

Para Referencia:

Tristán L.A. (1999-2002) *Contribución al estudio del error de medida*. Guía del usuario de Kalt-Criterial; notas sobre evaluación criterial 13. Instituto de Evaluación e Ingeniería Avanzada, S.C. México. 5 pp.

KALT-CRITERIAL SISTEMA DE CALIFICACIÓN DE PRUEBAS REFERIDAS A CRITERIO



GUIA DEL USUARIO

Versión 2. (1999-2002)

Familia de Programas KALT
Cordillera Occidental 635. Col. Lomas 4ª sección
C.P. 78216 San Luis Potosí, México
Tel.(444) 825-50-78
Página web: www.ieia.com.mx
E-MAIL informes@ieia.com.mx

NOTAS SOBRE EVALUACION CRITERIAL 13

CONTRIBUCIÓN AL ESTUDIO DEL ERROR DE MEDIDA

Presentación

El cálculo de la confiabilidad utilizando el coeficiente ALFA de Cronbach (o las aproximaciones de KUDER-RICHARDSON), conducen a una estimación de la consistencia interna de una prueba:

$$ALFA = \frac{N}{N-1} \left(1 - \frac{\sum pq}{VarX} \right) \quad (1)$$

Donde

N = número de reactivos

P = proporción de respuestas correctas del reactivo

Q = 1-p

VarX = varianza de puntajes de los sustentantes (en número de aciertos)

Se dice que los valores de ALFA fluctúan entre -1 y +1. La realidad es que van de $-\infty$ a +1.

No se tienen valores de referencia para la aceptación de ALFA. Se acostumbra decir que es deseable obtener valores cercanos a +1 (alta confiabilidad).

Problema

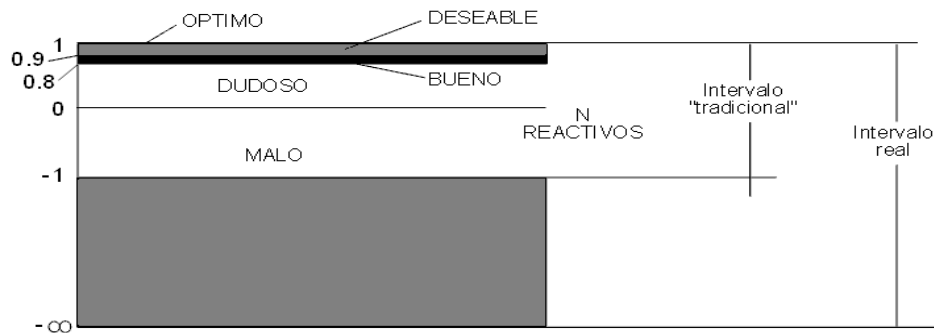
El problema que se plantea es: ¿cuál es el valor esperado de ALFA para una prueba de N reactivos?

Tal como se ha resuelto en la literatura, los valores esperados de ALFA, deben ser lo más cercano al máximo (ALFA=1). Sin embargo, el máximo de ALFA se obtiene en cualquiera de las siguientes posibilidades:

- A) El número de reactivos tiende a infinito. Esto es claro porque en lugar de hacer un muestreo se explora todo el dominio infinito de manera exhaustiva. Esta posibilidad no es práctica, pero se emplea de manera indirecta con el uso de la fórmula de Spearman-Brown que ofrece el número de reactivos necesarios para un valor de confiabilidad dado. Generalmente esto se traduce en que si se desea incrementar el valor de ALFA "basta" con incrementar el número de reactivos (con las mismas características de la prueba base).
- B) Todos los reactivos son de la misma dificultad y discriminan positivamente del mismo modo. Esto atenta contra la validez de medición porque no existe escala de reactivos (de fácil a difícil), ya que se tiene la prueba en una sola posición dentro de la escala. A pesar de esta deficiencia en validez, ALFA rinde un valor de 1. Esto se denomina "paradoja de la atenuación".

