

TECNICAS RAPIDAS DE ANALISIS DE ITEMS

Por *RUBEN FARRAY*

Evaluar convenientemente (1) es una de las responsabilidades principales de toda persona que tiene a su cargo una función docente. En una sencilla calificación quedan resumidos todos los esfuerzos de comunicación y motivación del profesor, y de interés y aprendizaje del estudiante.

Este artículo no tiene como objetivo indicar las pautas para efectuar una buena evaluación, sino mostrar unas técnicas que permitan al profesor juzgar si su evaluación ha sido adecuada. Este conjunto de técnicas recibe el nombre de análisis de items, entendiendo por item cualquier situación que presenta un examinador a un sujeto examinado y que exige una respuesta de éste último.

En líneas generales, las técnicas de análisis de items que aparecen en los textos clásicos de psicometría y estadística (Anastasi, 1968; Guilford, 1965; Magnusson, 1966; Wood, 1965) son tan laboriosas en cálculos que exigen un tiempo extraordinario de trabajo o la utilización de computadoras. Las técnicas que se presentan en este artículo permitirán a cualquier profesor efectuar un análisis de items con suficiente precisión estadística, en pocos minutos y sin necesidad de utilizar sistemas complejos.

El método rápido de análisis de items se aplica, por supuesto, solamente a las pruebas objetivas. Sin embargo, no hay que identificar las pruebas objetivas con las pruebas de selección múltiple o de apareamiento. Una prueba por temas bien elaborada, con una

clave de calificación, es también una prueba objetiva. Quedan únicamente excluidas de las pruebas objetivas, aquellas en que no existe una justificación lógica de la forma en que debe corregirse la respuesta.

El artículo estará dividido en tres partes. En la primera se identificarán los aspectos de un ítem que deben ser analizados. En la segunda se presentarán las técnicas rápidas para analizar cada aspecto. En la última parte aparecerá un resumen de los pasos a efectuar en el análisis de ítems, de forma que el interesado no tenga que estar recurriendo a la totalidad del texto en el momento de la aplicación. El apéndice está dirigido a personas interesadas en la estadística, pero su lectura no es indispensable para la comprensión del texto.

Aspectos de análisis

El primer tópico a considerar se refiere a cuáles aspectos de los ítems deben ser analizados cuando concluye una prueba. Estos son la dificultad y la validez de los ítems.

1.— Dificultad.— El análisis de la dificultad implica simplemente obtener una información sobre la cantidad de sujetos que no supieron contestar correctamente un ítem. Un ítem difícil es aquel que muy pocos han podido contestar correctamente y un ítem fácil es aquel en que muchos han podido dar una respuesta acertada.

La importancia de analizar la dificultad de los ítems se debe a que, de esa manera, puede conocerse su poder diferenciador, es decir, la facultad que tienen los ítems de distinguir entre los sujetos que conocen su respuesta y los que no la conocen. Para aclarar esta noción, en el cuadro I se presenta el ejemplo de un ítem contestado por cuatro sujetos, A, B, C y D. En estas circunstancias existen cinco eventualidades: que lo contesten incorrectamente los cuatro, sólo tres, sólo dos, sólo uno o ninguno. Estas eventualidades aparecen indicadas en el cuadro, con un ejemplo concreto para cada eventualidad; la letra "i" indica una respuesta incorrecta y la letra "c" una respuesta correcta. En la primera y última eventualidad, no es posible diferenciar ningún par de sujetos pues todos emitieron un mismo tipo de respuesta, incorrecta en el primer caso y correcta en el último caso. En la segunda eventualidad pueden observarse tres pares de sujetos que contestaron

diferentemente: AD, BC y CD; también en la cuarta eventualidad aparecen tres pares diferentes: AB, AC y AD. Por último, en el tercer caso pueden distinguirse cuatro pares de sujetos con respuesta diferentes: AC, AD, BC y BD.

