61.3] = -2.94$, $p < 0.01$, $d = -0.69$) y Habilidad General ($t[95] = -2.02$, $p < 0.05$, $d = -041$); todas las comparaciones favorecieron a las mujeres. Estos autores no protegieron sus comparaciones contra el error Tipo I en la situación de múltiples comparaciones (Brown & Melamed, 1990), dado que se hicieron seis comparaciones *t de Student* (seis escalas) efectuadas entre varones y mujeres, y sus conclusiones sobre la significancia estadística de sus resultados podrían ser diferentes a los publicados. Para avanzar un paso más en el análisis, se plantea la hipótesis si las diferencias de género fueron mayores en un área que en otra. Esta pregunta no se responde formalmente por los análisis de Merino y Díaz (2003), así que se hará uso de la técnica propuesta aquí. Obteniendo un *D* calculado entre (0.58) y el *D* de contraste, ($D_{n+m}[149.3] = 3.25$, $p = 0.56$), se concluye que los varones y mujeres difieren en la misma magnitud en ambas escalas ($d = 0.09$). Comparando las diferencias en las subescalas con aquellas reportadas en el puntaje total, estas tampoco fueron sustancialmente discrepantes: para Historia, ($D[183] = 0.34$, $p = 0.73$, $d = 0.05$) y para Música, ($D[149.3] = 0.94$, $p = 0.35$, $d = 0.15$). Una conclusión más general puede ser que las diferencias entre las mujeres obtienen puntajes moderadamente superiores a los varones en ambas escalas, pero estas diferencias son similares entre sí. Se debe anotar que, esta similaridad de las variables *t* pudo ser previamente reconocida mediante las estimaciones de magnitud del efecto (*d* de Cohen) reportadas en esta investigación.

Un último ejemplo proviene de los resultados descriptivos reportados por Ruiz, Godoy y Gavino (2008) respecto a la evaluación psicométrica del Cuestionario de

* sikayax@yahoo.com.ar

ISSN: 1729-4827 (Impresa)
ISSN: 2233-7666 (Digital)

Creencias Obsesivas (OCCWG, 2005). Ruiz et al. (2008) no hallaron diferencias estadísticamente significativas entre varones y mujeres, ni entre universitarios y comunidad general, y tales hallazgos son consistentes con otros estudios respecto a la similaridad de los puntajes entre estos grupos de adultos en constructos asociados (Fullana et al., 2005; Rasmussen, 1988; Rodríguez-Albertus, Godoy & Gavino, 2008). Ruiz et al. (2008) no describieron la prueba estadística para las diferencias, ni sus resultados cuantitativos, pero afortunadamente reportaron estadísticos descriptivos básicos para la muestra universitaria y de adultos de la comunidad en general. Una hipótesis nula adicional que se pudo probar es si las diferencias halladas entre estudiantes (n = 247) y la comunidad (n = 395) en la subescalas *Responsabilidad-Estimación de la Amenaza* (*OBQ-RH*) y en *Perfeccionismo-Intolerancia* a la *Incertidumbre* (*OBQ-PC*) son cero. Las diferencias entre los grupos en *OBQ-RH* ($t_{602.29}$ = 1.01) y *OBQ-PC* ($t_{581.64}$ = 1.47) no fueron estadísticamente significativas; luego, la comparación de estas diferencias usando los resultados de ambas *t de Student* fue D = 0.46 (gl = 1183.94), p = 0.64, d = 0.02, que indica que las diferencias entre los grupos en ambos puntajes son apenas distintas.

