

Tema 5

La fiabilidad en los tests referidos al criterio

Profesor-tutor: Jaime León

Abril, 2018

TEMA 5

- 1 INTRODUCCIÓN
- 2 TRN vs TRC
- 3 MODELO DE MILLMAN
- 4 FIABILIDAD DE LAS CLASIFICACIONES
- 5 MÉTODOS PARA ESTIMAR EL PUNTO DE CORTE

1. Introducción

Libro antiguo Pág. 249 – Libro nuevo Pág. 232

Un test referido al criterio (TRC) se utiliza para evaluar el status absoluto del sujeto con respecto a algún dominio bien definido.

3. Un test referido al criterio se utiliza para: a) evaluar el nivel absoluto del sujeto con respecto a algún dominio bien definido, b) poner de manifiesto las diferencias individuales en el rasgo que mide el test, c) obtener la puntuación verdadera del sujeto en el rasgo que está midiendo el test.

A

2.- Los tests referidos al criterio (TRC) permiten evaluar el nivel de rendimiento de los sujetos: a) en relación al obtenido por una muestra representativa de la población; b) en función de la aptitud media del grupo; c) en relación a unos estándares o dominio bien definidos

C



TEMA 5

- 1 INTRODUCCIÓN
- 2 TRN vs TRC
- 3 MODELO DE MILLMAN
- 4 FIABILIDAD DE LAS CLASIFICACIONES
- 5 MÉTODOS PARA ESTIMAR EL PUNTO DE CORTE

2. TRC vs TRN

Libro antiguo Pág. 249 – Libro nuevo Pág. 232

	Test Referido al Criterio (TRC)	Test Referido a las Normas (TRN)
Objetivo	Interpretación absoluta respecto a un dominio de conductas Puntuación → Test referido al dominio Apto-Rechazado → Test de maestría	Posición relativa respecto a un grupo normativo Describir rasgo
Ítems	Relevancia y representatividad respecto al objetivo del test	Que discriminen para maximizar las diferencias individuales

Los tests referidos al criterio: a) combinan las puntuaciones del test y del criterio; b) sólo tienen validez predictiva o relativa al criterio; c) no requieren la utilización de un grupo normativo.

C

9.- En los tests referidos al criterio: a) los ítems deben tener un alto poder discriminativo, b) es fundamental la representatividad de los ítems respecto a un dominio, c) se intenta maximizar las diferencias individuales

B

1.- En los tests referidos a las normas las puntuaciones de los sujetos se interpretan en relación: a) a un grupo de referencia; b) a unos criterios mínimos; c) al dominio de contenidos.

A

3.- En los tests referidos a normas (TRN) el objetivo es: a) medir el rendimiento de los sujetos en relación a unos criterios específicos, b) poner de manifiesto las diferencias individuales en aquello que mide el test, c) especificar de forma clara las conductas que se desean evaluar.

B

TEMA 5

- 1 INTRODUCCIÓN
- 2 TRN vs TRC
- 3 MODELO DE MILLMAN
- 4 FIABILIDAD DE LAS CLASIFICACIONES
- 5 MÉTODOS PARA ESTIMAR EL PUNTO DE CORTE

3. Modelo de Millman

Libro antiguo Pág. 252 – Libro nuevo Pág. 235

Objetivo → determinar la longitud del test (número de ítems)

Supuestos

Compuesto por ítems dicotómicos

Probabilidad de un sujeto de responder bien es constante en todos los ítems

Las respuestas son independientes

Los errores se ajustan al modelo binomial

En el modelo de Millman se parte del supuesto de que: a) los errores se ajustan a la curva normal; b) los ítems del test son dicotómicos; c) cada ítem del test tiene una probabilidad distinta de ser respondido correctamente.

B

El modelo de Millman está basado en: a) modelos bayesianos; b) el modelo de regresión; c) el modelo binomial.

C

Fórmula para calcular la longitud del test

$$n = \frac{p_c(1 - p_c)}{e^2}$$

n = número de ítems del test

p_c = proporción de aciertos para ser considerado apto

e^2 = error máximo admisible

3. Modelo de Millman

En una prueba de conocimientos se considera que para que los sujetos sean considerados como APTOS deberán acertar el 75% de sus ítems. Cuál deberá ser la longitud de la prueba si el error máximo que estamos dispuestos a admitir es de $e^2 = 0.0025$? a) 65; b) 75; c) 85

$$p_c = 0.75 \quad e^2 = 0.0025$$

$$n = \frac{p_c(1 - p_c)}{e^2}$$

$$n = \frac{0.75(1 - 0.75)}{0.0025}$$

$$n = 75$$

TEMA 5

- 1 INTRODUCCIÓN
- 2 TRN vs TRC
- 3 MODELO DE MILLMAN
- 4 FIABILIDAD DE LAS CLASIFICACIONES
- 5 MÉTODOS PARA ESTIMAR EL PUNTO DE CORTE

4. Fiabilidad en las clasificaciones

4.1 Dos aplicaciones del test

4.2 Una aplicación del test

4. Fiabilidad en las clasificaciones

4.1 Dos aplicaciones del test

Libro antiguo Pág. 253 – Libro nuevo Pág. 236

Un TRC es fiable si tras dos aplicaciones están asignados a la misma categoría (apto – rechazado, trastorno - sano)