Correctas	Incorrectas	Sujetos				Pares Diferentes
		A	B	C	D	
0	4	i	i	i	i	
1	3	i	i	i	i	AD BD CD
2	2	i	i	i	i	AC AD BC BD
3	1	i	c	c	c	AB AC AD
4	0	c	c	c	c	

Cuadro I. Pares de sujetos con respuestas diferentes como consecuencia de las diferentes formas en que cuatro sujetos pueden contestar un ítem.

Generalizando, si el ítem ha sido contestado por todos los sujetos o por ningún sujeto, su poder diferenciador es nulo, mientras que si ha sido contestado por la mitad de los sujetos, su poder diferenciador es máximo. Cuando se analice la dificultad de los ítems, se seleccionarán aquellos que han sido contestados incorrectamente por la mitad de los sujetos aproximadamente y se descartarán los que han sido contestados incorrectamente por casi todos los sujetos o por casi ninguno. En el cuadro II aparece indicado el poder diferenciador de los ítems, en función del número de respuestas incorrectas y para grupos de sujetos de tamaño diferente.

2.— *Validez*. El análisis de la validez de un ítem tiene por objetivo investigar si el ítem mide lo que pretende medir. Existen dos tipos de análisis de validez: externa e interna. En el primer caso se trata de conocer si el ítem mide lo que *pretende medir* la prueba, y en el segundo, si el ítem mide lo que *mide* la prueba.

Para aclarar la diferencia entre los dos tipos de validez de un ítem, se presentará un ejemplo a continuación. En el campo de las aptitudes verbales humanas, se encuentran dos que se miden frecuentemente, la comprensión verbal y la fluidez verbal. La comprensión verbal es la capacidad de un individuo para entender el

N	Número de respuestas incorrectas																				
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
8	0	7	12	15	16	15	12	7	8												
12	0	11	20	27	32	35	36	35	32	27	20	11	0								
16	0	15	28	39	48	55	60	63	64	63	60	55	48	39	28	15	0				
20	0	19	36	51	64	75	84	91	96	99	100	99	96	91	84	75	64	51	36	19	0

Cuadro II. Poder diferenciador de un ítem (Pares de sujetos diferentes) en función del número de respuestas incorrectas y del número de sujetos que contestaron el ítem (n).

significado de palabras; los items de sinónimos miden adecuadamente esta aptitud. La fluidez verbal es la capacidad de un individuo para elaborar palabras rápidamente; como ejemplos de items de fluidez verbal se encuentran las preguntas en que se solicita al sujeto escribir un cierto número de palabras que comiencen con una letra y terminen con otra. Un evaluador desea confeccionar una prueba de comprensión verbal. Si utiliza 50 items, 45 de comprensión verbal y 5 de fluidez verbal, los items de comprensión verbal tendrán buena validez externa e interna, pues miden lo que pretende medir la prueba y también miden lo que realmente está midiendo casi toda la prueba; los items de fluidez verbal tendrán una validez externa e interna inadecuadas. Por el contrario, si utilizan 45 items de fluidez verbal y 5 items de comprensión verbal, la validez externa de los items de fluidez verbal será inadecuada, pues la prueba pretende medir la comprensión verbal, pero la validez interna será adecuada, ya que casi toda la prueba está realmente midiendo la fluidez verbal; por el contrario, los cinco items de comprensión verbal tendrán buena validez externa y mala validez interna.

Para medir los dos tipos de validez se utilizan dos tipos de criterio: externo e interno. Un ejemplo de criterio externo puede ser la constitución de dos grupos en que el aspecto que se pretende medir aparezca en dos intensidades diferentes. En el ejemplo precedente se podría solicitar a profesores de literatura de muchos cursos diferentes que identifiquen estudiantes con un nivel alto (grupo A) y con un nivel bajo (grupo B) de comprensión verbal, después de haberles explicado con claridad lo que esta variable significa. A continuación se aplica la prueba a los dos grupos de estudiantes que constituyen el criterio. Aquellos items en que la cantidad de respuestas erróneas sea significativamente más pequeña en el grupo A que en el grupo B, tendrán buena validez externa; pero si la cantidad de respuestas incorrectas no es significativamente diferente en ambos grupos, o si el grupo A emite más respuestas incorrectas que el grupo B, entonces se considerará que el item tiene una validez externa inadecuada.

