LERTAP 5 (PAQUETE DE ANÁLISIS DE PRUEBAS DEL LABORATORIO DE INVESTIGACIONES EDUCATIVAS): SOFTWARE PARA EL ANÁLISIS DE ÍTEMS

Lertap 5 (Laboratory of Educational Research Test Analysis Package): Software for item analysis

Juan Sebastián Argüello Chiquiza/ jsarguelloc@unal.edu.co

Estudiantes de Psicología
Universidad Nacional de Colombia

EN la actualidad, la Teoría clásica de los test (TCT) continúa siendo la base teórica para la estimación de la calidad de la mayoría de las pruebas psicológicas usadas, adaptadas o construidas; ello, a pesar de los avances técnicos que la Teoría de la respuesta al ítem (TRI) ha aportado a las técnicas de la medición y evaluación del comportamiento y los rasgos psicológicos (Attorresi et al., 2009), lo cual resalta la vigencia de la TCT, y asienta a la TRI como una extensión de la misma, la cual no puede ser considerada como un modelo aparte (Herrera, Sánchez y Jiménez, 2001).

Desde la TCT, el análisis de ítems se centra en el análisis de dificultad, discriminación y flujo de opciones; a continuación, se describe cómo realizar cada uno de estos procedimientos en el programa Lertap.

ANÁLISIS DE DIFICULTAD Y DISCRIMINACIÓN

Los análisis de acierto-fallo son los primeros análisis de calidad llevados a cabo a nivel granular en la construcción, adaptación o validación de instrumentos (Hogan, 2004); estos procedimientos se centran en la observación del comportamiento de la opción correcta de un ítem, con respecto a la elección de cualquier otra opción de respuesta.

Por un lado, el índice de dificultad es la proporción de personas que contestan correctamente un ítem, su valor oscila entre 0 y 1; entre más cercano a 1, se interpreta como una mayor facilidad del mismo. Por otro lado, el índice de discriminación es considerado como la capacidad que tiene el ítem para diferenciar a las personas que tienen cierto atributo, rasgo o habilidad, con respecto a las que no lo tienen. Tradicionalmente, este índice ha sido calculado de dos maneras: la primera, la correlación del acierto o fallo de un ítem con la puntuación total del test completo; la segunda, la comparación de la dificultad del ítem en el grupo de personas de nivel superior e inferior.

En el primer método, se utiliza el estadístico de correlación biserial puntual (r_{pbis}); mientras que para el segundo método, se ordenan las personas de la muestra de estandarización por puntaje bruto obtenido, y se selecciona un porcentaje de personas con el mayor puntaje obtenido y otro porcentaje de personas con el menor puntaje obtenido (grupos extremos), normalmente, se aconseja que cada grupo obedezca al 27 % de las puntuaciones más bajas o las más altas; ambos métodos conllevan a valores de entre 0 y 1, los cuales se interpretan directamente (mayor o más cercano a uno significa mayor discriminación del ítem).

ANÁLISIS DE FLUJO DE OPCIONES

Los análisis de distractores tienden a ser análisis de segundo orden, puesto que son análisis aplicables en situaciones particulares.

i. Un distractor es cualquier opción de respuesta que no sea considerada una respuesta correcta; por lo tanto, la primera situación es que el ítem a evaluar tenga más de dos opciones de respuesta (Al tener solo dos opciones de respuesta, solo se puede hacer análisis de acierto y fallo).

ii. Al analizar los distractores, se intenta hacer una revisión del mal funcionamiento que pueda tener un ítem, por lo cual, estos procedimientos pueden ser utilizados en situaciones donde se quiere mejorar la calidad de los ítems, más que simplemente medir las propiedades psicométricas del reactivo (aquí llamadas análisis de acierto-fallo); por ejemplo, en los pilotajes o aplicación de versiones previas, realizados para la construcción de pruebas.

iii. Es una herramienta para saber cómo están respondiendo diferencialmente los grupos personas de mayor y menor puntaje. Esto sirve en las situaciones en donde queremos saber cómo se está llevando a cabo el proceso de respuesta diferencialmente entre grupos (Hogan, 2004).

De este modo, es importante recalcar que este tipo de análisis son comparativos del comportamiento de las opciones de respuesta distractoras, con respecto a la opción de respuesta correcta.