Estos tres métodos que miden la consistencia en las clasificaciones

Hambleton y Novick

Kappa de Cohen

Crocker y Algina

4. Fiabilidad en las clasificaciones

4.1 Dos aplicaciones del test

Hambleton y Novick

15.- Para hallar la consistencia en las clasificaciones de dos tests paralelos se aplican a una muestra de 10 sujetos obteniéndose los resultados que se muestran a continuación. Si para ser clasificado como APTO deben obtener al menos 7 puntos, ¿qué consistencia en las clasificaciones se obtendría al utilizar los tests?, ¿cuál sería la consistencia obtenida por mero azar?: a) 0,80 y 0,50; b) 0,74 y 0,25; c) 0,40 y 0,25

Test A: 10, 9, 7, 8, 6, 5, 9, 3, 6, 5

Test B: 9, 8, 7, 6, 6, 5, 10, 4, 7, 6

		Test B		
		Apto (1)	Rechazado (2)	
Test A	Apto (1)	$N_{11} = 4$	$N_{12} = 1$	$N_{A1} = 5$
	Rechazado (2)	$N_{21} = 1$	$N_{22} = 4$	$N_{A2} = 5$
		$N_{B1} = 5$	$N_{B2} = 5$	$N = 10$

Pág. 27 del formulario

$$P_c = \frac{N_{11} + N_{22}}{N}$$

$$P_c = \frac{4 + 4}{10}$$

$$P_c = 0.80$$

$$P_a = \frac{N_{B1} \cdot N_{A1} + N_{B2} \cdot N_{A2}}{N^2}$$

$$P_a = \frac{5 \cdot 5 + 5 \cdot 5}{100}$$

$$P_a = 0.50$$

4. Fiabilidad en las clasificaciones

4.1 Dos aplicaciones del test

Kappa de Cohen

18. Dos psicólogos forenses han evaluado a 100 presos con el objetivo de concederles la libertad condicional. En general, es más probable que los psicópatas cometan delitos violentos que los sujetos maniaco-depresivos o normales. Por tanto cada psicólogo ha puntuado a cada uno de los 100 presos en una escala de 1 a 3, donde 1 es psicópata; 2 es maniaco-depresivo; y 3 es normal. En la siguiente tabla se muestran los resultados:

		Psicólogo forense A		
		1	2	3
Psicólogo forense B	1	44	5	1
	2	7	20	3
	3	9	5	6

Pág. 28 del formulario

a) La probabilidad de clasificaciones consistentes esperadas por azar es 0.7, b) El índice *Kappa de Cohen* es igual a 0.41, c) El coeficiente *Kappa* es significativamente distinto de cero (NC 95%).

		Psicólogo A			
		1	2	3	
Psicólogo B	1	$N_{11} = 44$	$N_{12} = 5$	$N_{13} = 1$	$N_{B1} = 50$
	2	$N_{21} = 7$	$N_{22} = 20$	$N_{23} = 3$	$N_{B2} = 30$
	3	$N_{31} = 9$	$N_{32} = 5$	$N_{33} = 6$	$N_{B3} = 20$
		$N_{A1} = 60$	$N_{A2} = 30$	$N_{A3} = 10$	$N = 100$

4. Fiabilidad en las clasificaciones

4.1 Dos aplicaciones del test

		Psicólogo A			
		1	2	3	
Psicólogo B	1	$N_{11} = 44$	$N_{12} = 5$	$N_{13} = 1$	$N_{B1} = 50$
	2	$N_{21} = 7$	$N_{22} = 20$	$N_{23} = 3$	$N_{B2} = 30$
	3	$N_{31} = 9$	$N_{32} = 5$	$N_{33} = 6$	$N_{B3} = 20$
		$N_{A1} = 60$	$N_{A2} = 30$	$N_{A3} = 10$	$N = 100$

$$K = \frac{P_c - P_a}{1 - P_a}$$

$$P_c = \frac{N_{11} + N_{22} + N_{33}}{N}$$

$$P_c = \frac{44 + 20 + 6}{100}$$

$$P_c = 0.70$$

$$P_a = \frac{N_{A1} \cdot N_{B1} + N_{A2} \cdot N_{B2} + N_{A3} \cdot N_{B3}}{N^2}$$

$$P_a = \frac{60 \cdot 50 + 30 \cdot 30 + 20 \cdot 10}{100^2}$$

$$P_a = 0.41$$

$$K = \frac{P_c - P_a}{1 - P_a}$$

$$K = \frac{0.70 - 0.41}{1 - 0.41}$$

$$K = 0.49$$

$$IC = K \pm E_{max}$$

$$E_{max} = Z_c S_e$$

$$Z_c = 1.96$$

$$S_e = \sqrt{\frac{F_a}{N(N - F_a)}}$$

$$F_a = \frac{N_{B1} \cdot N_{A1} + N_{B2} \cdot N_{A2} + N_{B3} \cdot N_{A3}}{N}$$

$$F_a = 41$$

$$S_e = 0.08$$

$$E_{max} = Z_c S_e$$

$$E_{max} = 1.96 \cdot 0.08$$

$$E_{max} = 0.16$$

$$IC = K \pm E_{max}$$

$$IC = 0.49 \pm 0.16$$

$$IC = 0.33, 0.65$$

4. Fiabilidad en las clasificaciones

4.1 Dos aplicaciones del test

Crocker y Algina

Aplicados dos tests paralelos que tratan de diagnosticar el Trastorno de Hiperactividad con Déficit de Atención (TDAH), se ha encontrado la siguiente clasificación:

