

**SIGNIFICANCIA ESTADÍSTICA EN LA INVESTIGACIÓN  
PSICOEDUCATIVA**

por

Pedro A.Sánchez Escobedo, PhD  
Alvaro J. Arce Ferrer, PhD  
Facultad de Educación  
Universidad Autónoma de Yucatán

**SINOPSIS**

El presente trabajo discute la validez relativa de la significancia estadística como criterio e relevancia para evaluar los resultados de la investigación psicoeducativa. Se analizan los conceptos de significancia estadística y significancia práctica y se argumenta que la evaluación de los métodos y procedimientos de la investigación otorgan criterios de mayor utilidad que la significancia estadística para evaluar los resultados de la investigación. Se compara el concepto de significancia práctica en los paradigmas diferenciales de investigación cualitativa y cuantitativa.

**ABSTRACT**

This paper discusses the relative validity of statistical significance as a relevant criterion to assess results yielded by educational research. By differentiating between the concepts of statistical significance and practical significance support is given to the idea that the assessment of the procedures and methods is a better criterion than the statistical significance to evaluate research results.

## Significancia estadística en la investigación psicoeducativa

Afortunada o desafortunadamente, en la mayoría de las revistas especializadas de psicología y educación, uno de los criterios más importantes para la publicación de la investigación ha sido el análisis de la significancia estadística de los resultados. Es decir, los resultados de una investigación son o no, aceptados por la comunidad educativa si estos superan el umbral preestablecido por la probabilidad.

Esta situación, posiblemente ha influido en los investigadores de estas áreas, quienes con frecuencia perciben excesivas dificultades en el procesamiento de datos y análisis cuantitativo de sus trabajos empíricos así como en la validación de sus resultados.

Existe la creencia generalizada que una buena investigación científica debe producir resultados que sean estadísticamente significativos.

En lo que respecta al procesamiento y análisis de los datos, los procedimientos se han facilitado significativamente con la incorporación de avances tecnológicos en las escuelas y centros de investigación, principalmente la accesibilidad a computadoras y programas de gran potencia. Por ejemplo, la realización de un análisis multivariado factorial en los sesentas, requería de varios ayudantes de investigación que hicieran los cálculos durante varias semanas, mientras que hoy en día una computadora personal con un programa estadístico comercial de mediana complejidad puede hacer este análisis en cuestión de minutos.

En virtud de esta simplificación en el procesamiento y análisis de datos, el investigador puede dirigir ahora sus esfuerzos a cuestionar otros aspectos – de índole cualitativa – de la investigación científica que influyen considerablemente en la calidad

e importancia de una investigación tales como el diseño de la misma, la validez de los instrumentos, la fortaleza de su base conceptual, entre otros.

El presente trabajo analiza diferentes limitaciones de la significancia estadística como criterio para evaluar los resultados de la investigación psico-educativa y sugiere que debe reconsiderarse la importancia otorgada hasta hoy a la significancia estadística en esta área. Asimismo, se argumenta que la significancia estadística en la empresa científica debe evaluarse con base en su influencia para la generación de conocimientos nuevos, y en su papel para el desarrollo, comprobación o rechazo de teoría en psicología y educación. Se defiende la premisa de que la consideración de la metodología global empleada es un criterio de mucho mayor utilidad que el criterio de significancia estadística en la evaluación de la investigación en esta área.

Antes de analizar las limitaciones de este criterio en la evaluación de la investigación psico-educativa, se definirá el concepto de significancia estadística a fin de diferenciarlo de otro con el que comúnmente se le confunde: la significancia práctica.

La significancia estadística es un concepto matemático que se refiere a las probabilidades de que un resultado específico se deba -o no- al azar. Por ejemplo, cuando el resultado de una prueba estadística arroja una significancia  $\leq .01$ , esto quiere decir que existen una en cien posibilidades de que este resultado se deba al azar. ¿Qué nos dice esta significancia estadística de la significancia práctica de este resultado?. La respuesta es NADA (Felt, 1990). La significancia práctica se refiere a la utilidad de este concepto para refinar una teoría determinada a la capacidad de los resultados de aportar conocimientos nuevos o al efecto que tienen éstos en aumentar o disminuir la confianza del investigador en determinada explicación de un evento observado. Es decir, la

significancia práctica de los resultados de una investigación se juzga no con base en criterios matemáticos, sino que con base en su utilidad práctica, importancia empírica, base y aceptación teórica y comparación con otras investigaciones similares.