Según lo anterior, la dificultad de un distractor es la proporción de personas que responden a este sobre el total de la muestra de estandarización. De manera contraria a la estimación de dificultad de la clave, entre mayor sea este índice, más difícil es considerado un distractor; eso obedece a que es más difícil de descartar o más atractivo para las personas que no tienen la cantidad de atributo suficiente para llegar al respuesta correcta de un ítem.

Por su parte, los análisis de discriminación por correlación punto biserial, nos indican qué tan relacionada está la elección de cada opción de respuesta herrada con el puntaje total de la prueba de una persona. Por lógica, se espera que ningún distractor correlacione de manera altamente positiva con el puntaje total de la prueba, y mucho menos que este nivel de discriminación sea mejor que el proporcionado por la clave.

En los casos en los que se pueda notar una correlación de un distractor con el puntaje la prueba anormalmente alto, puede llevarse a cabo análisis de discriminación por comparación de grupos, comparar los índices de dificultad de los reactivos del grupo alto, en especial, con los demás grupos. Estos análisis ayudan a identificar cuando un distractor es más atractivo de lo esperado en el grupo alto (esto podría ser evidencia de que este grupo de mayor rasgo puede entender y procesar el ítem de una manera diferente a lo esperado).

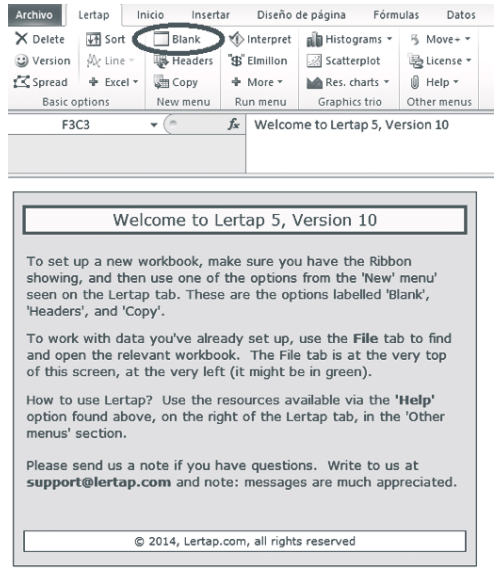
SOFTWARE

Lertap 5 es un Macro para Excel, desarrollado por Larry Richard Nelson, perteneciente a la Universidad de Tecnología de Curtin (institución pública ubicada en Perth, Australia. Diseñado especialmente para el análisis de ítems en quices exámenes, pruebas o encuestas (Curtin University of Technology, 2001). Debido a que el uso del software es bastante sistemático, este será expuesto a continuación, a través de un paso a paso.

Paso 1

Abrir el archivo del Macro y hacer clic en la opción Blank, para abrir una base de datos nueva y editable. Tal como se muestra en la imagen 5-1.

Imagen 5-1. (siguiente página)
Archivo Excel de inicio del macro de Lertap.

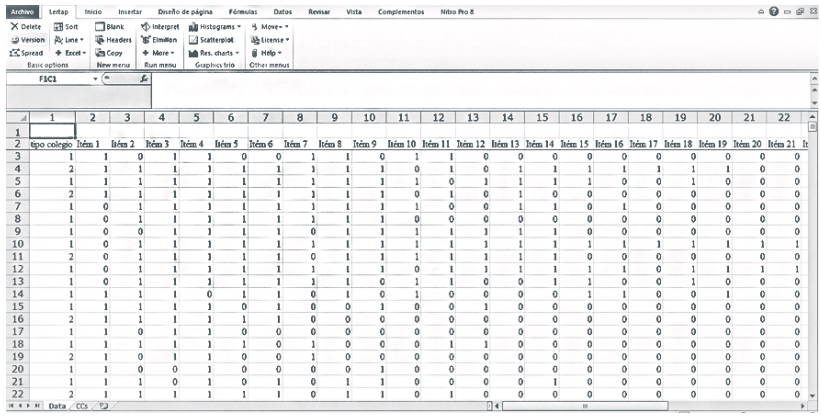


Fuente: Libro de Excel con el programa Letrap en el computador del autor.

Paso 2

Sobre este archivo nuevo, se insertará la base de datos, teniendo cuidado de que: la primera fila de respuestas quede sobre la fila tres (03) del archivo; la primera columna sea algún dato descriptivo (Imagen 5-2), y que no haya ninguna fila o columna vacía dentro de la base de datos.