El programa

Las fórmulas para estimar $D_{n\text{-}m}$ parecerán difíciles para el usuario, sin embargo, actualmente se ha creado un programa en MS Excel que funcionará en toda computadora que tenga instalado este componente de MS Office. El usuario requiere el ingreso de los resultados paramétricos de la *t de Student* aplicada a dos muestras (dependientes o independientes), así como el nivel *alfa* o error Tipo I, y los grados de libertad asociados a cada variable *t*. El valor *alfa* puede determinarse en los tradicionales valores 0.01 o 0.05

Consistente con las actuales exigencias de comunicación de resultados estadísticos (American Educational Research Association, 2006; Cumming & Finch, 2005; Fidler, 2002; Fidler, Cumming, Burgman, Thomason, 2004; Thompson, 1996; Thompson, 2002; Wilkinson & APA Task Force on Statistical Inference, 1999), el programa también reporta intervalos de confianza y un estimador de la magnitud del efecto (d : Cohen, 1988), este último basado en una transformación desde el D calculado, mediante la propuesta de Rosenthal y Rosnow (1991) para traducir una variable t hacia la d de Cohen para grupos de desigual tamaño:

$$d = \frac{t\left(n_1 + n_2\right)}{\sqrt{gl}\sqrt{n_1 n_2}} \qquad (6)$$

Con esta última información, se puede interpretar el grado en que ocurren las diferencias entre las variables t obtenidas; Cohen (1988) sugirió una flexible propuesta para los valores recomendados en la interpretación cualitativa de las diferencias: 0.20 (pequeño), 0.50 (moderado) y 0.80 (grande).

Anotaciones finales

La aplicación de esta técnica no garantiza óptimos resultados si se violan los presupuestos estadísticos para la aplicación de las técnicas paramétricas. La independencia de las observaciones en la muestra de estudio debe garantizarse empírica o conceptualmente, ya que las probabilidades deducibles de las distribuciones teóricas se respaldan en el muestreo de variables aleatorias. Otra condición necesaria pero no suficiente para la aplicación de la prueba *t de Student* es el presupuesto de distribución normal de los datos (Cohen, 2001; Howell, 1997; Kiess, 1996), una característica infrecuente en la investigación psicológica, incluso en muestras grandes (Micceri, 1989). Mientras que el alejamiento de la normalidad no sea severo, se puede asumir que no será un problema serio (Cohen, 2001; Micceri, 1989). La transformación de las variables hacia un mejor ajuste de la normalidad podría ser una opción razonable que el investigador debe evaluar antes de aplicar la técnica.

Por otro lado, para enfrentar la sensibilidad de la *t de Student* a la violación de algunos o todos sus presupuestos, se han estudiado ampliamente otras opciones; estas versiones de *t de Studen*t para superar problemas de heterogeneidad de varianza (Keselman, Wilcox, Taylor & Kowalchuk, 2000; Lix & Keselman, 1998; Ruxton, 2006) o para detectar los cambios asimétricos en la comparación de dos grupos independientes (Balkin & Mallows, 2001) podrían utilizarse para aplicar la técnica recuperada en este artículo.

* sikayax@yahoo.com.ar

ISSN: 1729-4827 (Impresa)
ISSN: 2233-7666 (Digital)

Otro aspecto que merece discusión y planificación es la presencia del error de medición en las mediciones. Si el error de medición en psicología es el talón de Aquiles de la información recolectada (Pedhazur & Schmelkin, 1991), su impacto debe ser previamente evaluado por el investigador, dado el sesgo producido sobre la significancia estadística (Charter, 1997) y en las diferencias estimadas en general (Feldt & Brennan, 1989; Stanley, 1971). Esta evaluación es recomendada por los Estándares de Medición (American Educational Research Association, American Psychological Association and National Council on Measurement in Education, 1999) y su conocimiento por parte del usuario debe darle un marco de moderación para interpretar de las diferencias halladas. De este modo, el usuario debe observar las estimaciones de confiabilidad en las mediciones aplicadas. Si esta información no aparece explícitamente, se puede recurrir a estimarla por un método propuesto por Magnuson (1990), pero esta práctica es solo aproximativa y teórica (Onwuegbuzie & Daniel, 2004) de los valores muestrales de confiabilidad que debería ser publicada por los autores.

Estando de acuerdo con Willson (1983), este procedimiento es menos poderoso comparado con un ANOVA de dos vías cuando se tienen los datos completos. Pero mayormente solo se tienen resultados publicados. Si los datos publicados contienen la información de tendencia central y dispersión, o únicamente los resultados de la prueba *t*, entonces la aplicación de la técnica no debe ser un problema para el investigador.