		Test A		Total
		Diagnóstico TDAH	No-diagnóstico TDAH	
Test B	Diagnóstico TDAH	34	10	44
	No-diagnóstico TDAH	15	41	56

Pág. 28 del formulario

17. El índice de Crocker y Algina es igual a: a) 0,46; b) 0,50; c) 0,75.

$$p^* = 2P_C - 1$$

$$P_C = \frac{N_{11} + N_{22}}{N}$$

$$P_C = \frac{34 + 41}{100}$$

$$P_C = 0.75$$

$$p^* = 2 \cdot 0.75 - 1$$

$$p^* = 0.5$$

4. Fiabilidad en las clasificaciones

4.1 Dos aplicaciones del test

4.2 Una aplicación del test

4. Fiabilidad en las clasificaciones

4.2 Una aplicación del test

Libro antiguo Pág. 259 – Libro nuevo Pág. 242

Estos métodos se utilizan cuando solo tenemos una prueba

Huynh

Subkoviak

Livingston

Da más importancia los sujetos que están lejos del punto de corte.

Al igual que el α de Cronbach varía entre 0 y 1.

Cuando $\alpha = 1 \rightarrow K$ de Livingstone = 1

Cuando la media del test = pto.corte $\rightarrow K$ de Livingstone = α

Cuando la media del test = 1 $\rightarrow K$ de Livingstone = 1

Cuando $\alpha < 1 \rightarrow K$ de Livingstone $> \alpha$

4. Fiabilidad en las clasificaciones

4.2 Dos aplicaciones del test

Se quiere llevar a cabo un estudio de las propiedades psicométricas de un nuevo test de inglés (X). Para ello se ha aplicado a una muestra de 6 sujetos (es un ejemplo) junto con una prueba ya validada (Y). Los resultados obtenidos se encuentran en la tabla adjunta:

Sujetos	X1	X2	X3	X4	Y
1	1	1	0	1	5
2	1	1	1	1	6
3	1	0	0	0	4
4	1	1	0	1	6
5	0	1	1	1	7
6	1	1	0	0	8

Si la puntuación de 2 es el punto de corte en el test para clasificar a los sujetos en aptos y no aptos, el coeficiente de Livingston será: a) 0,49; b) 0,39; c) 0,59

$$K_{XV}^2 = \frac{\alpha S_x^2 + (\bar{X} - C)^2}{S_x^2 + (\bar{X} - C)^2}$$

4. Fiabilidad en las clasificaciones

4.2 Dos aplicaciones del test

$$K_{XV}^2 = \frac{\alpha S_x^2 + (\bar{X} - C)^2}{S_x^2 + (\bar{X} - C)^2}$$

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum pq}{S_x^2} \right)$$

$$\sum pq = p_1q_1 + p_2q_2 + p_3q_3 + p_4q_4$$

$$\sum pq = \frac{5}{6} \cdot \frac{1}{6} + \frac{5}{6} \cdot \frac{1}{6} + \frac{2}{6} \cdot \frac{4}{6} + \frac{4}{6} \cdot \frac{2}{6}$$

$$\sum pq = 0.72$$

$$S_x^2 = 0.88$$

$$\alpha = \frac{5}{5-1} \left(1 - \frac{0.72}{0.89} \right)$$

$$\alpha = 0.24$$

$$K_{XV}^2 = \frac{0.24 \cdot 0.88 + (2.66 - 2)^2}{0.88 + (2.66 - 2)^2}$$

$$K_{XV}^2 = 0.49$$

TEMA 5

- 1 INTRODUCCIÓN
- 2 TRN vs TRC
- 3 MODELO DE MILLMAN
- 4 FIABILIDAD DE LAS CLASIFICACIONES
- 5 MÉTODOS PARA ESTIMAR EL PUNTO DE CORTE

5. Métodos para estimar el punto de corte

Libro antiguo Pág. 266 – Libro nuevo Pág. 250

Métodos

Valorativos

Combinados

Compromiso

Métodos Valorativos

Se basan en los ítems y no en los sujetos

Nedelsky → preguntas tipo test

Ebel

8. En los métodos para estimar los puntos de corte relativos en los tests referidos al criterio, el punto de corte se establece en relación al: a) grupo de sujetos evaluados; b) constructo medido de interés; c) grupo de expertos seleccionado.

C

El método de Nedelsky se utiliza con tests compuestos con: a) ítems de elección múltiple; b) ítems de tipo verdadero-falso; c) todo tipo de ítems.

A