PSICOMETRÍA

Tema 4.
Evaluación del instrumento de medida:
ANÁLISIS DE ÍTEMS

Salvador Chacón Moscoso
Susana Sanduvete Chaves

Agradecemos a Francisco Pablo Holgado Tello su inestimable colaboración en la elaboración de este material

ÍNDICE

TEMA 4: ANÁLISIS DE ÍTEMS

1. INTRODUCCIÓN

La calidad psicométrica puede evaluarse mediante el análisis de:

1. La alternativa correcta.

2. Las incorrectas.

Con el análisis de la **alternativa correcta**, se obtienen indicadores como:

1. Dificultad: puede estimarse a priori desde consideraciones teóricas. Su análisis empírico permite detectar ítems con una facilidad o dificultad extrema → revisar o eliminar.
2. Discriminación: ¿es capaz de diferenciar a los participantes de distinto nivel en la variable medida?
3. Fiabilidad y validez: reflejan la contribución del ítem a la fiabilidad y validez referida al test en su conjunto.

3

TEMA 4: ANÁLISIS DE ÍTEMS

1. INTRODUCCIÓN

La calidad psicométrica puede evaluarse mediante el análisis de:

1. La alternativa correcta.

2. Las incorrectas.

Análisis de las **incorrectas o análisis de distractores**: informa sobre la utilidad de cada alternativa incorrecta del ítem y su contribución a la calidad del mismo. Su función es atraer las respuestas de los participantes menos competentes en la variable medida.

4

TEMA 4: ANÁLISIS DE ÍTEMS

2. Parámetros de los ítems

5

TEMA 4: ANÁLISIS DE ÍTEMS

2.1. PARAMETROS DE LOS ÍTEMS. Dificultad del ítem

Dificultad del ítem: El índice más sencillo de obtener viene dado por la proporción de participantes que responden correctamente al ítem.

$$p = \frac{A}{N}$$

Donde:

A = número de participantes que respondieron correctamente al ítem

N = número de participantes que respondieron al ítem

p oscila entre 0 y 1.

Donde:

0 → ningún participante ha acertado el ítem (extremadamente difícil).

1 → todos los participantes han acertado el ítem (extremadamente fácil).

6

2.1. PARAMETROS DE LOS ÍTEMS.
Dificultad del ítem

Un ítem dicotómico de rendimiento en lengua se aplica a 10 estudiantes. Las respuestas al ítem se muestran en la siguiente tabla:

Alumno	1º	2º	3º	4º	5º	6º	7º	8º	9º	10º
Respuesta	1	1	0	1	1	0	0	1	1	0

Se aplica a otra muestra de participantes muy competentes en lengua:

Alumno	1º	2º	3º	4º	5º	6º	7º	8º	9º	10º
Respuesta	1	1	0	1	1	1	1	1	1	1

TEMA 4: ANÁLISIS DE ÍTEMS

2.1. PARAMETROS DE LOS ÍTEMS.
Dificultad del ítem

Un ítem dicotómico de rendimiento en lengua se aplica a 10 alumnos. Las respuestas al ítem se muestran en la siguiente tabla

Alumno	1º	2º	3º	4º	5º	6º	7º	8º	9º	10º
Resp.	1	1	0	1	1	0	0	1	1	0

p sería la proporción de participantes que han acertado (6/10)→ dificultad media

$$p = \frac{A}{N} = \frac{1+1+0+1+1+0+0+1+1+0}{10} = \frac{6}{10} = 0{,}6$$

Se aplica a otra muestra de participantes muy competentes en lengua:

Alumno	1º	2º	3º	4º	5º	6º	7º	8º	9º	10º
Resp.	1	1	0	1	1	1	1	1	1	1

En participantes más competentes *p*, lógicamente es mayor (9/10)→ dificultad baja

$$p = \frac{A}{N} = \frac{1+1+0+1+1+1+1+1+1+1}{10} = \frac{9}{10} = 0{,}9$$

Es decir, la dificultad del ítem va a depender de la muestra de participantes en la que se calcule ya que, si la muestra es competente, el ítem va a resultar fácil y viceversa

TEMA 4: ANÁLISIS DE ÍTEMS

2.1. PARAMETROS DE LOS ÍTEMS. Dificultad del ítem. Corrección de la dificultad en ítems de elección múltiple

Corrección de la dificultad: En ítems de elección múltiple los participantes con nula o poca competencia suelen responder al azar → el número de aciertos observados será mayor que el esperado según el nivel de competencia de los participantes. La dificultad se corrige según:

$$p_c = p - p_{azar} = \frac{A}{N} - \frac{\frac{E}{k-1}}{N} = \boxed{\frac{A - \frac{E}{k-1}}{N}}$$

Donde:
p_c = dificultad del ítem corregida
p = dificultad del ítem sin corregir
p_{azar} = proporción de respuestas correctas al azar
E = número de errores
k = número de alternativas de respuesta

Los supuestos para poder aplicarla son:

1. Existe un subgrupo de participantes, relativamente numeroso, que responden correctamente al ítem por pura **adivinación**.
2. Para este grupo de participantes, **todas** las alternativas tienen la misma probabilidad de ser elegidas.

Si no se sospecha sobre la existencia de participantes que responden al azar, conviene aplicar *p*

TEMA 4: ANÁLISIS DE ÍTEMS

9

2.1. PARAMETROS DE LOS ÍTEMS. Dificultad del ítem. Corrección de la dificultad en ítems de elección múltiple

En la tabla siguiente se muestran las distribuciones de frecuencias de las respuestas de 500 alumnos a cuatro ítems de elección múltiple con 4 alternativas de respuesta (* representa la opción correcta).

	Alternativas			
Ítem	A	B	C	D
1	5	200	105	**190***
2	64	**250***	80	106
3	5	**492***	0	3
4	50	65	**350***	35

- Calcular la dificultad de cada ítem (p y p_c).
- Ordenarlos de menor a mayor en función de p.

TEMA 4: ANÁLISIS DE ÍTEMS

10

2.1. PARAMETROS DE LOS ÍTEMS. Dificultad del ítem. Corrección de la dificultad en ítems de elección múltiple

ítem 1:

$$p = \frac{A}{N} = \frac{190}{500} = 0{,}38$$

$$p_c = p - p_{azar} = \frac{A - \frac{E}{k-1}}{N} = \frac{190 - \frac{310}{4-1}}{500} = 0{,}17$$

Ítem	Aciertos	Errores	p	p_c
3	492	8	0,98	0,97
4	350	150	0,70	0,60
2	250	250	0,50	0,33
1	190	310	0,38	0,17

La ordenación de los ítems de menor a mayor dificultad sería **3, 4, 2, y 1.**

La diferencia entre *p* y *p corregida* es mayor en los ítems más difíciles, ya que se supone que ha habido mayor adivinación al azar que en los fáciles.

TEMA 4: ANÁLISIS DE ÍTEMS

11

2.2. PARAMETROS DE LOS ÍTEMS. Discriminación del ítem

Discriminación del ítem: la capacidad del ítem para distinguir a los participantes de baja, media y alta puntuación en la variable medida. El criterio, o la variable medida puede ser interno o externo:

-*Criterio interno*: si ítem y test miden el mismo concepto, es esperable que los participantes que hayan obtenido una puntuación alta en el test, respondan correctamente al ítem, y viceversa.

-*Criterio externo*: el índice de discriminación es una medida del grado de validez del ítem referida a ese criterio, y su valor será proporcional al índice de validez del test.

Procedimientos estadísticos:

-Índice de discriminación D.

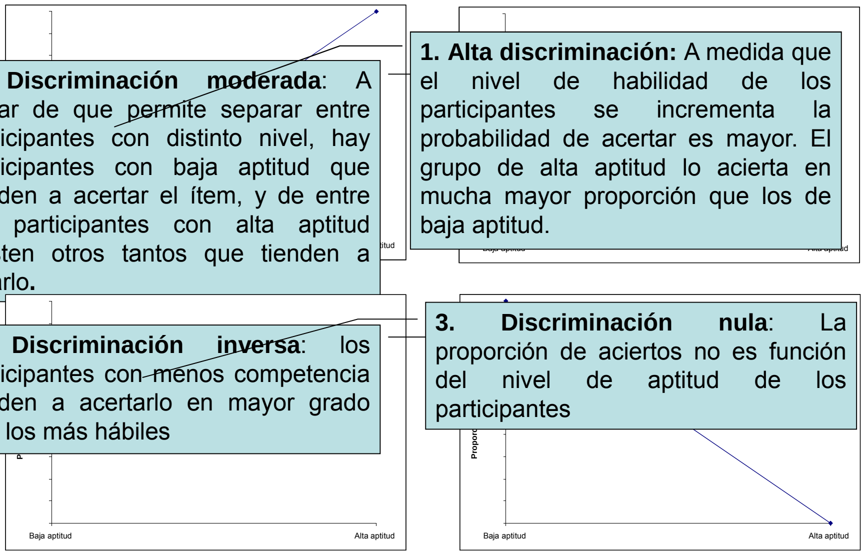
-Coeficientes de discriminación.