El criterio interno que se utiliza para analizar la validez interna es siempre el puntaje total en la prueba. Tomando un ejemplo análogo al anterior, los sujetos se pueden dividir en dos grupos, los que obtuvieron un puntaje más alto (grupo A) y los que obtuvieron un puntaje más bajo (grupo B). También, para que la

validez interna sea conveniente, el número de respuestas erróneas debe ser significativamente inferior en el grupo A con relación al grupo B; en el caso contrario, el ítem posee mala validez interna.

En las pruebas de conocimientos es prácticamente imposible medir la validez externa, ya que no existe forma de identificar con facilidad un criterio externo. En estas pruebas se acostumbra a analizar solamente la validez interna, elaborando el criterio interno en función de la calificación total en la prueba.

De todas formas, el problema de la validez externa queda obviado, porque en realidad los que obtienen mejor calificación son generalmente los que más saben y los que obtienen peor calificación son casi siempre los que saben menos, excepto en circunstancias extremadamente raras que en la práctica no se presentan.

En el cuadro III aparecen dos ejemplos de ítems de validez interna diferente. En ambos ítems se utilizaron 40 sujetos. En las líneas, estos sujetos se encuentran divididos en función de la calificación total de la prueba en dos grupos: los veinte de puntaje alto (A) y los veinte de puntaje bajo (B). En las columnas, los sujetos aparecen divididos en función del tipo de respuesta que emitieron: los que contestaron correctamente (C) y los que contestaron incorrectamente (I) al ítem. Así, en el ejemplo 1, de los veinte sujetos del grupo alto, 9 contestaron mal, y de los veinte sujetos del grupo bajo, 10 contestaron incorrectamente; en este caso, la validez interna del ítem es pobre. Por el contrario, en el ejemplo 2, la validez interna del ítem es elevada, ya que de los sujetos del grupo alto sólo uno contestó mal al ítem, mientras que de los sujetos del

	I	C			I	C	
A	9	11	20	A	1	19	20
B	10	10	20	B	17	3	20
	19	21			18	22	
	ítem 1				ítem 2		

Cuadro III. Ejemplos de dos ítems de validez interna diferente.

grupo bajo, 17 contestaron incorrectamente.

Técnicas de análisis

Antes de introducir las técnicas rápidas de análisis de items, debe advertirse que éstas son sólo aplicables a grupos de 10, 12, 14, 16, 18, 20, 24, 28, 32, 36, 40, y 50 sujetos. En su situación personal, el profesor deberá descartar al azar el excedente de sujetos. Si un profesor aplica una prueba a 34 estudiantes, deberá eliminar aleatoriamente 2, para conservar 32 sujetos; igualmente, si a un examen se presentan 47 estudiantes, el profesor deberá descartar 7 para el análisis de items y quedarse exactamente con 40 sujetos.

Para efectuar el análisis de items, el primer paso consiste en elaborar una matriz sujetos-items, es decir un cuadro en que aparecen los items en las líneas, los sujetos en las columnas, y las respuestas correctas (c) o incorrectas (i) en las casillas. En el cuadro IV aparece como ejemplo una matriz de una prueba de 25 items aplicados a 20 sujetos; los sujetos se encuentran indicados por las letras A a T. Como condición suplementaria, es indispensable que los sujetos se encuentren ordenados en función del puntaje total obtenido en la prueba, situando en la primera columna el que obtuvo menos puntos, y en la última columna el que obtuvo más. En el ejemplo del cuadro IV el sujeto E contestó bien a los items 1, 6 y 14, y contestó mal a los items 4, 12 y 25; igualmente, el sujeto N respondió correctamente a los items 3, 19 y 23, e incorrectamente a los items 5, 16 y 20.