Imagen 5-2.
Correcto posicionamiento de base de datos en lertap.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1																						
2	tipo colegio	Ítém 1	Ítém 2	Ítém 3	Ítém 4	Ítém 5	Ítém 6	Ítém 7	Ítém 8	Ítém 9	Ítém 10	Ítém 11	Ítém 12	Ítém 13	Ítém 14	Ítém 15	Ítém 16	Ítém 17	Ítém 18	Ítém 19	Ítém 20	Ítém 21
3	1	1	0	1	1	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0
4	2	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	0
5	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	1	0	0	0
6	2	1	1	1	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0
7	1	0	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	0	0	0	0	0
8	1	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
9	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0
10	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	2	0	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0
12	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1
13	1	0	1	1	1	1	1	1	1	0	1	1	0	0	1	1	0	0	1	0	0	0
14	1	1	1	1	0	1	1	0	1	0	1	0	0	0	0	1	1	0	0	1	0	0
15	1	1	1	1	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
16	2	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	1	1	1	1	1	1	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0
19	2	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	1	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
21	1	1	1	0	1	0	1	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0
22	2	1	1	1	1	1	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0

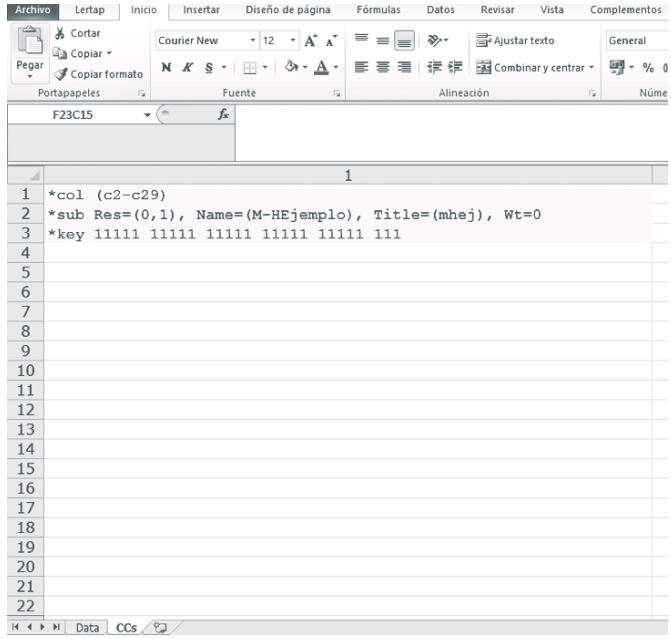
Fuente: Libro de Excel con el programa Letrap en el computador del autor.

Paso 3

Dirigirse a la pestaña "CCs" (Code cards, o tarjetas de código) (imagen 5-3), y digitar el código de análisis, según la ubicación de los ítems, especificando: columna de inicio y fin (Fila 1); el tipo de prueba (Fila 2), y las claves a cada ítem por orden de columna (Fila 3).

En la segunda fila del código, se pueden poner nombres abreviados y extensos de la prueba, así como el peso o la calificación de las opciones incorrectas en la opción "Wt". Aunque este punto puede ser el más complicado, en el archivo de inicio del macro, en la pestaña "Syntax", se explica cada uno de los códigos necesarios para el análisis de ítem.

Imagen 5-3.
Pestaña de códigos de análisis.

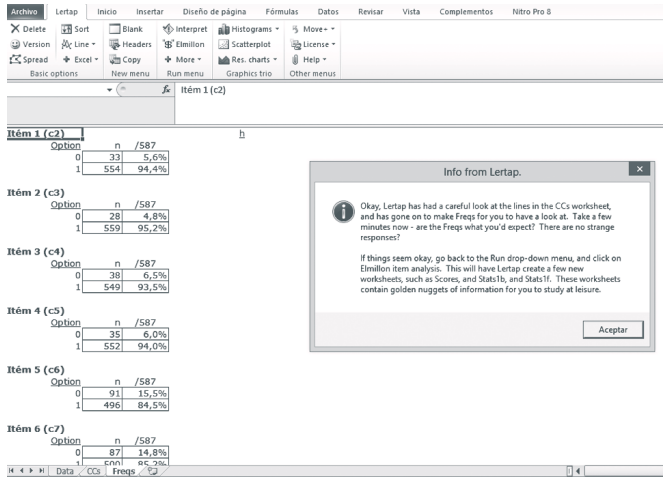


Fuente: Libro de Excel con el programa Letrap en el computador del autor.