TEMA 4: ANÁLISIS DE ÍTEMS

12

2.2. PARAMETROS DE LOS ÍTEMS. Discriminación del ítem

Discriminación: proporción de aciertos en función del nivel de aptitud de los participantes



1. Alta discriminación: A medida que el nivel de habilidad de los participantes se incrementa la probabilidad de acertar es mayor. El grupo de alta aptitud lo acierta en mucha mayor proporción que los de baja aptitud.

2. Discriminación moderada: A pesar de que permite separar entre participantes con distinto nivel, hay participantes con baja aptitud que tienden a acertar el ítem, y de entre los participantes con alta aptitud existen otros tantos que tienden a fallarlo.

3. Discriminación nula: La proporción de aciertos no es función del nivel de aptitud de los participantes

4. Discriminación inversa: los participantes con menos competencia tienden a acertarlo en mayor grado que los más hábiles

2.2. PARAMETROS DE LOS ÍTEMS. Índice de discriminación D

El índice de Discriminación D: se basa en la comparación del rendimiento de los grupos extremos (bajo y alto) en las puntuaciones en el test. Se compara el número de participantes que ha acertado el ítem en el grupo alto con respecto al grupo de baja aptitud.

$$\boxed{D = \frac{A_{alto} - A_{bajo}}{N_g}}$$

Donde:

$A_{alto}\, A_{bajo}$ = *número de participantes que han respondido correctamente en los grupos alto y bajo respectivamente.*

N_g = *número de participantes correspondientes al 27% de la muestra.*

Interpretación:

-Valores **altos:** los participantes del grupo alto obtienen más respuestas correctas que los del grupo bajo (discrimina adecuadamente)

-Valores **próximos a cero:** ambos grupos se encuentran próximos entre sí → no discrimina entre participantes de ambos grupos.

- Valores **negativos:** los participantes del grupo bajo obtienen más respuestas correctas que los del grupo alto → favorece la disminución de la precisión del instrumento de medida.

14

2.2. PARAMETROS DE LOS ÍTEMS. Índice de discriminación D

Índice de discriminación	Interpretación
$D \geq 0{,}40$	El ítem presenta un gran poder **discriminativo**
$0{,}30 \leq D \leq 0{,}39$	Discriminación **aceptable**
$0{,}20 \leq D \leq 0{,}29$	Discrimina **poco** y necesita una revisión
$0{,}10 \leq D \leq 0{,}19$	Ítems **no adecuados**, que deber ser modificados o eliminados del test
$D \leq 0{,}09$	Ítems que deben **eliminarse** directamente

TEMA 4: ANÁLISIS DE ÍTEMS

15

2.1., 2.2. PARAMETROS DE LOS ÍTEMS. Ejemplo

Se presentan las respuestas de los participantes de una muestra a un ítem de cuatro alternativas. De los 63 participantes se seleccionaron: 27% superior (17) y el 27% inferior.

	Alternativas			
Grupo	A	B	C*	D
Superior	1	0	**13**	3
Medio	3	4	**10**	12
Inferior	2	5	**5**	5

1. Calcular la dificultad corregida y el índice de discriminación D del ítem.

TEMA 4: ANÁLISIS DE ÍTEMS

16

2.1., 2.2. PARAMETROS DE LOS ÍTEMS. Ejemplo

Se presentan las respuestas de los participantes de una muestra a un ítem de cuatro alternativas. De los 63 participantes se seleccionaron: 27% superior (17) y el 27% inferior.

	Alternativas			
Grupo	A	B	**C***	D
Superior	1	0	**13**	3
Medio	3	4	**10**	12
Inferior	2	5	**5**	5

1. Calcular la dificultad corregida y el índice de discriminación D del ítem.

$$p_c = \frac{A - \frac{E}{k-1}}{N} = \frac{28 - \frac{35}{4-1}}{63} = 0{,}26$$

$$D = \frac{A_{alto} - A_{bajo}}{N_g} = \frac{13-5}{17} = 0{,}47$$

1. El índice de dificultad es 0,26, lo que indica que es relativamente difícil.

2. El índice D, es de 0,47, lo que demuestra que diferencia adecuadamente entre participantes de distinto nivel en el rasgo medido

TEMA 4: ANÁLISIS DE ÍTEMS

2.2. PARAMETROS DE LOS ÍTEMS. Coeficientes de discriminación. Correlación biserial-puntual

El grado de discriminación también puede medirse mediante el coeficiente de correlación. Tiene la ventaja, respecto al índice D, que tiene en cuenta a la **totalidad de la muestra**.

1. Correlación Biserial-puntual: el ítem es una variable dicotómica y la puntuación en el test es continua

$$r_{bp} = \frac{(\overline{X}_c - \overline{X})}{S_x}\sqrt{\frac{p}{q}}$$

Donde:

$\overline{X}_c$ = media de las puntuaciones obtenidas en el test por los participantes que han respondido correctamente al ítem.

$\overline{X}$ = media de las puntuaciones en el test calculada con todos los participantes de la muestra

S_x = desviación típica de las puntuaciones en el test con todos los participantes

p = dificultad del ítem

q = 1-p

18

TEMA 4: ANÁLISIS DE ÍTEMS

2.2. PARAMETROS DE LOS ÍTEMS.
Coeficientes de discriminación. Correlación biserial-puntual

Interpretación:

- Los valores oscilan entre -1 y +1. Es poco probable que sea menor que -0.10, o superior a 0.75.
- Se ha de corregir la asociación espuria que aparece al calcular la puntuación del test incluyendo la del ítem→ se correlaciona el ítem consigo mismo y por tanto se infla artificialmente el valor.

También se puede aplicar la siguiente fórmula de corrección:

$$\boxed{r_{bp(c)} = \frac{r_{bp}S_x - S_i}{\sqrt{S_i^2 + S_x^2 - 2r_{bp}S_iS_x}}}$$

A medida que el número de ítems aumenta r_{bp} se aproxima a $r_{bp(c)}$

TEMA 4: ANÁLISIS DE ÍTEMS

19

2.2. PARAMETROS DE LOS ÍTEMS.
Coeficientes de discriminación. Correlación biserial-puntual

Ejemplo. La siguiente tabla muestra las respuestas de 5 personas a 4 ítems. Calcular la correlación biserial-puntual del ítem 2.

	Ítems			
Participantes	**1**	**2**	**3**	**4**
A	0	1	0	1
B	1	1	0	1
C	1	1	1	1
D	0	0	0	1
E	1	1	1	0

TEMA 4: ANÁLISIS DE ÍTEMS

20

2.2. PARAMETROS DE LOS ÍTEMS.
Coeficientes de discriminación. Correlación biserial-puntual

	Ítems				Total		
Participantes	**1**	**2**	**3**	**4**	**X**	**(X-i)**	**$(X\text{-}i)^2$**
A	0	1	0	1	2	1	1
B	1	1	0	1	3	2	4
C	1	1	1	1	4	3	9
D	0	0	0	1	1	1	1
E	1	1	1	0	3	2	4
∑						**9**	**19**

2.2. PARAMETROS DE LOS ÍTEMS.
Coeficientes de discriminación. Correlación biserial-puntual

- Los participantes que respondieron correctamente el ítem fueron A, B, C y E; su media es:

$$\overline{X}_c = \frac{1+2+3+2}{4} = 2$$

- La media total es:

$$\overline{X}_T = \frac{9}{5} = 1{,}8$$

- La desviación típica del test es:

$$S_X = \sqrt{\frac{\sum X^2}{N} - \overline{X}^2} = \sqrt{\frac{19}{5} - 1{,}8^2} = \sqrt{0{,}56} = 0{,}75$$