1.— *Dificultad.* Para analizar la dificultad de los items, sólo es necesario contar el número de sujetos que respondieron erróneamente a cada item. Este dato aparece en la columna f_i (frecuencia de incorrectas) del cuadro; por ejemplo, los items 8, 11 y 22 obtuvieron respectivamente 7, 3 y 19 respuestas incorrectas.

A continuación se entra en la tabla que aparece en el cuadro V (cfr. Apéndice, punto A) en la columna correspondiente al número de sujetos utilizados (10, 12, 14, 16, 18, 20, 24, 28, 32, 36, 40, 50). Se considerarán con buena dificultad aquellos items cuya f_i cae entre los valores extremos máximo y mínimo (incluyendo estos valores) que aparecen en la tabla, en las líneas. Así, si se utilizan 32 sujetos, se aceptarán todos los items cuya f_i se encuentren entre 26 y 6, y se descartarán los items que hayan alcanzado una f_i de 27 ó más, o de 5 ó menos.

Iten	Sujetos																f _i	b _i	Dd	Dv				
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P					Q	R	S	T
1	i	c	i	c	c	c	c	c	c	c	c	c	i	c	c	c	c	c	c	c	3		R	
2	i	c	c	c	c	c	c	c	c	c	c	c	c	c	c	c	c	c	c	c	1		R	
3	i	c	i	c	c	c	c	c	c	c	c	c	c	c	c	c	c	i	c	c	3		R	
4	i	c	i	i	i	i	c	i	c	c	c	c	c	c	c	c	c	c	c	c	6	6	A	A
5	i	i	i	i	i	c	c	i	c	c	(i)	c	c	i	i	c	i	i	i	c	12	6	A	R
6	c	i	c	c	c	c	c	c	(i)	c	c	c	c	c	c	c	c	c	c	c	2		A	
7	c	c	i	i	i	i	i	i	c	c	c	c	c	i	c	c	c	c	c	i	8	6	A	A
8	i	c	c	i	i	i	c	i	c	(i)	c	c	c	c	c	i	c	c	c	c	7	5	A	R
9	c	c	i	c	c	c	c	c	c	c	c	i	c	i	c	c	c	c	c	c	3		A	
10	i	i	c	c	c	c	c	i	(i)	c	c	i	c	c	c	c	c	c	c	c	7	5	A	R
11	i	i	c	c	c	c	c	c	c	c	c	i	c	c	c	c	c	c	c	c	3		A	
12	i	i	i	c	i	i	c	c	c	c	c	c	i	c	c	c	c	i	c	c	7	5	A	R
13	i	i	c	c	i	c	i	i	c	c	c	c	c	c	c	c	c	c	i	c			A	A
14	i	i	i	i	c	c	c	c	i	i	i	i	c	c	c	c	c	c	c	c	10	7	A	A
15	c	c	c	i	c	i	i	e	(i)	c	c	c	c	c	c	c	c	c	c	c	4	3	A	A
16	i	i	i	i	i	c	c	c	i	c	i	i	i	c	i	c	c	c	c	c	11	6	A	R
17	i	c	i	i	c	i	i	c	c	i	i	i	i	i	i	i	i	c	i	c	14	6	A	R
18	i	i	i	i	i	i	i	i	(i)	(i)	c	c	i	i	i	i	i	i	c	i	17		A	
19	i	i	i	i	i	i	c	i	i	c	i	i	c	i	c	c	c	c	c	c	13	9	A	R
20	c	i	c	c	i	c	i	i	(i)	(i)	c	c	c	i	i	c	i	c	c	c	9	4	A	R
21	c	i	c	c	i	c	i	c	c	i	c	c	c	c	c	c	c	c	c	c	5	4	A	R
22	i	i	i	i	i	c	i	i	i	i	i	i	i	i	i	i	i	i	i	i	19		A	
23	c	i	c	i	i	c	i	i	c	c	c	c	c	c	i	c	c	c	c	c	6	5	A	A
24	i	i	i	i	c	i	i	c	i	i	i	i	i	i	c	i	i	c	i	c	15	8	A	R
25	i	i	i	i	i	i	i	c	c	i	i	i	i	i	i	i	i	i	i	i	18		A	
	18	16	16	14	13	12	12	12	10	10	10	10	9	9	8	8	7	6	5	4				

