

Confiabilidad y coeficiente Alpha de Cronbach

*Milton Quero Virla**

1. Generalidades de la confiabilidad

Dos características deseables en toda medición son la confiabilidad y la validez; al referirse a cualquier instrumento de medición en el campo de las ciencias sociales y de la conducta, se consideran estas dos cualidades como aspectos claves de la llamada “solidez psicométrica” del instrumento (Cohen y Swerdlik, 2001). En esta Nota nos ocuparemos fundamentalmente de la confiabilidad y específicamente del coeficiente α (alpha) de Cronbach.

Algunos autores encuentran cierta similitud entre las acepciones cotidianas y técnicas de la confiabilidad (Cozby, 2005; Kerlinger y Lee, 2002; Cohen y Swerdlik, 2001). Variando un poco el ejemplo dado por el primero de estos autores, en una pareja uno de los dos puede considerar altamente confiable al otro, en el sentido de la puntualidad, porque generalmente llega a las citas a la hora convenida. Vale comentar, que en esta situación un solo encuentro con retardo, aunque sea el primero de los encuentros, no permite estimar la confiabilidad. No se deceptione casi a priori!

La confiabilidad o fiabilidad, se refiere a la consistencia o estabilidad de una medida. Una definición técnica de confiabilidad que ayuda a resolver tanto problemas teóricos como prácticos es aquella que parte de la investigación de qué tanto error de medición existe en un instrumento de medición, considerando tanto la varianza sistemática como la varianza por el azar (Kerlinger y Lee, 2002). Dependiendo del grado en que los errores de medición estén presentes en un instrumento de medición, el instrumento será poco o más confiable.

A partir de estas consideraciones, los autores definen la confiabilidad como la ausencia relativa de errores de medición en un instrumento de medida. Expresado más explícitamente, un puntaje observado o medido es la suma de un puntaje real o verdadero más un puntaje de error o error de medición (Magnusson,

* Estudios doctorales en Ciencias de la Educación. MSc Estadística Aplicada. Licenciado Educación Mención Matemática y Física. Profesor de Educación Integral Especialidad Matemática. Profesor Agregado (Ordinario) del Departamento de Matemáticas de la Facultad Experimental de Ciencias, LUZ. Coordinador de la Maestría en Ciencias de la Educación Mención Gerencia Educativa URBE. Maracaibo, Venezuela. Correo electrónico: miltonqv@hotmail.com.

1978). El tratamiento dado a la confiabilidad sobre estas premisas se enmarca dentro de la llamada teoría clásica de las pruebas (Carmines and Zeller, 1979; Kerlinger y Lee, 2002; Magnusson, 1978). Un modelo alternativo a este de las puntuaciones reales o clásico, es el conocido como teoría de la generalización, atribuida a Cronbach en 1970 y a sus colegas en 1972, teoría basada en la idea de que las puntuaciones de una persona varían de una prueba a otra (Cohen y Swerdlik, 2001) rebelándose contra el concepto de una puntuación verdadera.

Diversas definiciones de la confiabilidad son dadas partiendo de ciertas características del instrumento de medición. Sin embargo, una característica común a varios de ellos es la de considerar que mientras la validez se refiere a que se mide lo que se desea medir, la confiabilidad se refiere a la exactitud con que un instrumento de medida mide lo que mide (Magnusson, 1978). Tan solo denota algo que es consistente, no necesariamente consistentemente bueno o malo, sino tan solo consistente (Cohen y Swerdlik, 2001).

2. Fuentes de Inconsistencia

Entre los factores que reducen la confiabilidad o consistencia de un instrumento de medición pueden considerarse (Nunnally, 1970): a. instrucciones no estandarizadas, las originadas por el encuestador si las instrucciones del instrumento de medición son dadas oralmente sin tener un conjunto estándar de instrucciones que pueda leerse invariablemente; b. errores en el registro de puntajes de respuesta, vinculadas a la transcripción “a mano” de los distintos puntajes de respuesta; c. errores debido al ambiente de medición, relacionados con la aplicación de los instrumentos en condiciones ambientales distintas (iluminación, niveles de ruido, confort, etc.) a diversos grupos de encuestados; d. errores debidos al muestreo de contenido, presentes en situaciones cuando el objetivo del instrumento es estimar un puntaje de los encuestados respecto al dominio de ciertos contenidos; e. errores debidos a fluctuaciones en los encuestados, estrechamente vinculados al individuo encuestado, como distracciones momentáneas, equivocación al momento de marcar una respuesta en la hoja, un lápiz o bolígrafo roto o defectuoso, etc.

3. El Coeficiente α (alpha) de Cronbach

La confiabilidad de una medición o de un instrumento, según el propósito de la primera y ciertas características del segundo, puede tomar varias formas o expresiones al ser medida o estimada: coeficientes de precisión, estabilidad, equivalencia, homogeneidad o consistencia interna, pero el denominador común es que todos son básicamente expresados como diversos coeficientes de correlación.

En el caso específico del coeficiente de confiabilidad vinculado a la homegenidad o consistencia interna, se dispone del coeficiente α (alpha), propuesto por Lee J. Cronbach (1916-2001) en el año 1951. Se ha demostrado que este coeficiente representa una generalización de las populares fórmulas KR-20 y KR-21 de consistencia interna, desarrolladas en 1937 por Kuder y Richardson (Kerlin-

ger y Lee, 2002), las cuales eran solo aplicables a formatos binarios de calificación o de respuesta (dicotómicas). Por lo tanto, con la creación del α de Cronbach, los investigadores fueron capaces de evaluar la confiabilidad o consistencia interna de un instrumento constituido por una escala Likert, o cualquier escala de opciones múltiples. Se afirma que varias modificaciones de las fórmulas de Kuder y Richardson se han propuesto en el transcurso del tiempo, pero la que ha recibido mayor aceptación hasta la fecha es el estadígrafo α Cronbach (Cohen, y Swerdlik, 2001).