2.2. PARAMETROS DE LOS ÍTEMS.
Coeficientes de discriminación. Correlación biserial-puntual

$$p = \frac{4}{5} = 0,8$$

$$q = \frac{1}{5} = 0,2$$

$$r_{bp} = \frac{\overline{X}_c - \overline{X}_T}{S_X}\sqrt{\frac{p}{q}} = \frac{2-1,8}{0,75}\sqrt{\frac{0,8}{0,2}} = 0,54$$

TEMA 4: ANÁLISIS DE ÍTEMS

23

2.2. PARAMETROS DE LOS ÍTEMS.
Coeficientes de discriminación. Correlación biserial

2. Correlación Biserial: ítem es una variable continua que ha sido dicotomizada y la puntuación en el test es continua

$$\boxed{r_b = \frac{(\overline{X}_c - \overline{X})}{S_x}\frac{p}{y}}$$

Donde:

$\overline{X}_c$ = *media de las puntuaciones obtenidas en el test por los participantes que han respondido correctamente al ítem.*

$\overline{X}$ = *media de las puntuaciones en el test calculada con todos los participantes de la muestra*

S_x = *desviación típica de las puntuaciones en el test con todos los participantes*

p= dificultad del ítem

y= valor de la ordenada que le corresponde a la puntuación típica en la curva normal que deja por debajo un área igual a p (se mira en tablas)

24

Con los datos del ejemplo anterior, calcular la correlación biserial del ítem 3.

TEMA 4: ANÁLISIS DE ÍTEMS

2.2. PARAMETROS DE LOS ÍTEMS.
Coeficientes de discriminación. Correlación biserial

	Ítems				Total		
Participantes	1	2	3	4	X	(X-i)	$(X\text{-}i)^2$
A	0	1	0	1	2	2	4
B	1	1	0	1	3	3	9
C	1	1	1	1	4	3	9
D	0	0	0	1	1	1	1
E	1	1	1	0	3	2	4
∑						**11**	**27**

TEMA 4: ANÁLISIS DE ÍTEMS

2.2. PARAMETROS DE LOS ÍTEMS.
Coeficientes de discriminación. Correlación biserial

- Los participantes que respondieron correctamente el ítem fueron C and E; sus medias son:

$$\overline{X}_c = \frac{3+2}{2} = 2{,}5$$

- La media total es:

$$\overline{X}_T = \frac{11}{5} = 2{,}2$$

- La desviación típica del test es:

$$S_X = \sqrt{\frac{\sum X^2}{N} - \overline{X}^2} = \sqrt{\frac{27}{5} - 2{,}2^2} = \sqrt{5{,}4 - 4{,}84} = \sqrt{0{,}56} = 0{,}75$$

TEMA 4: ANÁLISIS DE ÍTEMS

2.2. PARAMETROS DE LOS ÍTEMS. Coeficientes de discriminación. Correlación biserial

$$p = \frac{2}{5} = 0{,}4$$

$$r_b = \frac{\overline{X}_c - \overline{X}_T}{S_X}\frac{p}{y} = \frac{2{,}5 - 2{,}2}{0{,}75}\frac{0{,}4}{0{,}3863} = 0{,}4 * 1{,}03 = 0{,}41$$

Debido a que el valor *p* = 0,4 no aparece en la primera columna de la tabla, buscamos su complementario (0,6), que se asocia con una *y* = 0,3863.

TEMA 4: ANÁLISIS DE ÍTEMS

27

2.2. PARAMETROS DE LOS ÍTEMS. Coeficientes de discriminación. Correlación biserial-puntual y biserial

La **biserial** y la **biserial-puntual**, se pueden relacionar mediante:

$$\boxed{r_{bp} = r_b \frac{y}{\sqrt{pq}}}$$

Ventajas e inconvenientes:

1. La **biserial-puntual** es fiel reflejo de la contribución del ítem, ya que no supone ninguna habilidad continua subyacente.
2. La **biserial** se ve menos influida por la dificultad del ítem, y tiende a ser invariante de una aplicación a otra.

TEMA 4: ANÁLISIS DE ÍTEMS

28

2.2. PARAMETROS DE LOS ÍTEMS.
Coeficientes de discriminación. Coeficiente de correlación ϕ

-Coeficiente phi (ϕ): se necesita calcular la correlación entre la ejecución en un ítem dicotómico y un test en el que la muestra de participantes se divide en función de un punto de corte entre aptos-no aptos; clínicos-no clínicos, etc.→ utilizar tablas de contingencia:

		ítem		
		+	-	
Test	-	a	b	a+b
	+	c	d	c+d
		a+c	b+d	

Donde:

a = número de participantes con –NO ÉXITO- en el test y que aciertan el ítem

b = número de participantes con –NO ÉXITO- en el test y que no aciertan el ítem

c = número de personas con –ÉXITO- en el test y que aciertan el ítem.

d= número de personas con –ÉXITO- en el test y que no aciertan el ítem.

TEMA 4: ANÁLISIS DE ÍTEMS

29

2.2. PARAMETROS DE LOS ÍTEMS.
Coeficientes de discriminación. Coeficiente de correlación ϕ

Coeficiente **phi** (ϕ): para el cálculo se procede mediante la siguiente fórmula:

$$\Phi = \frac{bc - ad}{\sqrt{(a+b)(a+c)(c+d)(b+d)}}$$

		ítem		
		+	-	
Test	-	a	b	a+b
	+	c	d	c+d
		a+c	b+d	

TEMA 4: ANÁLISIS DE ÍTEMS

30

2.2. PARAMETROS DE LOS ÍTEMS.
Coeficientes de discriminación.
Coeficiente de correlación ϕ

Ejemplo. La tabla que se presenta a continuación contiene el resultado de 50 estudiantes en un examen y en un ítem de dicho examen. Calcular el coeficiente de correlación ϕ.

		Ítem 5 (X)	
		1 (acertado)	0 (fallido)
Test (Y)	0 (suspenso)	5	10
	1 (aprobado)	30	5

2.2. PARAMETROS DE LOS ÍTEMS.
Coeficientes de discriminación. Coeficiente de correlación ϕ

		Ítem 5 (X)		
		1 (acertado)	0 (fallido)	
Test (Y)	0 (suspenso)	5 (a)	10 (b)	15
	1 (aprobado)	30 (c)	5 (d)	35
		35	15	

$$\Phi = \frac{bc - ad}{\sqrt{(a+b)(a+c)(c+d)(b+d)}} = \frac{30*10 - 5*5}{\sqrt{15*35*35*15}} = \frac{275}{525} = 0{,}52$$

Existe una alta correlación entre el ítem y el test; es decir, los participantes que acertaron el ítem generalmente aprobaron el examen, mientras que quienes lo fallaron suspendieron en su mayoría.

2.2. PARAMETROS DE LOS ÍTEMS.
Discriminación de los ítems en diseños pre-post test

Discriminación de los ítems en diseños pre-post test: A veces, también es necesario determinar el grado de discriminación de un ítem entre un grupo de participantes que ha pasado por un proceso de instrucción. Disponemos de varios índices como:

1. *Índice de discriminación D*: diferencia entre la proporción de participantes que contestan un determinado ítem correctamente después y antes de recibir dicho período de instrucción.

$$\boxed{D = Ppost - Ppre}$$

Donde:

Ppost = proporción de participantes que contestan correctamente un ítem después de recibir la instrucción.

Ppre= proporción de participantes que contestan correctamente un ítem antes de recibir la instrucción.

2.2. PARAMETROS DE LOS ÍTEMS.
Discriminación de los ítems en diseños pre-post test.

Ejemplo: 40 personas participaron en un curso de formación. En la prueba previa al curso, 10 personas respondieron correctamente al ítem 3; tras el curso, 35 personas respondieron correctamente a este mismo ítem. Calcular el índice de discriminación D.

2.2. PARAMETROS DE LOS ÍTEMS.
Discriminación de los ítems en diseños pre-post test

$$D = Ppost - Ppre = 0{,}875 - 0{,}25 = 0{,}625$$

$$Ppost = \frac{35}{40} = 0{,}875$$

$$Ppre = \frac{10}{40} = 0{,}25$$

2.2. PARAMETROS DE LOS ÍTEMS.
Poder discriminativo de los ítems en las escalas de actitudes

Discriminación en las escalas de actitudes: existen procedimientos basados en métodos correlacionales; y métodos basados en la división de grupos extremos de actitud.