Cuadro IV. Ejemplo de análisis de ítems de una prueba de 25 preguntas contestadas por 20 sujetos.

En el ejemplo del cuadro IV se utilizaron 20 sujetos; en consecuencia los valores extremos son 16 y 4. En estas circunstancias serán aceptados los items 4, 5, 7, 8, 10, 12, 13, 14, 15, 16, 17, 19, 20, 21, 23, y 24, y serán descartados los restantes, pues la f_i se encuentra o por encima de 16, o por debajo de 4. Los items aceptados y rechazados han sido indicados por las letras A y R en la columna Dd (decisión sobre la dificultad dd los items).

Valores extremos	n											
	10	12	14	16	18	20	24	28	32	36	40	50
Máximo	8	10	11	13	15	16	20	23	26	30	33	42
Mínimo	2	2	3	3	3	4	4	5	6	6	7	8

Cuadro V. Valores extremos de f_i en función de n.

2.— *Validez.* El procedimiento para el análisis de la validez de los items es un poco más laborioso. La técnica que se presenta en los párrafos siguientes se fundamenta en el análisis estadístico de la correlación tetracórica.

El primer paso consiste en dividir el grupo de sujetos en dos partes exactamente iguales. En el cuadro IV, esta división se ha operado trazando una doble línea vertical entre los sujetos J y K; así, de cada lado de la doble línea se encuentra la misma cantidad de sujetos. Los sujetos A a J constituirán el grupo bajo (B), mientras que los sujetos de K a T integrarán el grupo alto (A).

En segundo lugar deben contarse todas las respuestas incorrectas (i) de cada sujeto. Este resultado aparece en la línea P_i (puntaje de incorrectas). Puede observarse en el cuadro IV que los sujetos E, J y R obtuvieron respectivamente 13, 10 y 6 respuestas negativas. Si de lado y lado de la doble línea se encuentran sujetos con la misma cantidad de respuestas incorrectas, delimítese esta zona. En el ejemplo, los sujetos I, J, K y L obtuvieron todos diez respuestas negativas; es por esta razón por lo que se trazaron dos líneas verticales, la primera a la izquierda del sujeto I y la segunda a la

derecha del sujeto L. La zona delimitada por estas dos líneas será denominada zona mediana. Esta zona mediana está compuesta de dos partes: la que se encuentra a la izquierda de la doble línea o zona mediana baja y la que se encuentra a la derecha de la doble línea o zona mediana alta. En el ejemplo, la zona mediana baja se encuentra compuesta por los sujetos I y J, y la zona mediana alta por los sujetos K y L. Para cada ítem, póngase entre paréntesis todas las respuestas incorrectas (i) de la zona mediana baja que puedan ser sustituidos por respuestas correctas (c) de la zona mediana alta (2); por ejemplo, en el ítem 5 aparece una "i" en la zona mediana baja, que ha sido puesto entre paréntesis, ya que puede ser sustituida por una de las dos "c" de la zona mediana alta; también en el ítem 18, las dos "i" de la zona mediana baja han sido colocadas entre paréntesis, pues pueden ser intercambiadas por las dos "c" de la zona mediana alta del mismo ítem; sin embargo, en el ítem 10, aunque aparecen dos "i" en la zona mediana baja, sólo una ha podido ponerse entre paréntesis, ya que en la zona mediana alta aparece una sola "c".