Este documento presenta los resultados de un análisis distinto. El planteamiento parte del hecho de que si se sigue el esquema tradicional entonces no es factible acotar el valor de ALFA, porque este coeficiente depende de muy diversas variables: el número de reactivos, las dificultades de los mismos y la forma en que los sujetos responden al cuestionario. El único parámetro que puede conocerse a priori es el número de reactivos, los demás elementos, siendo variables, no permiten identificar patrones de funcionamiento de la confiabilidad. En este enfoque clásico se busca obtener la máxima confiabilidad, independientemente de la validez de medida; como no hay ninguna restricción, bajo esta óptica se puede llegar a valores de ALFA=1 en pruebas que no tienen ninguna validez o que estén formadas por infinitos reactivos.

Lo que se acostumbra es decir que valores de ALFA=0.9 o mayores son “Deseables”, que valores superiores a 0.8 son “buenos” y debajo de 0.8 la confiabilidad de la prueba es dudosa, por lo que cada evaluador debe decidir si acepta el valor de ALFA en función de los propósitos de su prueba.



En este enfoque, el valor de ALFA queda muy vagamente definido, no se relaciona con la validez de medida de la prueba y queda a criterio muy abierto para el evaluador.

Por estas razones el esquema tradicional no funciona. Es necesario acotar ALFA utilizando alguna condición conveniente de acuerdo con un diseño o siguiendo alguna especificación o criterio externo. Esto cambia el planteamiento de esta forma: Dada una prueba de tamaño infinito, de N reactivos, que tenga validez de medida, ¿cuál es el máximo de ALFA? La condición que se está imponiendo es “validez de medida” para N reactivos.

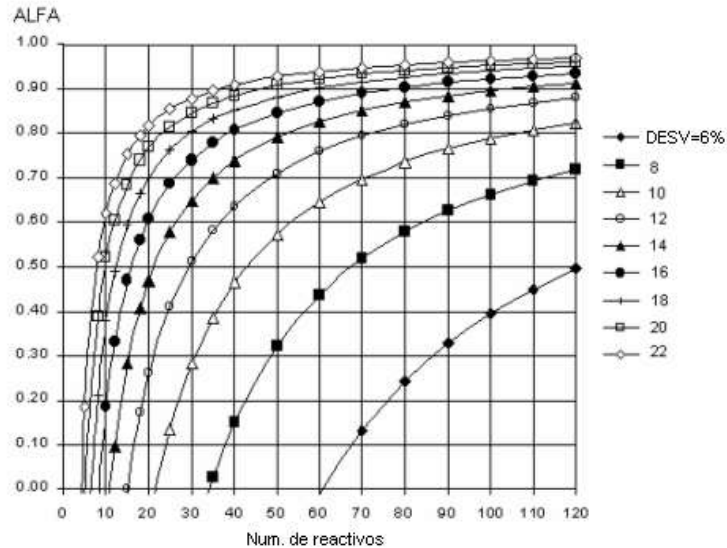
¿Cómo se establece esta restricción? Si se acepta una prueba con validez de medida de acuerdo con el diseño 20-80, esto permite definir la distribución uniforme de N reactivos, con ello se puede calcular el numerador Σpq , y por lo tanto, queda acotado el valor de ALFA para una N dada y para un valor de la varianza de las respuestas de los sujetos.

El requisito de “validez” de medida” es una condición necesaria y suficiente para el cálculo de ALFA, de tal modo que se puede predecir el valor “óptimo” o “deseable” para consistencia interna de una prueba que cumple el diseño con la recta 20-80. Sin el uso de esta condición el cálculo de valores óptimos de ALFA carece de sentido.

A continuación se presentan los resultados obtenidos con ALFA partiendo de esta condición necesaria y suficiente.

Resultados

1. CURVAS DE VALORES OPTIMOS DE ALFA DE CRONBACH PARA PRUEBAS QUE CUMPLEN LA VALIDEZ DE DISEÑO EN RECTA 20-80 (dependiendo de la varianza de respuestas de los sujetos).



Ejemplo:

Se tiene un examen de 20 reactivos. ¿Cuál es el valor esperado para ALFA si la desviación estándar de las calificaciones de los sujetos es: (a) 10%?, (b) 16%?