1. **Métodos correlacionales: coeficiente de correlación de Pearson o índice de homogeneidad (IH)**. Si las correlaciones obtenidas son nulas o bajas estaría indicando que el elemento no mide la misma dimensión de actitud → ítems que habría que eliminar.

$$\boxed{R_{jX} = \frac{N\sum JX - \sum J \sum X}{\sqrt{\left[N\sum J^2 - (\sum J)^2\right]\left[N\sum X^2 - (\sum X)^2\right]}}}$$

2.2. PARAMETROS DE LOS ÍTEMS.
Poder discriminativo de los ítems en las escalas de actitudes

Es necesario corregir la puntuación del test, eliminando la puntuación de los participantes en el test, o aplicar la fórmula:

$$R_{j(X-j)} = \frac{R_{jX} S_X - S_j}{\sqrt{S_X^2 + S_j^2 - 2R_{jX} S_X S_j}}$$

Eliminar ítems cuyo índice de homogeneidad sea inferior a 0,20

TEMA 4: ANÁLISIS DE ÍTEMS

37

2.2. PARAMETROS DE LOS ÍTEMS.
Poder discriminativo de los ítems en las escalas de actitudes

Ejemplo. La tabla que se presenta a continuación muestra las respuestas de 5 participantes a 4 ítems de actitud. Calcular la discriminación del ítem 4 utilizando el coeficiente de correlación de Pearson.

	Ítems			
Participantes	X_1	X_2	X_3	X_4
A	2	4	4	3
B	3	4	3	5
C	5	2	4	3
D	3	5	2	4
E	4	5	2	5

TEMA 4: ANÁLISIS DE ÍTEMS

38

2.2. PARAMETROS DE LOS ÍTEMS.
Poder discriminativo de los ítems en las escalas de actitudes

Ejemplo. La tabla que se presenta a continuación muestra las respuestas de 5 participantes a 4 ítems de actitud. Calcular la discriminación del ítem 4 utilizando el coeficiente de correlación de Pearson.

	Ítems				X_T	X_4X_T	X^2_4	X^2_T
Participantes	X_1	X_2	X_3	X_4				
A	2	4	4	3	13	39	9	169
B	3	4	3	5	15	75	25	225
C	5	2	4	3	14	42	9	196
D	3	5	2	4	14	56	16	196
E	4	5	2	5	16	80	25	256
				20	72	292	84	1042

39

TEMA 4: ANÁLISIS DE ÍTEMS

2.2. PARAMETROS DE LOS ÍTEMS.
Poder discriminativo de los ítems en las escalas de actitudes

- La correlación o IH entre el ítem 4 y la puntuación total del test es:

$$R_{jx} = \frac{N\sum JX - \sum J \sum X}{\sqrt{[N\sum J^2 - \left(\sum J\right)^2 *][N\sum X^2 - \left(\sum X\right)^2]}} =$$

$$\frac{5*292 - 20*72}{\sqrt{[5*84 - 20^2][5*1042 - 72^2]}} = 0,88$$

- Se trata de un resultado inflado, puesto que el ítem 4 está incluido en la puntuación total. Por ello, habría de hacerse la corrección:

40

TEMA 4: ANÁLISIS DE ÍTEMS

2.2. PARAMETROS DE LOS ÍTEMS.
Poder discriminativo de los ítems en las escalas de actitudes

– Medias y desviaciones típicas para el ítem 4 y para la puntuación total:

$$\overline{X}_4 = \frac{\sum X_4}{N} = \frac{20}{5} = 4 \qquad S_{X_4} = \sqrt{\frac{\sum X_4^2}{N} - \overline{X}_4^2} = \sqrt{\frac{84}{5} - 4^2} = \sqrt{0,8} = 0,89$$

$$\overline{X}_T = \frac{\sum X_T}{N} = \frac{72}{5} = 14,4 \qquad S_{X_T} = \sqrt{\frac{\sum X_T^2}{N} - \overline{X}_T^2} = \sqrt{\frac{1042}{5} - 14,4^2} = \sqrt{1,04} = 1,02$$

$$R_{j(X-j)} = \frac{R_{jX} S_X - S_j}{\sqrt{S_X^2 + S_j^2 - 2R_{jX} S_X S_j}} = \frac{0,88 * 1,02 - 0,89}{\sqrt{1,04 + 0,80 - 2 * 0,88 * 1,02 * 0,89}} = 0,01$$

TEMA 4: ANÁLISIS DE ÍTEMS

41

2.2. PARAMETROS DE LOS ÍTEMS.
Poder discriminativo de los ítems en las escalas de actitudes

- La gran diferencia encontrada en el resultado tras aplicar la corrección se debe al bajo número de ítems del ejemplo.
  – A medida que aumenta el número de ítems, esta diferencia va disminuyendo porque la influencia del ítem en la puntuación total va siendo menor. A partir de más de 25 ítems, los resultados encontrados son muy similares.

TEMA 4: ANÁLISIS DE ÍTEMS

42

2.2. PARAMETROS DE LOS ÍTEMS.
Poder discriminativo de los ítems en las escalas de actitudes

TEMA 4: ANÁLISIS DE ÍTEMS

Discriminación en las escalas de actitudes:

2. Método de división en grupos extremos: se eligen dos grupos extremos (25 ó 27%, generalmente) de actitud a partir de las puntuaciones en la escala total; y se comparan los elementos. Si no son discriminativos, no encontraremos diferencias significativas. Para ello usamos una diferencia de medias mediante *T de Student.*

$$T = \frac{\overline{X}_{sj} - \overline{X}_{ij}}{\sqrt{\frac{(n_s - 1)S_{sj}^2 + (n_i - 1)S_{ij}^2}{n_s + n_i - 2}\left[\frac{1}{n_s} + \frac{1}{n_i}\right]}}$$

-Donde:

$\overline{X}_{sj}$ *=es la media de las puntuaciones obtenidas en el elemento j por aquellos participantes, que en la escala total, obtuvieron puntuaciones más* ***altas***

$\overline{X}_{ij}$ *=es la media de las puntuaciones obtenidas en el elemento j por aquellos participantes, en la escala total, obtuvieron puntuaciones más* ***bajas***

S_{sj}^2 *=es la varianza de las puntuaciones obtenidas en el elemento j por los participantes del* ***grupo superior***

S_{ij}^2 *=es la varianza de las puntuaciones obtenidas en el elemento j por los participantes del* ***grupo inferior***

n_s *y* n_i*= son respectivamente el número de participantes que componen el grupo superior e inferior*

43

2.2. PARAMETROS DE LOS ÍTEMS.
Poder discriminativo de los ítems en las escalas de actitudes

TEMA 4: ANÁLISIS DE ÍTEMS

Ejemplo: utilizando los mismos datos del ejemplo anterior, calcular la *T* de Student para el ítem 2. ¿La media del grupo superior es significativamente mayor que la media del grupo inferior? (α = 0,05).

- Para calcular la discriminación del ítem 2 utilizando la *T* de Student, tenemos que hacer dos grupos con las puntuaciones extremas. Por razones didácticas, en este ejemplo cada grupo está conformado por sólo dos personas.

	Participantes	**X_2**
Grupo superior	E (16)	5
	B (15)	4
	Participantes	**X_2**
Grupo inferior	A (13)	4
	C (14)	2

44

2.2. PARAMETROS DE LOS ÍTEMS.
Poder discriminativo de los ítems en las escalas de actitudes

	Participantes	X_2	X_2^2
Grupo superior	E (16)	5	25
	B (15)	4	16
	∑	9	41
	Participantes	X_2	X_2^2
Grupo inferior	A (13)	4	16
	C (14)	2	4
	∑	6	20

$$\overline{X}_{sj} = \frac{\sum X_{sj}}{n_s} = \frac{9}{2} = 4{,}5$$

$$\overline{X}_{ij} = \frac{\sum X_{ij}}{n_i} = \frac{6}{2} = 3$$

TEMA 4: ANÁLISIS DE ÍTEMS

45

2.2. PARAMETROS DE LOS ÍTEMS.
Poder discriminativo de los ítems en las escalas de actitudes

$$S_{sj}^2 = \frac{\sum X_{sj}^2}{n_s} - \overline{X}_{sj}^2 = \frac{41}{2} - 4{,}5^2 = 20{,}5 - 20{,}25 = 0{,}25$$

$$S_{ij}^2 = \frac{\sum X_{ij}^2}{n_i} - \overline{X}_{ij}^2 = \frac{20}{2} - 3^2 = 10 - 9 = 1$$

$$T = \frac{\overline{X}_{sj} - \overline{X}_{ij}}{\sqrt{\frac{(n_s-1)S_{sj}^2 + (n_i-1)S_{ij}^2}{n_s+n_i-2}\left[\frac{1}{n_s}+\frac{1}{n_i}\right]}} = \frac{4{,}5-3}{\frac{(2-1)0{,}25+(2-1)1}{2+2-2}\left(\frac{1}{2}+\frac{1}{2}\right)} = 1{,}9$$

Una cola: $T_{(\alpha, ns+ni-2)} = T_{(0,05,\ 2+2-2)} = T_{(0,05,\ 2)} = 2{,}92$

1,9 < 2,92 – Se acepta la hipótesis nula. La media del grupo superior no es significativamente mayor que la media del grupo inferior. El ítem no discrimina adecuadamente.