En tercer lugar deberán contarse, *sólo para los ítems que tengan buena dificultad* (pues los otros ya han sido rechazados) todas las "i" del grupo bajo (es decir, de los sujetos que se encuentran a la izquierda de la doble línea) *excluyendo aquellos que se encuentren entre paréntesis*. Estos resultados han sido registrados en la columna fbi (frecuencia de incorrectas del grupo bajo) en el cuadro IV. Los ítems 7, 20 y 23 alcanzaron una fbi de 6, 4 y 5.

Por último deberá utilizarse la tabla que aparece en el cuadro VI (cfr. Apéndice, punto B). Esta tabla presenta, en las columnas, el número total de sujetos examinados (n), y en las líneas, los valores f_i . Para cada ítem, deberá entrarse en la tabla en la columna correspondiente al número de sujetos examinados y descender hasta encontrar la línea correspondiente al valor f_i del ítem analizado; el número encontrado en este punto será denominado valor significativo de fbi. En este caso, si el valor fbi del ítem es igual o superior al valor significativo de fbi dado por la tabla, entonces el ítem tendrá buena validez. En el ítem 5 del ejemplo anterior, se entra en la tabla por la columna 20 y se baja hasta encontrar la línea correspondiente a una f_i igual a 12; en este caso el valor significativo de fbi es de 8, y como el ítem 5 presenta una fbi igual a 6, se considerará que este ítem tiene mala validez. Por el contrario, en el caso del ítem 19, se descende en la columna 20 hasta encontrar la

f ₁	n												
	10	12	14	16	18	20	24	28	32	36	40	50	
42													23
41													23
40													23
39													22
38													22
37													21
36													21
35													21
34													20
33												19	20
32												18	19
31												18	19
30										17		18	18
29										17		17	18
28										16		17	17
27									15	16		16	16
26									15	15		16	16
25									14	15		15	15
24								13	14	14		15	15
23								13	14	14		14	14
22								13	13	13		14	14
21							12	12	13	13		13	13
20							12	12	12	13		13	13
19							11	12	12	12		12	12
18							11	11	11	12		12	12
17						10	10	11	11	11		11	11
16					9	10	10	10	10	10		11	11
15					9	9	10	10	10	10		10	10
14				8	9	9	9	9	9	9		9	9
13				8	8	8	8	9	9	9		9	9
12			7	8	8	8	8	8	8	8		8	8
11		6	7	7	7	7	8	8	8	8		8	8
10		6	7	7	7	7	7	7	7	7		7	7
9		6	7	7	7	7	7	7	7	7		7	7
8	5	6	6	6	6	6	6	6	6	6		6	6
7	5	6	6	6	6	6	6	6	6	6		6	6
6	5	5	5	5	5	5	5	5	5	5		5	5
5	5	5	5	5	5	5	5	4					
4	4	4	4	4	4	4	4						
3	3	3	3	3	3								
2	2	2											

Cuadro VI. Valores significativos de f_{b1} , en función de f_1 y de n.

En definitiva, deberán aceptarse solamente los items que tengan buena dificultad y buena validez, es decir aquellos que aparezcan con A en las columnas Dd y Dv al mismo tiempo. En el ejemplo presentado, sólo seis items presentan esta condición; estos son los items 4, 7, 13, 14, 19 y 23.

Con relación a los items rechazados, el profesor deberá efectuarse interrogantes de orden didáctico; un item rechazado corresponde generalmente a un aspecto del programa que ha sido desarrollado deficientemente por el profesor. En consecuencia, estos items no deben entrar en el cálculo del puntaje total de la prueba, o deben computarse como correctos para todos los estudiantes.

Resumen

Para efectuar un análisis de items de pruebas de conocimientos, síganse las reglas que aparecen a continuación:

1.— Corregir toda la prueba y ordenar los sujetos en función del puntaje total, de menor a mayor.