Pese a las enormes diferencias entre los conceptos de significancia práctica y estadística, la significancia estadística continua siendo un criterio pivotal para la aceptación de trabajos y la publicación de reportes de investigación por lo que conviene revisar las ventajas y desventajas de este criterio y reflexionar acerca de las consecuencias de su uso.

La crítica hacia la consideración de la significancia estadística como criterio principal de la calidad y/o efectividad de una investigación no es nueva. En 1968, Lykken publicó un interesante artículo acerca de la significancia estadística en la investigación psicológica que cuestionaba las prácticas existentes para evaluar los resultados del trabajo de investigación.

Lykken en su artículo, analizó un estudio publicado por Sapolsky (1964), quien proponía la “teoría cloacal del nacimiento”. En síntesis, Sapolsky proponía que los pacientes que presentan problemas de alimentación, como anorexia y bulimia, presentaban una tendencia a la impregnación oral, manifestada por comer compulsivamente aquellos que desean embarazarse y por anorexia aquellos que no. Sapolsky propuso que estos pacientes estarían inclinados a ver animales con cloaca. Como ranas en el test de Rorschach. Es decir la teoría argumentaba que los pacientes con desórdenes alimenticios mantenían una creencia inconsciente de impregnación oral. Para probar su teoría, Sapolsky administró la prueba de Rorschach a un grupo **de**

¿¿¿?? En un Hospital psiquiátrico. 19 de los 31 que visualizaron una rana en el test tenían, de hecho, algún desorden alimenticio comparados con los 5 del grupo control. El análisis estadístico de los grupos reportó una chi cuadrada significativa.

Lykken, dudando de la validez teórica del postulado de Sapolsky encuestó a 20 colegas quienes estimaron nula o improbable la factibilidad de la teoría cloacal de Sapolsky, aún después de haber analizado los resultados estadísticamente significativos de su investigación. Lykken sugirió entonces que el criterio del comité editorial del *Journal of Consulting Psychology* –una revista de psicología con reconocido prestigio– adolecía de igualar la significancia estadística con la significancia práctica. Es decir, los miembros del comité editorial, con base en una Chi cuadrada significativa, estuvieron dispuestos a creer debido a la significancia estadística de la prueba, que la impregnación oral lleva a un sujeto a visualizar ranas en el Rorschach, simplemente porque las ranas tienen cloaca (cosa que muchos pacientes ignoran).

Lykken propuso como criterio alternativo para juzgar una investigación la siguiente regla: “Cuando una predicción o hipótesis derivada de una teoría es confirmada mediante un experimento, debe resultar en un incremento no-trivial de la confianza en esta teoría” (p.152) y sugirió la corroboración múltiple y la repetición constructiva, que pone a prueba los métodos utilizados, como alternativas para probar una teoría. Aduciendo a la significancia práctica, Lykken concluyó: “El valor de una investigación puede determinarse más que por los resultados estadísticos, por la evaluación habilidosa y subjetiva de la coherencia y racionalidad de la teoría, del grado

de control experimental empleado, de la sofisticación de las técnicas de medición y por la importancia práctica o científica del fenómeno estudiado” (p. 158-59).

Antes de elaborar más acerca de estos criterios alternativos para juzgar la investigación psico-educativa en términos cualitativos más que cuantitativos, enumeraremos algunas consideraciones que deben tenerse en cuenta al analizar la parte cuantitativa de la investigación.

En primer término, para juzgar la significancia debe establecerse el poder estadístico de la prueba, el cual depende sobre todo del tamaño de la muestra y en menor medida del nivel alfa preestablecido (Siegel, 1990). Una prueba pudiera arrojar un resultado estadísticamente significativo al  $\alpha = .0001$ , sin embargo, tener un poder estadístico bajo. Pese a la importancia de este criterio, ¡el poder estadístico casi nunca se reporta o calcula en las investigaciones!.