TEMA 4: ANÁLISIS DE ÍTEMS

46

2.2. PARAMETROS DE LOS ÍTEMS.
Poder discriminativo de los ítems en las escalas de actitudes

Cuando se violan los supuestos de normalidad y/o igualdad de varianzas entre las poblaciones de las que proceden las muestras, podemos utilizar la ***U* de Mann-Whitney**

Procedimiento:

1. Se hace una ordenación conjunta de todos los participantes, asignando el orden 1 a la puntuación más baja.
2. Se suman por separado los órdenes de cada grupo y se halla la ***U*** para cada uno de ellos mediante la fórmula
3. Se consulta en las tablas el valor crítico de ***U*** (si es mayor indica que hay diferencias significativas).

$$U_s = n_s n_i + \frac{n_s(n_s+1)}{2} - R_s$$

$$U_i = n_s n_i + \frac{n_i(n_i+1)}{2} - R_i$$

-Donde:

R_s *=la suma de los órdenes del grupo* ***superior***

R_i *=la suma de los órdenes del grupo* ***inferior***

2.3., 2.4. PARAMETROS DE LOS ÍTEMS.
Fiabilidad y validez del ítem

Índices de fiabilidad y validez de los ítems: pueden interpretarse directamente en relación con las propiedades psicométricas más relevantes del test (fiabilidad y validez). Son estadísticos que están en función de la correlación del ítem con el test o criterio, respectivamente.

$$\boxed{IF = S_j r_{jX}}$$

Donde:

S_j*= desviación típica del ítem.*

r_{jX}*=correlación del ítem con la puntuación en el* ***test***.

$$\boxed{IV = r_{jY}}$$

Donde:

r_{jY}*=correlación del ítem con la puntuación en el* ***criterio***.

2.3., 2.4. PARAMETROS DE LOS ÍTEMS. Fiabilidad y validez del ítem

- Ejemplo: Con la información que se presenta a continuación, calcular el IF y el IV del ítem 4.

	p	$R_{bp\ test}$	$R_{bp\ criterio}$
Ítem 4	0,47	0,5	0,4

TEMA 4: ANÁLISIS DE ÍTEMS

49

2.3., 2.4. PARAMETROS DE LOS ÍTEMS. Fiabilidad y validez del ítem

$$IF = S_j r_{jX} = 0{,}5 * 0{,}5 = 0{,}25$$

$$S_j^2 = pq = 0{,}47 * 0{,}53 = 0{,}25$$

$$q = 1 - p = 1 - 0{,}47 = 0{,}53$$

$$S_j = \sqrt{S_j^2} = \sqrt{0{,}25} = 0{,}5$$

$$IV = r_{jY} = 0{,}4$$

TEMA 4: ANÁLISIS DE ÍTEMS

50

2.3., 2.4. PARAMETROS DE LOS ÍTEMS.
Fiabilidad y validez del ítem

Relación entre validez y fiabilidad: La validez del test se puede expresar a través de los coeficientes de fiabilidad y validez de los ítems:

Importancia de la ecuación porque expresa la validez del test a partir de:

$$r_{xy} = \frac{\sum_{j=1}^{n} S_j r_{jy}}{\sum_{j=1}^{n} S_j r_{jx}} = \frac{\sum_{j=1}^{n} S_j IV_j}{\sum_{j=1}^{n} IF_j}$$

1. La discriminación de los ítems (r_{jx})
2. La validez de los ítems (r_{jy})
3. La dificultad de los ítems (S^2_j=pq)

Paradoja: si queremos seleccionar ítems para maximizar la fiabilidad del test tendremos que elegir aquellos cuya discriminación (r_{jx}) sea alta. Pero ello, implica reducir el coeficiente de validez del test porque ésta aumenta a medida que los índices de validez de los ítems son elevados y los de fiabilidad bajos. Por tanto, incrementar la validez, o la fiabilidad del test a partir de la selección de los ítems, es una cuestión que ha de ser sometida al criterio del investigador

TEMA 4: ANÁLISIS DE ÍTEMS

51

2.5. PARAMETROS DE LOS ÍTEMS.
Comparación de las respuestas a los ítems

Supongamos que se han aplicado dos ítems a una muestra de 200 participantes, y que los datos se distribuyen según la siguiente tabla:

	Ítem 1	
Ítem 2	Acierto	Error
Acierto	65 (a)	35 (b)
Error	35 (c)	65 (d)

¿Es equivalente el grado de dificultad de los ítems? (NC = 95%). Podemos utilizar la *Chi-cuadrado* propuesta por Harris y Pearlman (1977). La hipótesis que planteamos es si la ejecución en ambos ítems es igual o no.

$$\chi^2 = \frac{(|b-c|-1)^2}{b+c}$$

$X^2 \leq X^2_{(\alpha,\ (filas-1)*(columnas-1))}$: Se acepta la hipótesis nula. El grado de dificultad de ambos ítems es equivalente.

$X^2 > X^2_{(\alpha,\ (filas-1)*(columnas-1))}$: Se rechaza la hipótesis nula. El grado de dificultad de ambos ítems es diferente.

TEMA 4: ANÁLISIS DE ÍTEMS

52

2.5. PARAMETROS DE LOS ÍTEMS.
Comparación de las respuestas a los ítems

$$\chi^2 = \frac{(|b-c|-1)^2}{b+c} = \frac{(|35-35|-1)^2}{70} = 0{,}014$$

$$\chi^2_{(0,05;1)} = 3{,}84$$

$$NC = 95\%\ (0{,}014 \le 3{,}84)$$

Dado que a un nivel de confianza del 95% el valor obtenido es menor que el de las tablas, podemos concluir que la **dificultad** de ambos ítems es igual, o que no presentan valores significativamente distintos

TEMA 4: ANÁLISIS DE ÍTEMS

53

3. ANÁLISIS DE DISTRACTORES

ANÁLISIS DE DISTRACTORES:

Si tras análisis de ítems, aparecen elementos poco discriminativos o con dificultad extrema, pero que deseamos mantener → Revisión de las alternativas o distractores.

El **objetivo de un distractor** es atraer la atención de los participantes con nivel *medio o bajo* en el constructo y que no tienen suficiente nivel para responder correctamente al ítem. Se consideran buenos si:

1. Son elegidas por un mínimo de participantes (al menos el **10%).**
2. Son aproximadamente igual de atractivas para los participantes (**misma probabilidad**).
3. El rendimiento medio en el test de los participantes que han elegido cada distractor sea inferior al de los participantes que han seleccionado la correcta→ la media en el test de los participantes que eligen la correcta es **superior** a la media de los participantes que seleccionan la incorrecta.
4. Que discriminen entre los participantes de bajo, medio, y alto nivel en el rasgo medido, pero en sentido contrario a como lo hace la alternativa correcta→ índice de discriminación alto y **negativo** (**poder discriminativo**)

TEMA 4: ANÁLISIS DE ÍTEMS

54

3. ANÁLISIS DE DISTRACTORES.
3.1. Misma probabilidad de los distractores

- Los distractores son igualmente probables si son seleccionados por un número mínimo de participantes y si son igual de atractivos para quienes no saben la respuesta correcta.

- *X*²:

$$\chi^2 = \sum_{j=1}^{k} \frac{(FT_i - FO_i)^2}{FT_i}$$

FT_i = Frecuencias teóricas (esperadas).

FO_i = Frecuencias observadas.