2.— Elaborar una matriz donde aparezcan los sujetos en las columnas, los items en las líneas, y las respuestas de los sujetos a los items en las casillas.

3.— Contar el número de respuestas negativas de cada item e inscribir este resultado en la columna f_i . Utilizar, en la tabla del cuadro V la columna correspondiente al total de sujetos empleados en el análisis; determinar si la f_i de cada item cae dentro (aceptación) o fuera (rechazo) de los valores extremos que indica la tabla. Indicar la aceptación (a) o el rechazo (R) de cada item en la columna Dd.

4.— Contar el número de respuestas negativas de cada sujeto, inscribir este resultado en la línea P_j e indicar el punto medio dibujando una doble línea que divida los sujetos en dos grupos iguales. Si de cada lado del punto medio aparecen sujetos con la misma P_j , dibujar dos líneas que encierren estos sujetos, delimitando así las zonas medianas alta y baja. Para cada item, poner entre paréntesis las "i" de la zona mediana baja que pueden ser intercambiadas por "c" de la zona mediana alta.

5.— Contar, sólo en dos items de buena dificultad, el número de “i” que se encuentran del lado bajo (izquierdo) del punto medio, excluyendo las que se encuentren entre paréntesis; inscribir este resultado en la columna f_{bi} . Utilizar, en la tabla del cuadro VI, la columna correspondiente al total de sujetos empleados (n); para cada item, encontrar la línea de la tabla que corresponda a f_i ; en la intersección de la columna y de la línea se encontrará el valor significativo de f_{bi} ; determinar si la f_{bi} del item es igual o superior a (aceptación) o inferior (rechazo) al valor significativo de f_{bi} que aparece en la tabla. Indicar la aceptación (A) o el rechazo (R) de cada item en la columna Dv.

6.— Calcular de nuevo el puntaje de cada sujeto sin considerar las respuestas a los items rechazados por razones de mala dificultad o de mala validez.

REFERENCIAS

- Anastasi, A. *Psychological Testing*. London: Collier—Macmillan, 1968.
- Guilford, J. P. *Fundamental statistics in psychology and education*. New York: McGraw—Hill, 1965.
- Magnussn, D. *Test Theory*. Reading, Massachusetts: Addison—Wesley, 1966.
- Wood, D. A. *Elaboracion de tests psicológicos. México: Trillas, 1965.*

APENDICE

A. Construcción de la tabla de valores extremos de f_i en función de n .

Para elaborar esta tabla, se presupuso normalidad en la distribución latente o teórica de cada item y se tomó como criterio de valores extremos una desviación típica de cada lado de la medida teórica; estos valores encierran aproximadamente el 68o/o del área central de la distribución.

Como una desviación típica por encima de la medida incluye el 84.13o/o de los casos y una desviación típica por debajo de la media incluye el 15.87o/o de los casos, los valores máximo (Mx) y mínimo (Mn) de la tabla se calcularon mediante las fórmulas:

$$M_x - 0.8413 n \quad (1.a)$$

$$M_n - 0.1587 n \quad (2.a)$$

B. Construcción de la tabla de valores significativos de f_{bi} en función de f_i y de n .

Esta tabla exigió la elaboración de dos tablas previas.

1.- Tabla de r_t en función de P_i y de P_{bi}

La fórmula de la correlación tetracórica es

$$r_t = \cos \left(\frac{180^\circ}{1 + \sqrt{\frac{f_{ac}f_{bi}}{f_{ai}f_{bc}}}} \right) \quad (1.b)$$

donde f_{ac} = número de sujetos que obtuvieron un puntaje alto en la prueba y que contestaron correctamente al ítem

f_{ai} = número de sujetos que obtuvieron un puntaje alto en la prueba y que contestaron incorrectamente al ítem

f_{bc} = número de sujetos que obtuvieron un puntaje bajo en la prueba y que contestaron correctamente al ítem

f_{bi} = número de sujetos que obtuvieron un puntaje bajo en la prueba y que contestaron incorrectamente al ítem (3)