Paso 4

Hacer clic en el menú del macro "Interpret", el cual se encuentra al lado de la opción "Blank" (véase imagen 5-1), para que el software calcule las frecuencias de las opciones de respuesta de cada ítem en una nueva pestaña (como se muestra en la imagen 5-4). El software mostrará una ventana de advertencia, en donde pide revisar los datos antes de continuar con el proceso de análisis.

Imagen 5-4.
Pestaña de frecuencias.



Fuente: Libro de Excel con el programa Letrap en el computador del autor.

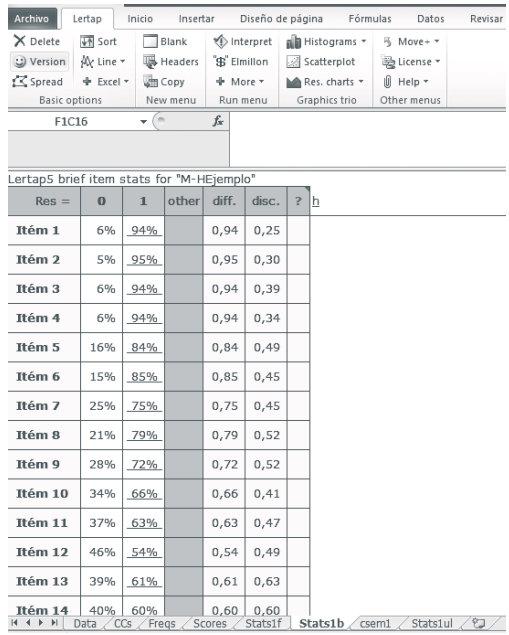
Paso 5

Si los datos están bien tabulados y son correctos en la pestaña "Freqs", hacer clic en el menú del macro "Elmillon", con el cual se abrirán una serie de pestañas (más adelante explicadas), con todo el análisis de Ítems; antes de eso, dependiendo del tamaño de la muestra de la aplicación, preguntará en cuantos grupos de igual tamaño se dividirá la muestra, con el fin de hacer los análisis de comparación de grupos, en caso de tener una gran muestra, el programa automáticamente divide la muestra en 5 grupos.

Pestañas

La primeras dos pestañas que aparece son "Stats1b" y "Stasts1f". Ambas contienen los estadísticos de cada ítem, con la única diferencia de que la primera es un resumen de los resultados en general (Brief) (imagen 5-5) y, en la segunda, hay un compilado de los estadísticos analizados en los ítems (Full), ello incluye media puntuación Z, correlación biserial y número de personas que eligieron cada opción de respuesta; además de la dificultad (p) y la discriminación ítem prueba (pb(r)) (imagen 5-6), así como los estadísticos descriptivos de la calificación del test total (imagen 5-7) y el Coeficiente de alfa de Cronbach total y depurado, si se elimina ítems (imagen 5-8) .

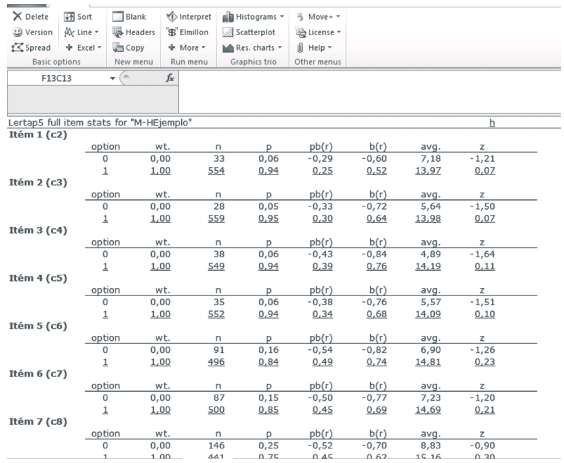
Imagen 5-5.
Pestaña de estadísticos de los ítems abreviada.