TEMA 4: ANÁLISIS DE ÍTEMS

55

3. ANÁLISIS DE DISTRACTORES.
3.1. Misma probabilidad de los distractores

- Grados de libertad: K -1 (K = número de alternativas incorrectas).

- Ho: $FT_i = FO_i$ (en los participantes que no saben la respuesta correcta, la elección de cualquier distractor es igual de atractiva).

- Conclusión:
  - $\boxed{\chi_O^2 \leq \chi_{(\alpha,k-1)}^2}$ → Se acepta la hipótesis nula. Los distractores son igualmente probables.
  - $\boxed{\chi_O^2 > \chi_{(\alpha,k-1)}^2}$ → Se rechaza la hipótesis nula. Los distractores no son igualmente probables.

TEMA 4: ANÁLISIS DE ÍTEMS

56

3. ANÁLISIS DE DISTRACTORES.
3.1. Misma probabilidad de los distractores

Ejemplo. Determinar si las alternativas incorrectas son igual de atractivas (α = 0,05).

	A	B*	C
Número de respuestas	136	142	92

TEMA 4: ANÁLISIS DE ÍTEMS

57

3. ANÁLISIS DE DISTRACTORES.
3.1. Misma probabilidad de los distractores

$$\chi^2 = \sum_{j=1}^{k} \frac{(FT_i - FO_i)^2}{FT_i} = \frac{(114-136)^2 + (114-92)^2}{114} =$$

$$= \frac{22^2 + 22^2}{114} = \frac{484+484}{114} = \frac{968}{114} = 8{,}49$$

$$FT_i = \frac{136+92}{2} = \frac{228}{2} = 114$$

Para ser igualmente probables, cada distractor habría de ser elegido por 114 participantes.

TEMA 4: ANÁLISIS DE ÍTEMS

58

3. ANÁLISIS DE DISTRACTORES.
3.1. Misma probabilidad de los distractores

$$\chi^2_{(\alpha,k-1)} = \chi^2_{(0,05;2-1)} = \chi^2_{(0,05;1)} = 3{,}84$$

8,49 > 3,84 → Se rechaza la hipótesis nula. Las alternativas incorrectas no son igualmente atractivas para los participantes, aunque cumplieron el criterio de haber sido seleccionadas por un mínimo del 10% de la muestra (N).

$$N = 136 + 142 + 92 = 370$$

$$10\% = \frac{370 * 10}{100} = 37$$

$$136 > 37$$

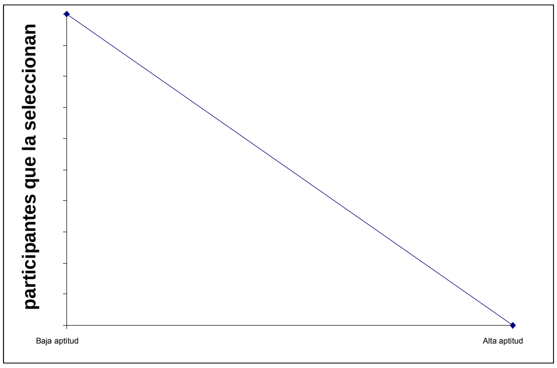
$$92 > 37$$

TEMA 4: ANÁLISIS DE ÍTEMS

59

3. ANÁLISIS DE DISTRACTORES.
3.2. Poder discriminativo de los distractores

Discriminación alternativas incorrectas: A medida que la aptitud aumenta, la alternativa incorrecta es seleccionada por menos participantes



TEMA 4: ANÁLISIS DE ÍTEMS

60

3. ANÁLISIS DE DISTRACTORES
3.2. Poder discriminativo de los distractores

			Alternativas del ítem				
			A*	B	C	D	E
Nivel de aptitud	Superior		65	11	64	32	28
	inferior		15	30	20	68	67
Estadísticos		P	0,30	0,08	0,19	0,21	0,22
		$\overline{X}$	11,1	8,3	13,2	8,9	7,8
		D	0,25	-0,09	0,22	-0,18	-0,19

1. Alternativa B: *no funciona* como un distractor eficaz ya que no es elegida por un mínimo de personas (0,08).

2. Alternativa C: *no funciona* como un distractor eficaz ya que la media es superior (13,2) a los que han elegido la correcta→ atrae a participantes competentes

3. Alternativas D y E Funcionan adecuadamente ya que son respondidas por un mínimo de personas y sus índices de discriminación son negativos (-,18; y -,19)→ atraen a los participantes menos competentes de la muestra.

TEMA 4: ANÁLISIS DE ÍTEMS

61

3. ANÁLISIS DE DISTRACTORES
3.2. Poder discriminativo de los distractores

Ejemplo. La tabla que se presenta a continuación contiene las respuestas de 5 participantes a 4 ítems. Entre paréntesis se muestra la alternativa seleccionada por cada participante. La respuesta correcta está marcada con un asterisco. Calcular la discriminación del distractor b en el ítem 3.

	Ítems			
Participantes	**1(a*)**	**2(b*)**	**3(a*)**	**4(c*)**
A	0 (b)	1	0 (b)	1
B	1	1	0 (b)	1
C	1	1	1	1
D	0 (c)	0 (a)	0 (b)	1
E	1	1	1	0 (b)

TEMA 4: ANÁLISIS DE ÍTEMS

62

3. ANÁLISIS DE DISTRACTORES
3.2. Poder discriminativo de los distractores

	Ítems				Total		
Participantes	1(a*)	2(b*)	3(a*)	4(c*)	X	(X-i)	$(X\text{-}i)^2$
A	0 (b)	1	0 (b)	1	2	2	4
B	1	1	0 (b)	1	3	3	9
C	1	1	1	1	4	3	9
D	0 (c)	0 (a)	0 (b)	1	1	1	1
E	1	1	1	0 (b)	3	2	4
Σ						11	27

$$r_{bp} = \frac{\overline{X}_1 - \overline{X}_T}{S_X}\sqrt{\frac{p}{q}} = \frac{2-2,2}{0,75}\sqrt{\frac{0,4}{0,6}} = -0,22$$

Al encontrarse un resultado negativo, el distractor puede considerarse bueno (fue principalmente elegido por participantes con bajo nivel de conocimiento).

TEMA 4: ANÁLISIS DE ÍTEMS

63

3. ANÁLISIS DE DISTRACTORES
3.2. Poder discriminativo de los distractores

Cálculos:

- Media de las puntuaciones en el test de los participantes que **seleccionaron la alternativa b** en el ítem 3 (participantes A, B y D):

$$\overline{X}_1 = \frac{2+3+1}{3} = \frac{6}{3} = 2$$

- Los demás cálculos se realizan como es usual (nada cambia):

$$\overline{X}_T = \frac{11}{5} = 2,2$$

$$S_X = \sqrt{\frac{27}{5} - 2,2^2} = \sqrt{5,4-4,84} = \sqrt{0,56} = 0,75$$

$$p = \frac{2}{5} = 0,4$$

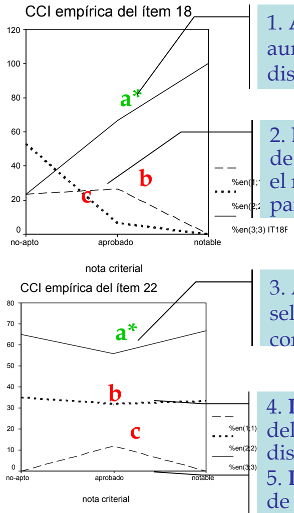
$$q = 1 - p = 1 - 0,4 = 0,6$$

TEMA 4: ANÁLISIS DE ÍTEMS

64

3. ANÁLISIS DE DISTRACTORES.
3.2. Poder discriminativo de los distractores. Utilidad de los métodos gráficos

REPRESENTACIÓN GRÁFICA:



1. **Alternativa A (correcta)**: a medida que el nivel de aptitud aumenta, es seleccionada por más participantes → discriminación positiva

2. **Distractores B y C**: la tendencia es la contraria. En niveles de aptitud bajo, son igualmente seleccionadas, y a medida que el nivel de aptitud aumenta la seleccionan cada vez menos participantes→ discriminación negativa

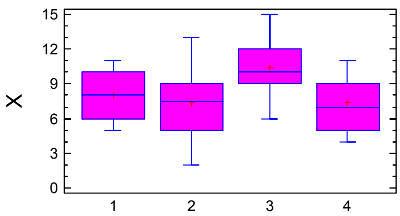
3. **Alternativa A (correcta)**: aproximadamente es igualmente selecciona entre participantes poco competentes y muy competentes → discriminación baja o próxima a cero.