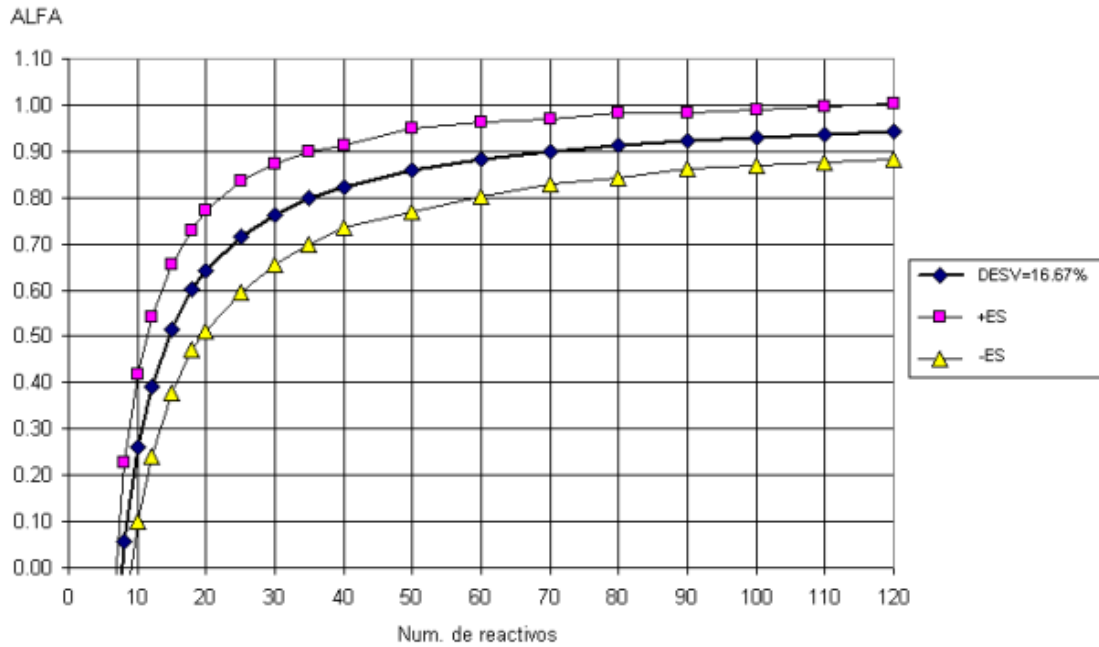
Solución:

- Para $desv=10\%$ la curva NO proporciona valores positivos para ALFA. Esto indica que No se debe usar esta prueba porque la confiabilidad de una prueba cuya validez de medida cumple las condiciones de diseño 20-80, sería negativa.
- Para $desv=16\%$ el valor de $ALFA=0.61$

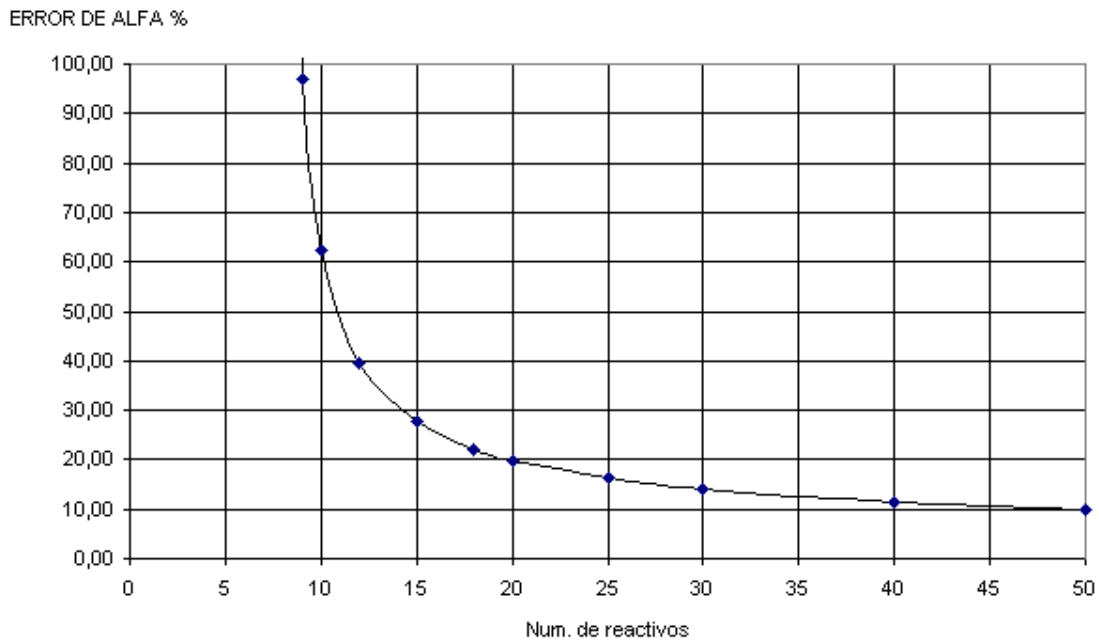
IMPORTANTE: El valor de confiabilidad de una prueba de 20 reactivos que cumple el diseño 20-80, con desviación estándar de 16% en las puntuaciones de los alumnos, debe estar cercana al valor 0.61. Por lo tanto, si el evaluador consigue modificar la distribución de los reactivos de la prueba (sin cambiar su número) hasta llegar a un valor de $ALFA=0.9$ o mayor, en este caso, se sabría que la prueba está "sobrediseñada" hacia la confiabilidad y, por lo tanto, ha perdido fuerza en el único posible parámetro cuantitativo que existe para la validez: el ajuste al diseño de la recta 20-80.