Res =	0	1	other	diff.	disc.	?	h
Ítem 1	6%	94%		0,94	0,25		
Ítem 2	5%	95%		0,95	0,30		
Ítem 3	6%	94%		0,94	0,39		
Ítem 4	6%	94%		0,94	0,34		
Ítem 5	16%	84%		0,84	0,49		
Ítem 6	15%	85%		0,85	0,45		
Ítem 7	25%	75%		0,75	0,45		
Ítem 8	21%	79%		0,79	0,52		
Ítem 9	28%	72%		0,72	0,52		
Ítem 10	34%	66%		0,66	0,41		
Ítem 11	37%	63%		0,63	0,47		
Ítem 12	46%	54%		0,54	0,49		
Ítem 13	39%	61%		0,61	0,63		
Ítem 14	40%	60%		0,60	0,60		

Fuente: Libro de Excel con el programa Letrap en el computador del autor.

Imagen 5-6.
Pestaña de estadísticos completos de los ítems, parte superior.



Ítem 1 (c2)

option	wt.	n	p	pb(r)	b(r)	avg.	z
0	0,00	33	0,06	-0,29	-0,60	7,18	-1,21
1	1,00	554	0,94	0,25	0,52	13,97	0,07

Ítem 2 (c3)

option	wt.	n	p	pb(r)	b(r)	avg.	z
0	0,00	28	0,05	-0,33	-0,72	5,64	-1,50
1	1,00	559	0,95	0,30	0,64	13,98	0,07

Ítem 3 (c4)

option	wt.	n	p	pb(r)	b(r)	avg.	z
0	0,00	38	0,06	-0,43	-0,84	4,89	-1,64
1	1,00	549	0,94	0,39	0,76	14,19	0,11

Ítem 4 (c5)

option	wt.	n	p	pb(r)	b(r)	avg.	z
0	0,00	35	0,06	-0,38	-0,76	5,57	-1,51
1	1,00	552	0,94	0,34	0,68	14,09	0,10

Ítem 5 (c6)

option	wt.	n	p	pb(r)	b(r)	avg.	z
0	0,00	91	0,16	-0,54	-0,82	6,90	-1,26
1	1,00	496	0,84	0,49	0,74	14,81	0,23

Ítem 6 (c7)

option	wt.	n	p	pb(r)	b(r)	avg.	z
0	0,00	87	0,15	-0,50	-0,77	7,23	-1,20
1	1,00	500	0,85	0,45	0,69	14,69	0,21

Ítem 7 (c8)

option	wt.	n	p	pb(r)	b(r)	avg.	z
0	0,00	146	0,25	-0,52	-0,70	8,83	-0,90
1	1,00	441	0,75	0,45	0,62	15,16	0,30

Fuente: Libro de Excel con el programa Letrap en el computador del autor.

Imagen 5-7.
Pestaña de estadísticos completos de los ítems, parte media..

Lertap5 full item stats for "M-HEjemplo"

Summary statistics

number of scores (n):	587	
lowest score found:	0,00	(0,0%)
highest score found:	28,00	(100,0%)
median:	14,00	(50,0%)
mean (or average):	13,58	(48,5%)
standard deviation:	5,31	(19,0%)
standard deviation (as a sample):	5,31	(19,0%)
variance (sample):	28,23	

number of subtest items:	28	
minimum possible score:	0,00	
maximum possible score:	28,00	
reliability (coefficient alpha):	0,89	
index of reliability:	0,94	
standard error of measurement:	1,80	(6,4%)

item difficulty bands

,00: Ítem 26 Ítem 27 Ítem 28
,10: Ítem 21 Ítem 22 Ítem 23 Ítem 24 Ítem 25
,20: Ítem 19
,30: Ítem 17 Ítem 18 Ítem 20
,40: Ítem 16
,50: Ítem 12 Ítem 15
,60: Ítem 10 Ítem 11 Ítem 13 Ítem 14
,70: Ítem 7 Ítem 8 Ítem 9
,80: Ítem 5 Ítem 6
,90: Ítem 1 Ítem 2 Ítem 3 Ítem 4

item discrimination bands

,00:
,10: Ítem 28

Data / CCs / Freqs / Scores / **Stats1f** / Stats1b / csem1 / Stats1ul

Fuente: Libro de Excel con el programa Letrap en el computador del autor.

Imagen 5-8.
Pestaña de estadísticos completos de los ítems, parte Baja.