4. **Distractor B**: es igualmente seleccionada indistintamente del nivel de aptitud de los participantes (línea plana) → discriminación próxima a cero.
5. **Distractor C**: Prácticamente a lo largo de todo el continuo de aptitud lo identifican como falso, y por tanto no lo seleccionan ni los participantes menos competentes.

TEMA 4: ANÁLISIS DE ÍTEMS

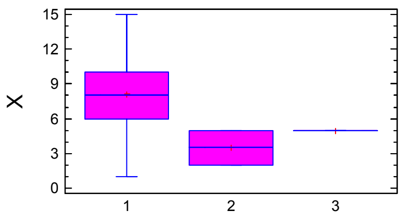
3. ANÁLISIS DE DISTRACTORES.
3.2. Poder discriminativo de los distractores. Utilidad de los métodos gráficos

Diagrama de caja y bigotes de un ANOVA de un BUEN ítem



La media de los participantes que han seleccionado la opción correcta (*3*) es más alta en el test que los que han seleccionado las incorrecta y estas a su vez no difieren entre si.

Diagrama de caja y bigotes de un ANOVA de un MAL ítem



-La opción *4* no ha sido seleccionada.
-La opción *3* sólo la elige un participante.
-La opción *2* presenta una variabilidad pequeña
-La opción correcta (*1*) ha sido respondida indistintamente por participantes de baja y alta aptitud.

66

TEMA 4: ANÁLISIS DE ÍTEMS

TEMA 4: ANÁLISIS DE ÍTEMS

4. Funcionamiento diferencial de los ítems

67

TEMA 4: ANÁLISIS DE ÍTEMS

4.1. EL FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM. Conceptos de sesgo, impacto y funcionamiento diferencial del ítem.

FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM:

¿miden todos los ítems de la misma manera a todos los **grupos** de participantes a los que se han administrado?→en caso contrario, estaríamos perjudicando sistemáticamente a determinados participantes en función de su pertenencia a determinados grupos.

Martínez-Arias (1995): *"Las diferencias entre grupos encontradas en los resultados de tests de aptitudes y rendimiento ¿reflejan diferencias reales entre los grupos o están causadas por* ***fuentes sistemáticas*** *de variación ajenas al constructo que mide el test?"*

Fuentes sistemáticas de variación, son las que originan el sesgo, y afecta directamente a la validez de constructo, → la introducción de componentes irrelevantes para el constructo.

68

4. EL FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM. Conceptos de sesgo, impacto y funcionamiento diferencial del ítem.

Ejemplo: Se ha aplicado un test de 35 elementos que intenta medir el rendimiento en matemáticas de los alumnos de 2ª de la ESO→ factores como los contenidos elegidos pueden hacer que las niñas tengan más probabilidades de éxito en responder correctamente a los ítems. ¿Factores ajenos al rendimiento en matemáticas están propiciando que dentro del mismo nivel en el rasgo medido niños y niñas obtengan resultados distintos?

El **Sesgo** analiza las posibles fuentes de variación que puedan beneficiar más a un subgrupo de participantes que a otros. Un Ítem está sesgado, en la medida en que para dos participantes o grupos con el mismo valor en la variable medida, se generan mediciones distintas.

El estudio del sesgo se realiza a través de distintas técnicas englobadas bajo el epígrafe de **Funcionamiento Diferencial del Ítem (DIF)**: Se dice que un ítem presenta DIF si comparando grupos de participantes en función de una característica sociodemográfica (generalmente), se observa que participantes con el mismo nivel en la variable medida tienen diferentes posibilidades de acertar el ítem.

4.1. EL FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM. Conceptos de sesgo, impacto y funcionamiento diferencial del ítem.

Conviene diferenciar claramente entre impacto y Funcionamiento diferencial:

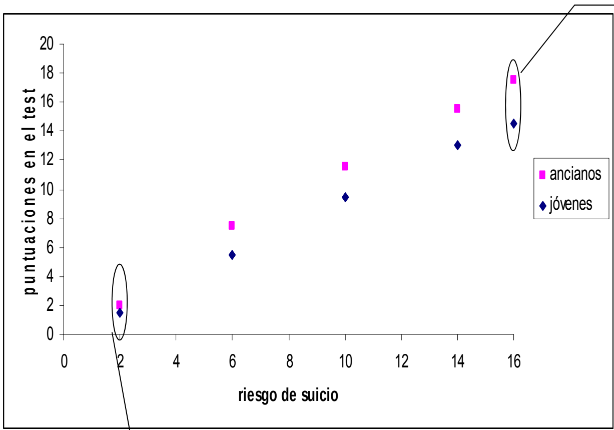
Impacto: Se dice que un ítem presenta impacto cuando existen diferencias **reales** en la puntuación media obtenida en ese ítem por dos grupos de participantes con **distinto nivel** en el rasgo o característica que mide el test.

Funcionamiento Diferencial: Un ítem presenta DIF cuando existen diferencias en la puntuación media obtenida en ese ítem por dos grupos de participantes pero con el **mismo nivel** en el rasgo o característica evaluada por el test (por ejemplo anglosajones e hispanos que responden a un test sobre matemáticas realizado en inglés).

La presencia de impacto, implica diferencias reales entre los participantes de ambos grupos, mientras que el DIF indica que no son reales, sino que se deben al instrumento de medida utilizado

4.1. EL FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM. Conceptos de sesgo, impacto y funcionamiento diferencial del ítem.

Ejemplo: test para detectar el riesgo de suicidio entre pacientes clínicos está sesgado en función de la edad.



1. Cuando el riesgo de suicidio es 16, los jóvenes obtienen en el test una puntuación menor que los ancianos. → participantes jóvenes que procuran atención psicológica urgente no la reciban.

2. Cuando el riesgo es despreciable. El test ofrece puntuaciones similares para ambos grupos

TEMA 4: ANÁLISIS DE ÍTEMS

71

4.2. EL FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM. Detección del DIF.

Evaluación del DIF → cuidadoso análisis por parte de varios expertos del contenido del ítem que pueda estar perjudicando sistemáticamente a un grupo respecto a otros. A esto se le denomina DIF (sustantivo), que hay que completar con procedimientos estadísticos o DIF estadístico:

Procedimientos estadísticos:

1. **χ^2 de los aciertos**: dividir las puntuaciones de los dos grupos analizados en distintos niveles (5 y 10, normalmente). Si el ítem no está sesgado es de esperar que las proporciones de aciertos sean iguales en los distintos niveles para los dos grupos. Aplicar χ^2 propuesta por Scheuneman (1979)

$$\chi^2 = \sum_{j=1}^{k} \frac{(FO_i - FT_i)^2}{FT_i}$$

$X^2_{(\alpha,\, k-1)}$

k = número de categorías en que se dividen las puntuaciones de X

TEMA 4: ANÁLISIS DE ÍTEMS

72

EL FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM. Detección del DIF, $\chi 2$ de los aciertos

1. Se aplica un test a 400 participantes, 200 hombres y 200 mujeres, cuyas puntuaciones se dividieron en 5 categorías. Se desea estudiar el posible sesgo de un ítem que por incluir estímulos visuales más familiares a los hombres pudiera estar sesgando a las mujeres (NC: 99%).

X	Número de participantes			Aciertos			Proporción de aciertos
	Mujer	Hombre	Total	Mujer	Hombre	Total	
20-24	20	15	35	15	10	25	25/35=0,71
15-19	100	105	205	70	85	155	155/205=0,76
10-14	50	40	90	10	30	40	40/90=0,44
5-9	20	30	50	5	20	25	25/50=0,50
0-4	10	10	20	0	5	5	5/20=0,25
Total	**200**	**200**	**400**	**100**	**150**	**250**	

1. Expresa el continuo de aptitud dividido en 5 categorías.

2. Expresa la proporción de aciertos. Es decir, el nº de aciertos entre el total de participantes en cada categoría.

TEMA 4: ANÁLISIS DE ÍTEMS

73

EL FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM. Detección del DIF, $\chi 2$ de los aciertos

Se necesita la frecuencia esperada en hombres y mujeres. Para ello, se multiplica la **proporción de aciertos de cada nivel por la frecuencia de** mujeres y hombres respectivamente.