MORALEJA: No se debe buscar un valor de ALFA cercano a 1, en cambio, se debe buscar acercarse a la curva de valores óptimos de ALFA, dados el número de reactivos y la desviación estándar de las puntuaciones de los sujetos.

2. INTERVALO DE ACEPTACION PARA LOS VALORES IDEALES DE CONFIABILIDAD DE PRUEBAS QUE CUMPLEN EL DISEÑO 20-80, CON DESVIACION ESTANDAR 16.67% EN LAS PUNTUACIONES DE LOS SUJETOS



3. ERROR EN PORCENTAJE DEL VALOR DE ALFA PARA PRUEBAS CON DESVIACION ESTANDAR DE 16.67%



Esta curva se relaciona con el error de medida y representa una cota superior que puede compararse contra los valores dados por una prueba de intervalo de confianza utilizando t de Student (ver Nota sobre Evaluación Criterial N.5).

4. TABLA DE VALORES DE CONFIABILIDAD PARA PRUEBAS QUE CUMPLEN EL DISEÑO 20-80, CON DESVIACION ESTANDAR DE 16.67% EN LOS SUJETOS.

No. de reactivos	ALFA	ERROR	ERROR %	No. de reactivos	ALFA	ERROR	ERROR %
2	-3.76		-	110	0.94	0.06	6.23
3	-1.92		-	120	0.94	0.06	5.94
4	-1.07		-	130	0.95	0.05	5.68
5	-0.59		-	140	0.95	0.05	5.46
6	-0.30		-	150	0.95	0.05	5.26
7	-0.09		-	160	0.96	0.05	5.08
8	0.06	0.17	297.50	170	0.96	0.05	4.92
9	0.17	0.16	97.14	180	0.96	0.05	4.77
10	0.26	0.16	62.39	190	0.96	0.04	4.63
12	0.39	0.15	39.47	200	0.97	0.04	4.51
15	0.52	0.14	27.60	210	0.97	0.04	4.40
18	0.60	0.13	22.19	220	0.97	0.04	4.29
20	0.64	0.13	19.91	230	0.97	0.04	4.19
25	0.72	0.12	16.27	240	0.97	0.04	4.10
30	0.76	0.11	14.08	250	0.97	0.04	4.01
40	0.82	0.09	11.46	260	0.97	0.04	3.93
50	0.86	0.09	9.90	270	0.97	0.04	3.85
60	0.88	0.08	8.84	280	0.98	0.04	3.78
70	0.90	0.07	8.06	290	0.98	0.04	3.71
80	0.91	0.07	7.46	300	0.98	0.04	3.65
90	0.92	0.06	6.97	1000	0.99	0.02	1.97
100	0.93	0.06	6.57	5000	1.00	0.01	0.88