Para determinar el coeficiente α Cronbach el investigador calcula la correlación de cada reactivo o ítem con cada uno de los otros, resultando una gran cantidad de coeficientes de correlación. El valor de α es el promedio de todos los coeficientes de correlación (Cozby, 2005). Visto desde otra perspectiva, el coeficiente α Cronbach puede considerarse como la media de todas las correlaciones de división por mitades posibles, otro método de cálculo de consistencia interna, las buenas junto las malas (Cohen y Swerdlik, 2001).

4. Coeficiente α de Cronbach y análisis de Ítems

Algunos paquetes estadísticos permiten hacer análisis de consistencia interna de los ítems o reactivos de un instrumento (Levy y Varela, 2003). Estos análisis tratan de determinar el grado en el que están relacionados recíprocamente los reactivos o ítems (Brown, 1980), y pueden hacerse a partir de una sola aplicación de una forma única de una prueba o instrumento (Cohen y Swerdlik, 2001).

La descripción resultante de los ítems o reactivos es resumida con el término homogeneidad, la cual es el grado en que una prueba o instrumento mide un solo factor, es decir, el grado en que los reactivos en una escala son unifactoriales. Más precisamente, el coeficiente de consistencia interna dependerá directamente de las correlaciones entre los ítems o reactivos, esto es, del grado en que los ítems midan la misma variable. Mientras más homogéneos sean los ítems, mayor será el valor de la consistencia interna para un número dado de ítems (Magnusson, 1978).

Relacionando estas interpretaciones y volviendo al asunto de realizar un análisis de consistencia interna de los ítems mediante un software, los autores exponen un proceso por etapas en las que paso a paso, se deben eliminar del conjunto de ítems aquellos que hacen disminuir la consistencia interna global; es decir, aquellos menos consistentes. El proceso se detiene cuando la escala no mejora o mantiene su nivel de consistencia interna, con la eliminación de algún otro ítem (Levy y Varela, 2003). Estos autores, coincidiendo con Cozby (2005), sugieren utilizar un segundo criterio para la depuración de ítems: correlación del ítem con el total calculado como suma de todos los ítems menos el calculado, denominada correlaciones reactivo-total.

5. Previsiones prácticas

Mientras que el conocimiento del grado de validez de un instrumento permite usar significativamente los datos obtenidos con él, el conocimiento de su confiabilidad permite usar los datos correctamente (Magnusson, 1978). Desde

esta perspectiva práctica, muchas previsiones deben tomarse al momento de estimar e interpretar un coeficiente de consistencia interna. En este espacio nos ocuparemos de sólo dos de ellas.

Primeramente, es oportuno destacar, y ya algo se ha dejado ver al respecto, que de manera “natural” el coeficiente de consistencia interna α de Cronbach expresa el grado en que los ítems miden la misma variable: homogeneidad. De tal manera que su utilidad original se orienta a calcular la confiabilidad de un instrumento cuyos ítems o reactivos conformen un único dominio, esto es, una variable o rasgo único.

De esta manera, si su instrumento consta de o mide varias variables, es decir, tiene implícitamente subinstrumentos o subpruebas individuales, la recomendación es disponer de medidas de confiabilidad para cada subprueba (Brown, 1980). Más específicamente, los resultados del estudio deben reportar tantos α como subpruebas y variables o rasgos se tengan. El usuario de los resultados compuestos o combinados de un instrumento heterogéneo, debe ejecutar siempre verificaciones para ver si se dan estimaciones de confiabilidad para cada subprueba. Estos y otros detalles pueden consultarse en Brown (1980) y Nunnally y Bernstein (1995).

En segundo lugar, se requiere comentar respecto a cómo interpretar un valor específico de α , esto es, cuál es el nivel satisfactorio de fiabilidad. Los investigadores Carmines y Zeller (1979) consideran, que como regla general, las confiabilidades no deben ser inferiores a 0.80. Otros valores referenciales pueden encontrarse en Kerlinger y Lee (2002). Sin embargo, estos últimos autores mencionan que Nunnally (1978) afirma que un nivel satisfactorio de confiabilidad depende de cómo se utilice la medida.

Para finalizar, y como conclusión general de esta Nota, citemos otra vez a Carmines y Zeller: “... the most important thing to remember is to report the reliability of the scale and how it was calculated” (p. 51).

Referencias Bibliográficas

- Brown, Frederick G. (1980). **Principios de la Medición en Psicología y Educación**. Editorial El Manual Moderno. México.
- Carmines, Edward and Zeller, Richard. (1979). **Reliability and Validity Assessment**. SAGE Publications. USA.
- Cohen, Ronald y Swerdlick, Mark. (2001). **Pruebas y Evaluación Psicológicas. Introducción a las Pruebas y a la Medición**. McGraw Hill. México.
- Cozby, Paul C. (2005) **Métodos de Investigación del Comportamiento**. McGraw Hill. México.
- Kerlinger, Fred y Lee, Howard. (2002). **Investigación del Comportamiento. Métodos de Investigación en Ciencias Sociales**. McGraw Hill. México.

- Lèvy M, Jean Pierre y Varela M., Jesús. (2005). **Análisis Multivariable para las Ciencias Sociales**. Pearson. Prentice Hall. España.
- Magnusson, David. (1978). **Teoría de los Test**. Editorial Trillas. México.
- Nunnally, Jum C. Jr. (1970). **Introduction to Psychological Measurement**. McGraw Hill. Japan.
- Nunnally, Jum C. y Bernstein, Ira H. (1995). **Teoría Psicométrica**. McGraw Hill. México.