X	*Frecuencias esperadas (H_0)*	
	Mujeres	Hombres
20-24	20*0,71=**14,20**	15*0,71=**10,65**
15-19	100*0,76=**76**	105*0,76=**79,8**
10-14	50*0,44=**22**	40*0,44=**17,6**
5-9	20*0,50=**10**	30*0,50=**15**
0-4	10*0,25=**2,50**	10*0,25=**2,50**

De esta forma, obtenemos los valores teóricos, es decir, aquellos que debieron haberse obtenido en caso de que la **proporción de aciertos fuese la misma para ambos grupos**.

TEMA 4: ANÁLISIS DE ÍTEMS

74

4.2. EL FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM. Detección del DIF, χ2 de los aciertos

Finalmente, una vez conocidas las frecuencias empíricas y teóricas podemos calcular el valor de Chi-cuadrado empírico

X	Frecuencias empíricas y teóricas			
	Mujeres		Hombres	
	Empíricas	Teóricas	Empíricas	Teóricas
20-24	15	14,20	10	10,65
15-19	70	76	85	79,80
10-14	10	22	30	17,6
5-9	5	10	20	15
0-4	0	2,5	5	2,5

$$\chi^2 = \frac{(15-14{,}2)^2}{14{,}2} + \ldots + \frac{(5-2{,}5)^2}{2{,}5} = 25{,}34$$

$$\chi^2_{(0{,}99;4)} = 13{,}28$$

Otenemos un valor de 25,34 que comparado con el teórico para un N.C del 99% y 4 g.l encontramos que es mayor → ha de rechazarse la hipótesis nula. Es decir, **ITEM SESGADO**

$\boxed{NC = 99\% \ (25{,}34 \geq 13{,}28) \rightarrow \text{rechazamos la hipótesis nula : ítem sesgado}}$

TEMA 4: ANÁLISIS DE ÍTEMS

75

4.2. EL FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM. Detección del DIF, el procedimiento de Mantel-Haenszel.

Procedimientos estadísticos:

2. El procedimiento de Mantel-Haenszel: dado su sencillez y buenos resultados es uno de los métodos más utilizados en la actualidad. Puede considerarse una extensión de los métodos basados en la Chi-Cuadrado como el anterior.

1. Seleccionar la variable externa que se sospeche que pueda estar generando DIF. Se diferencia entre grupo de referencia (GR) que usualmente es el grupo favorecido; y grupo focal (GF), usualmente el grupo desfavorecido.
2. Subdividir a los participantes de ambos grupos en función de la puntuación empírica obtenida en el test en subgrupos de habilidad homogénea.
3. Calcular el número de respuestas correctas e incorrectas por cada grupo y nivel de habilidad (*k*), y se organiza mediante una tabla de contingencia de 2*2 (tantas como niveles de habilidad, *k*).

TEMA 4: ANÁLISIS DE ÍTEMS

76

4.2. EL FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM. Detección del DIF, el procedimiento de Mantel-Haenszel.

	Correctas	Incorrectas	
GR	A_K	B_K	N_R
GF	C_K	D_K	N_F
	N_1	N_2	N_K

1. A= nº de participantes del G.R que aciertan el ítem en el nivel *k* de habilidad.

4. Estimar la cantidad de funcionamiento diferencial mediante la siguiente expresión:

$$\alpha_{MH} = \frac{\sum_{k=1}^{n} \frac{A_k D_k}{N_k}}{\sum_{k=1}^{n} \frac{B_k C_k}{N_k}}$$

5. Interpretar los resultados: Los valores de alfa oscilan entre 0 e infinito. Valores mayores que 1 indican que el ítem favorece al grupo de referencia. Valores menores que 1 al focal. Y valores próximos a 1 indica que no hay DIF.

TEMA 4: ANÁLISIS DE ÍTEMS

77

4.2. EL FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM. Detección del DIF, el procedimiento de Mantel-Haenszel.

Se sospecha que en un ítem de matemáticas esté sesgado en contra de las niñas. Para descartar esa posibilidad se llevó a cabo un análisis del FDI. Se formaron 4 grupos de aptitud. Los resultados se muestran en la siguiente tabla, donde se muestran las respuestas correctas e incorrectas en cada grupo de aptitud tanto en niños como en niñas:

	Niños		Niñas	
X	**Correctas**	**Incorrectas**	**Correctas**	**Incorrectas**
0-10	**1**	**8**	**2**	**8**
11-20	**13**	**58**	**10**	**50**
21-30	**30**	**51**	**19**	**84**
31-40	**69**	**15**	**47**	**35**

TEMA 4: ANÁLISIS DE ÍTEMS

78

4.2. EL FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM. Detección del DIF, el procedimiento de Mantel-Haenszel.

TEMA 4: ANÁLISIS DE ÍTEMS

A partir de la tabla se forman tantas tablas de contingencia de 2*2 como grupo de aptitud:

	Niños (G.R)		Niñas (G.F)	
X	Correctas	Incorrectas	Correctas	Incorrectas
0-10	1	8	2	8
11-20	13	58	10	50
21-30	30	51	19	84
31-40	69	15	47	35

(0-X-10)	Correctas	Incorrectas	
GR	1	8	
GF	2	8	
			19

En el nivel de aptitud comprendido entre 0 y 10 hay 19 participantes. En el G.R 1 ha acertado el ítem, mientras que del G.F han sido 2.

79

4.2. EL FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM. Detección del DIF, el procedimiento de Mantel-Haenszel.

TEMA 4: ANÁLISIS DE ÍTEMS

Construimos tantas tablas de contingencia como niveles de aptitud; y aplicamos la fórmula de Mantel-Haenzel:

$$\alpha_{MH} = \frac{\frac{1*8}{19} + \ldots + \frac{69*35}{166}}{\frac{2*8}{19} + \ldots + \frac{47*15}{166}} = 2{,}28$$

Dado que está por encima de 1, podemos concluir que existe funcionamiento diferencial que beneficia al grupo de referencia. Es decir, el ítem favorece a los niños y perjudica a las niñas con el mismo nivel en el rasgo medido.

80

4.2. EL FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM. Detección del DIF, el procedimiento de Mantel-Haenszel.

Limitaciones:

1. No detecta DIF no uniforme, que a diferencia del uniforme en el que el ítem perjudica sistemáticamente al grupo focal en todos los niveles de aptitud, el no uniforme para unas categorías perjudica a un grupo y para otra perjudica al otro grupo.
2. A medida que se reduce el número de categorías, aumenta la probabilidad de catalogar ítems con DIF.

5. BIBLIOGRAFÍA COMENTADA

1. Barbero, I., García, E. Vila, E., y Holgado, F.P. (2010). *Psicometría: Problemas resueltos*. Madrid: Sanz y Torres.
Se trata de un libro de ejercicios y problemas en el que se incluye el desarrollo de la solución. El alumno podrá completar desde un punto de vista aplicado los conceptos y contenidos vistos en la parte teórica; así como adquirir las destrezas necesarias para la resolución de problemas.

2. Barbero, I. (Coord.) **, Vila, E. y Holgado, F.P.** (2010). Psicometría. Madrid: Sanz y Torres.
En el **capítulo 8** se realiza una profusa introducción al concepto de análisis de los ítems y se presentan con muchos ejemplos cada los parámetros de los ítems más relevantes. El capítulo termina con una sencilla explicación del FDI.

3. Fidalgo, A.M. (1996). Funcionamiento diferencial de los ítems. En J. Muñiz (Coord.), *Psicometría*. Madrid: Síntesis.
Este Capítulo puede servir para preparar los contenidos relacionados con el estudio del DIF, particularmente los apartados 9.4 y 9.5 en los que se abordan sus causas y se presentan algunas técnicas estadísticas para su detección.

5. BIBLIOGRAFÍA COMENTADA

4. García Cueto, E. (1993). *Introducción a la Psicometría.* Madrid: Siglo XXI.
El Capítulo 12 se presentan los índices estadísticos más habituales. Un buen resumen de los principales contenidos con algunos ejercicios de autoevaluación al final del tema.

5. Martínez Arias, R. (1995). *Psicometría: Teoría de los Tests Psicológicos y Educativos.* Madrid: Síntesis.
El Capítulo 18 se puede utilizar como referencia para la preparación de los aspectos teóricos y aplicados. Se presentan los principales índices estadísticos clásicos para la evaluación de la calidad de los ítems, tanto en los TRN como en TRC.