En segundo lugar, debe considerarse la posibilidad de la existencia de un error de tipo I (rechazar la hipótesis nula cuando es verdadera), ya que la probabilidad de la existencia de este es un resultado específico, independientemente del valor alfa, es de .5 (Forsyth, 1988). Es decir, la probabilidad de error tipo I en una decisión estadística es de .5. Por otra parte, muchas de las teorías en psicología y educación predicen solamente la dirección de las relaciones entre variables, las diferencias entre grupos o el efecto de un tratamiento. De hecho, ya que la hipótesis nula nunca es estrictamente verdadera, estas predicciones estadísticas poseen un 50% de probabilidades de ser confirmadas aún cuando en realidad la hipótesis nula sea verdadera (Lykken, 1968; Forsyth, 1988).

Asimismo, debe considerarse que las pruebas estadísticas de uso común en la investigación psicoeducativa se basan en modelos que exigen el cumplimiento de supuestos tales como la igualdad de las varianzas entre los grupos comparados, la realización de observaciones independientes y la distribución normal de las medidas, entre otros. Sin embargo, dichos supuestos rara vez son verificados.

Por último, las variables de interés para la investigación psicoeducativa son reportadas con frecuencia como significativas. No obstante, este hallazgo debe tomarse con cautela ya que la mayor parte de éstas tienden a estar positivamente correlacionadas (su relación rara vez es totalmente aleatoria o cercana al cero).

Considerando las limitaciones antes mencionadas, es decir la falta de determinación del poder estadístico, la incertidumbre en la verificación de principios, la alta probabilidad de error tipo I y las características de las variables del área, llevan a la conclusión que se ha sobrestimado el valor de la significancia estadística como criterio de aceptación de resultados. Existe pues, la necesidad de expandir nuestros criterios de juicios para diferenciar una buena investigación psicoeducativa de una investigación de poca relevancia. Más allá del aspecto cuantitativo, existen criterios de índole cualitativa, que aunque subjetivos en su naturaleza, pudieran superar las limitaciones del enfoque matemático. Howe y Eisenhart (1988), sugieren explorar enfoques de investigación de índole cualitativa y afirman "...ya que toda la investigación científica es inherentemente interpretativa" (p.3). Howe y Eisenhart sugieren 5 estándares alternativos para juzgar la investigación psicoeducativa:

1. Congruencia entre las preguntas de investigación, la colección de los datos y las técnicas de análisis.

2. Aplicación efectiva de técnicas de análisis
3. Alerta y coherencia en los supuestos y bases del estudio.
4. El papel de la investigación en el mejoramiento de las prácticas educativas
5. Los métodos con los que se maneja los sujetos participantes en la investigación, es decir las prácticas éticas.

Es decir, más allá de conocer si una investigación arrojó o no resultados estadísticamente significativos a un nivel alfa predeterminado y arbitrario, debemos juzgar en forma global y extensiva los métodos, materiales y procedimientos empleados.

Cooper (1982) sugiere que la revisión de la metodología utilizada en un artículo y el análisis de cada una de las etapas del proceso, constituye en sí un proceso riguroso de indagación científica y afirma que la operacionalización de las variables, las funciones, las fuentes de varianza, la validez de los supuestos y las limitaciones del estudio son fundamentales de establecer cuando se analiza la calidad y trascendencia de los resultados de una investigación.

## **REFERENCIAS**

Cooper, H.M. (1982) Scientific guidelines for conducting integrative research reviews.

Review of Educational Research. 52(2), 291-302.

Felt, L. (1991). Experimental designs. Notes. The University of Iowa.

Forsyth, R.F. (1988) Introduction to Statistics. Mc Graw Hill.

Howe, K. & Eisenhart, M. (1988). Standards for qualitative (and quantitative) research.

A prolegomenon. In. J. Jager (Ed.) Complementary methods in research in

Education. Sage.

Lykken, D.T. (1968) Statistical significance in psychological research. Psychological Bulletin. /0(3). 131-159.

Sapolsky, A.Q. (1964) An effort at studying Rorschach content symbolism: The frog response. Journal of Consulting Psychology. 28, 469-472.

Siegel, S. (1990) Estadística no-paramétrica. México D.F: Trillas