

Confiabilidad en el punto de corte: re-análisis de Alvarado y colaboradores

Reliability for cut-off scores: re-analysis of Alvarado et al

Sr. Editor:

Deseamos difundir un método de estimación de la confiabilidad, para instrumentos cuyo uso se extienda a realizar conclusiones categóricas (por ejemplo, *burnout vs no burnout*, riesgo vs no riesgo), que es un aspecto característico de las pruebas de tamizaje; por ejemplo, el *Alcohol Use Identification Test*¹. El cálculo de la confiabilidad por consistencia interna mediante el coeficiente alfa² es apropiado cuando se satisfacen varias presunciones estadísticas, y el puntaje se interpreta de manera normativa (test referido a normas) y dimensional, es decir, como una expresión de un atributo continuo. Pero cuando el resultado se expresa como una clasificación, la interpretación cambia hacia un test referido a un criterio³. Este tipo de test es muy conocido en el ámbito educativo, cuando se toman decisiones sobre si los estudiantes aprueban o no aprueban los contenidos enseñados⁴. Es razonable que esta propuesta no sea conocida en ciencias de la salud debido al marco clínico de sus instrumentos.

Por ejemplo, en un reciente estudio de validación, Alvarado y cols.⁵ crearon puntos de corte para discriminar a los sujetos con algún nivel de dependencia alcohólica; aunque tal instrumento no medía conocimientos o habilidades (como se hace en el área educativa) sino una condición clínica (presenta o no presenta), el uso categórico es el mismo que en la situación educativa (aprueba o no aprueba).

Por lo tanto, en este tipo de tests con puntos de cortes, la estimación de la confiabilidad más apropiada apunta a la clasificación realizada, lo que conduce a que se estime la confiabilidad alrededor del punto de corte elegido para separar a los sujetos^{3,6}. Existen varios modelos para representar esta confiabilidad, como los de consistencia de las desviaciones, de los dominios muestreados y de las decisiones categóricas⁶. El método que se informa ahora corresponde al modelo basado en la desviación del punto de corte, creado por Livingston⁷, y desarrollado como una extensión de la teoría clásica de los test para pruebas referidas a criterios. La ecuación es la siguiente:

$$K2 = \frac{r_{xx}DE^2 + (\bar{X} - C)^2}{DE^2 + (\bar{X} - C)^2}$$

En esta fórmula, r_{xx} es la confiabilidad para pruebas normativas (por ejemplo, el coeficiente alfa), DE^2 es la varianza del puntaje, \bar{X} es el puntaje promedio, y C el puntaje identificado como punto de corte. Se trata de una razón entre el desvío del puntaje verdadero respecto al punto de corte, sobre el desvío del puntaje observado sobre el mismo criterio³. Tomando los resultados de Alvarado y colaboradores, aportamos el cálculo del coeficiente de Livingston ($K2$) a los puntos de corte obtenidos por los autores para el AUDIT (Tabla 1).

Tabla 1. Coeficiente de confiabilidad criterial correspondiente a los puntajes del AUDIT, con datos de Alvarado et al. (2009)⁵

Puntaje AUDIT	K2
3	0,752
4	0,748
5	0,770
6	0,807
7	0,846
8	0,878
9	0,904
10	0,924
11	0,938
12	0,949
13	0,958
14	0,965
15	0,970

K2: Coeficiente de confiabilidad criterial de Livingston.

Desde la Tabla 1 se puede reconocer que los puntajes más extremos aseguran más confiabilidad alrededor del punto de corte. Esto es razonable pues los casos más severos deben ser identificados con más precisión que los casos cercanos a la media. Este resultado debe añadirse a los criterios que los autores usaron para determinar el punto de corte apropiado en su muestra.

Finalmente, el presente aporte lleva a dos consideraciones adicionales que creemos importantes:

1) Independientemente del modelo elegido para estimar la confiabilidad de la clasificación realizada en los puntos de corte, parece necesario que los estudios de validación no solo exijan una evidencia de la sensibilidad y especificidad diagnóstica efectivas de un instrumento⁸, sino también de la confiabilidad en este tipo de decisión; 2) Considerando que el error de medición es ubicuo a un proceso de medición y es un componente del puntaje obtenido, la interpretación del puntaje de un sujeto debe incluir alguna información del error de medición, y el método presentado puede tener un papel relevante ahí.

César Merino-Soto¹, Marisol Angulo-Ramos²
¹Magíster en Psicología Educativa. Instituto de Investigación de Psicología, Universidad de San Martín de Porres. Lima, Perú.
²Licenciada en Enfermería. Instituto Superior Tecnológico El Libertador. Lima, Perú.

Referencias

1. Saunders JB, Aasland OG, Babor TF, De La Fuente JR, Grant M. Development Of The Alcohol Use Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction* 1993; 88: 791-804.
2. Cronbach, LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 6: 297-334.
3. Gempp R, Saiz JL. El coeficiente K2 de Livingston y la habilidad de una decisión dicotómica en un test psicológico. *Univ Psychol* 2014; 13: 217-26.
4. Woolfolk A. *Psicología Educativa* (11 edición). México, Pearson Educación; 2010.
5. Alvarado M, Garmendia M, Acuña G, Santis R, Arteaga O. Validez y confiabilidad de la versión chilena del Alcohol Use Disorders Identification Test (AUDIT). *Rev Med Chile* 2009; 137: 1463-8.
6. Mateo J. La fiabilidad y la validez desde la perspectiva criterial. En: Abalde E., Muñoz J, Editores, *Metodología educativa I. Jornadas de Metodología de Investigación Educativa (A Coruña, 23-24 abril 1991)*. Coruña: Universidade da Coruña; 1992. p. 77-86.
7. Livingston SA. A criterion-referenced application of classical test theory. *J Educ Meas* 1972; 19: 13-26.
8. Valenzuela L, Cifuentes L. Validez de estudios de tests diagnósticos. *Rev Med Chile* 2008; 136: 401-4.

Correspondencia a:
 César Merino-Soto
 Universidad de San Martín de Porres.
 Calle Enrique Palacios 430, Chorrillos, Lima 9 – Perú.