Expresándola en proporciones, la fórmula (1.b) se presenta

$$r_t = \cos \left(\frac{180^\circ}{1 + \sqrt{\frac{P_{ac}P_{bi}}{P_{ai}P_{bc}}}} \right) \quad (2.b)$$

cuyas proporciones pueden presentarse en una matriz 2 x 2 de la siguiente forma:

	I	C	
A	P_{ai}	P_{ac}	P_a
B	P_{bi}	P_{bc}	P_b
	P_i	P_c	

En toda matriz 2 x 2, si se conocen los cuatro subtotales y un elemento de la matriz, esta puede reconstruirse por completo. Pero como por definición, el grupo alto y el grupo bajo contienen la misma cantidad de sujetos, $P_a = P_b = 0.5$ y como también $P_c = 1 - P_i$, entonces la matriz puede quedar enteramente determinada conociendo solamente P_i y P_{bi} .

Se elaboraron así todas las matrices posibles haciendo asumir sucesivamente a P_i todos los valores entre 0.01 y 0.99, y a P_{bi} todos los valores entre 0.01 y 0.99 que permitiera la matriz.

Utilizando la fórmula (2.b) se calculó, para cada matriz, la correlación tetracórica, obteniendo así la tabla de r_t en función de P_i y de P_{bi} (tabla A).

2.- Tabla de r_t' en función de p_i y de n .

Para conocer si una correlación tetracórica es significativa, se puede utilizar el error típico de esta correlación, cuya fórmula es

$$s_{r_t} = \frac{\sqrt{P_a P_b} \sqrt{P_c P_i}}{y_{ab} y_{ci} \sqrt{n}} \quad (3.b)$$

donde P_a , P_b , P_c y P_i tienen el mismo significado que en la fórmula (2.b)

y_{ab} — ordenada de la curva normal que divide P_a y P_b

y_{ci} — ordenada de la curva normal que divide P_c y P_i

n — total de sujetos.

Como $P_a = P_b = 0.5$

$P_c = 1 - P_i$

y_{ab} — se encuentra especificada por P_a

y_{ci} — se encuentra especificada por P_c

entonces el error típico de la correlación tetracórica está completamente determinado a partir de P_i y de n .

Si se parte de la hipótesis de una correlación significativa y positiva (test a una cola) y si se acepta una probabilidad de error de 0.05, entonces los valores significativos de la correlación tetracórida (r_t') pueden calcularse según la fórmula

$$r_t' = 1.645 s_{r_t}$$

Así, haciendo asumir sucesivamente a P_i todos los valores posibles entre 0.01 y 0.99, a n los valores 10, 12, 14, ... 50, y multiplicando todas las s_{r_t} resultantes por 1.645, se obtuvo la tabla de r_t' en función de P_i y de n (tabla B).

3.— *Elaboración final.*

Se tomaron todos los valores r_t , de la tabla B y se localizaron, en la tabla A los valores P_{bi} correspondientes. De esta forma se identificaron los valores P_{bi} que determinarían una correlación tetracórida significativa y positiva con una probabilidad de error de 0.05, conociendo P_i y n .

Se transformaron las P_i y las P_{bi} en f_i y f_{bi} respectivamente, multiplicándolos por n , obteniendo así la tabla definitiva.

N O T A S

1

(1) El término evaluación se encuentra aquí restringido a la obtención de información sobre el aprendizaje efectuado a largo plazo; este tipo de información es la que se intenta medir en las pruebas parciales y finales del sistema de evaluaciones de la UNPHU.

(2) Este procedimiento se utiliza para eliminar la influencia del ítem del puntaje total de la prueba, ya que se ha podido demostrar que esta influencia aumenta espúreamente la correlación tetracórida.

(3) En muchos textos de estadística se acostumbra identificar estas cuatro frecuencias, por las cuatro primeras letras del alfabeto.