Este contexto es típico en estudios meta-analíticos, así que el principal requerimiento es que el investigador pueda percibir la posibilidad de aplicar la técnica y expresarlo en su planteamiento del problema. Si el planteamiento del problema y el diseños son correctos, entonces los resultados de esta técnica (y de cualquier otra en general), podrán ser interpretables con confianza. Por tal motivo, ni la complejidad computacional ni las soluciones estadísticamente elegantes pueden reemplazar el juicio del investigador para decidir y percibir la mejor situación del uso de técnicas como la presentada en este artículo.

Referencias

American Educational Research Association (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher, 35*(6), 33-40.

American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Balkin, S. D. & Mallows, C. L. (2001). An adjusted, asymmetric two-sample t test. *The American Statistician, 55*(3), 203-206.

Braden, J. P. (1986). Testing correlations between similar measures in small samples. *Educational and Psychological Measurement, 46,* 143-148.

Brown, S. R. & Melamed, L. E. (1990). *Experimental design and analysis*. Thousand Oaks, CA: Sage.

Charter, R. A. (1997). Effect of measurement error on test of statistical significance. *Journal of Clinical and Experimental Neuropsychology, 19*(3), 458-462.

Chen, P. Y. & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures*. Thousand Oaks, CA: Sage.

Cohen, B. H. (2001). *Explaining psychological statistics.* (2nd. ed.). New York: John Wiley & Sons.

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

Cronbach, L. J. &Meehl, P. E. (1955). Construct validity inpsychological tests. *Psychological Bulletin, 52*, 281-302.

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

Dayton, M. C. (1968). *Self-concept of ability scale: Technical manual*. Research and Demonstration Center of the Interprofessional Research Commission on Pupil Personnel Services, University of Maryland.

Dunn, L. (1965). *Peabody Picture Vocabulary Test.* Circle Pines, MN: American Guindance Service.

Feldt, L. S. & Brennan, R. L. (1989). Reliability. In R. Linn (Ed.), *Educational Measurement* (pp. 105-146). New York: Macmillan.

Fidler, F. (2002). The 5th edition of the APA Publication Manual: Why its statistics recommendations are so controversial. *Educational and Psychological Measurement, 62*, 749-770.

Fidler, F., Cumming, G., Burgman, M. & Thomason, N. (2004). Statistical reform in medicine, psychology and ecology. *Journal of Socio Economics, 33*, 615-630.

Fullana, M. A., Tortella-Feliu, M., Caseras, X., Andión, O., Torrubia, R. & Mataix-Cols, C. (2005). Psychometric properties of Spanish version of the Obsessive-Compulsive Inventory-Revised in a non-clinical sample. *Journal of Anxiety Disorders, 19,* 893-903.

* sikayax@yahoo.com.ar

ISSN: 1729-4827 (Impresa)
ISSN: 2233-7666 (Digital)

Ghosh, K. B. (1975). On the distribution of the difference of two t-variables. *Journal of the American Statistical Association, 70,* 463-469.

Hayes, A. F. & Cai, L. (2007). Further evaluating the conditional decision rule for comparing two independent means. *British Journal of Mathematical and Statistical Psychology, 60*(2), 217-244.

Hazzard, A., Christiensen, A. & Margolin, G. (1983). Children's perceptions of parental behaviors. *Journal of Abnormal Child Psychology, 11*(1), 49-60.

Howell, D. C. (1997). *Statistical methods for psychology.* (4th. ed.). Belmont, CA: International Thomson Publishing.

Keselman, H. J., Wilcox, R. R., Taylor, J. & Kowalchuk, R. K. (2000). Tests for mean equality that do not require homogeneity of variances: Do they really work? *Communications in Statistics: Simulation and Computation, 29*(3), 875-895.

Kiess, H. (1996). *Statistical concepts for Behavioral Sciences.* (2nd ed.). Boston: Allyn & Bacon.

Levy, S. (1983). Use of the Peabody Picture Vocabulary Test with low-functioning autistic children. *Psychology in Schools, 19,* 24-27.

Levy, S. (1983). Interpreting the differences: A reply to Willson. *Psychology in the Schools, 20,* 252-253.

Lix, L. M. & Keselman, H. J. (1998). To trim or not to trim: Tests of mean equality under heteroscedasticity and nonnormality. *Educational and Psychological Measurement, 58,* 409-429.

Magnus, J. R. (2000). On the sensitivity of the t-statistic. In A. W. A. Ullah & A. Chaturvedi (Ed.), *Handbook of Applied Econometrics and Statistical Inference*. New York: Marcel Dekker.

Magnusson, D. (1990). *Teoría de los tests*. México, D.F.: Trillas.

Merino, C. & Díaz, M. (2003). Validez de constructo y confiabilidad de la Escala de Autoconcepto sobre las Habilidades de M. C. Dayton. *Revista de Investigación en Psicología, 6*(2), 68-79.

Merino, C., Cohen, B. & Diaz, M. (2003). De los niños a los padres: El inventario de percepción de conductas parentales. *Personas, 6,* 135-149.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*, 156-166.

Michell, J. (1986). Measurement scales and statistics: a clash of paradigms. *Psychological Bulletin, 100,* 398-407.

OCCWG (2005). Psychometric validation of the Obsessive Beliefs Questionnaire and the Interpretation of Intrusions Inventory-Part 2: Factor analyses and testing of a brief version. *Behaviour Research and Therapy, 43,* 1527-1542.

Onwuegbuzie, A. J. & Daniel, L. G. (2004). Reliability generalization: The importance of considering sample specificity, confidence intervals, and subgroup differences. *Research in the Schools, 11*(1), 61-71.

Pedhazur, E. J. & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Pitman, E. J. G. (1939). A note on normal correlation. *Biometrika, 31*, 9-12.

Rasmussen, J. L. (1988). Evaluation of small-sample statistics that test whether variables measure the same trait. *Applied Psychological Measurement, 12,* 177-187.

Rodríguez-Albertus, M., Godoy, A. & Gavino, A. (2008). Propiedades psicométricas de la versión española del Inventario de Interpretación de Intrusiones (III). *Ansiedad y Estrés, 14*(2-3), 187-198.

Rosenthal, R. & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis.* (2nd ed.). New York: McGraw Hill.

Ruiz, C., Godoy, A. & Gavino, A. (2008). Propiedades psicométricas de la versión española del Cuestionario de Creencias Obsesivas (OBQ). *Ansiedad y Estrés, 14*(2-3), 175-185.

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. *Behavioral Ecology, 17*(4), 688-690.

Sarle, W. S. (1995). Measurement theory: Frequently asked questions. *Disseminations of the International Statistical Applications Institute* (4th ed.), Vol.1, (pp. 61-66). Wichita: ACG Press.

Stanley, J. (1971). Reliability. In R. Thorndike (Ed.), *Educational measurement* (pp. 356-442). Washington, DC: American Council on Education.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggestedreforms. *Educational Researcher, 25*(2), 26-30.

Thompson, B. (2002). «Statistical», «practical» and «clinical»: How many kinds of significance do counselors need to consider? *Journal of Counseling and Development, 80,* 64-71.

Thorndike, R. L. (1989). *Psicometría aplicada*. México, D. F.: Limusa.

Velleman, P. F. & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *The American Statistician, 47,* 65-72.

Wilkinson, L. & APA Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594-604.

Willson, V. L. (1983). A test for the differences between two groups t-distributed statistics. *Psychology in Schools, 20,* 250-251.

Wright, B. D. (1997). S. S. Stevens Revisited. *Rasch Measurement Transactions 11*, 552-553.

* Instituto de Investigación de Psicología, Universidad de San Martín de Porres, Perú.
** Texas A & M University, EE.UU.

* sikayax@yahoo.com.ar

ISSN: 1729-4827 (Impresa)
ISSN: 2233-7666 (Digital)