Lertap5 full item stats for "M-HEjemplo"

item discrimination bands

,00:
,10: Ítem 28
,20: Ítem 1 Ítem 26 Ítem 27
,30: Ítem 2 Ítem 4 Ítem 21 Ítem 22 Ítem 23 Ítem 24 Ítem 25
,40: Ítem 3 Ítem 5 Ítem 6 Ítem 7 Ítem 10 Ítem 11 Ítem 12 Ítem 17
,50: Ítem 8 Ítem 9 Ítem 16 Ítem 18 Ítem 19 Ítem 20
,60: Ítem 13 Ítem 14 Ítem 15
,70:
,80:
,90:

alpha figures (alpha = .8856)

without	*alpha*	*change*
Ítem 1	0,885	0,000
Ítem 2	0,885	-0,001
Ítem 3	0,883	-0,002
Ítem 4	0,884	-0,002
Ítem 5	0,881	-0,005
Ítem 6	0,882	-0,004
Ítem 7	0,882	-0,004
Ítem 8	0,880	-0,006
Ítem 9	0,880	-0,006
Ítem 10	0,883	-0,003
Ítem 11	0,881	-0,004
Ítem 12	0,881	-0,005
Ítem 13	0,877	-0,009
Ítem 14	0,877	-0,008
Ítem 15	0,877	-0,009
Ítem 16	0,880	-0,005
Ítem 17	0,881	-0,005
Ítem 18	0,878	-0,007
Ítem 19	0,880	-0,005
Ítem 20	0,879	-0,007
Ítem 21	0,883	-0,002

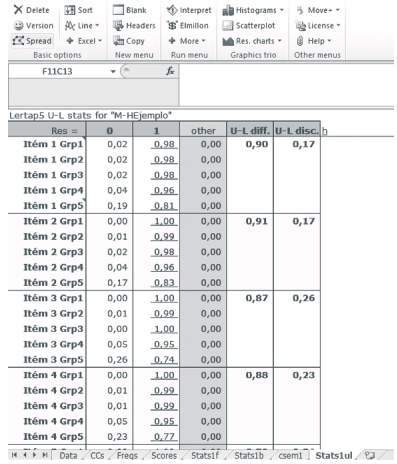
Data / CCs / Freqs / Scores / **Stats1f** / Stats1b / csem1 / Stats1ul

Fuente: Libro de Excel con el programa Letrap en el computador del autor.

La pestaña “Csem1” contiene algunas estimaciones del Error Estándar de Medida, por método clásico de la TCT y algunos métodos con correcciones. Sin embargo, acá no se analizará dado que estos resultados sobrepasan los fines de esta didáctica.

La ultima pestaña, “Stats1ul”, muestra los estadísticos para la comparación de grupos (Upper-Lower groups) (imagen 5-9); estos son porcentaje de elección de cada opción de respuesta (o dificultad de cada una de ellas), comparación de dificultades del grupo alto y bajo (índice de discriminación); por último, la comparación de discriminación en el grupo alto y bajo. A través de cometarios, el software indica cuál debe ser interpretado como el grupo de alto puntaje y el de bajo puntaje. Al final de esta pestaña, aparecen estadísticos descriptivos de las puntuaciones de cada uno de los grupos.

Imagen 5-9.
Pestaña de estadísticos de comparación de grupos altos y bajos.



Res =	0	1	other	U-L diff.	U-L disc.
Itém 1 Grp1	0,02	0,98	0,00	0,90	0,17
Itém 1 Grp2	0,02	0,98	0,00		
Itém 1 Grp3	0,02	0,98	0,00		
Itém 1 Grp4	0,04	0,96	0,00		
Itém 1 Grp5	0,19	0,81	0,00		
Itém 2 Grp1	0,00	1,00	0,00	0,91	0,17
Itém 2 Grp2	0,01	0,99	0,00		
Itém 2 Grp3	0,02	0,98	0,00		
Itém 2 Grp4	0,04	0,96	0,00		
Itém 2 Grp5	0,17	0,83	0,00		
Itém 3 Grp1	0,00	1,00	0,00	0,87	0,26
Itém 3 Grp2	0,01	0,99	0,00		
Itém 3 Grp3	0,00	1,00	0,00		
Itém 3 Grp4	0,05	0,95	0,00		
Itém 3 Grp5	0,26	0,74	0,00		
Itém 4 Grp1	0,00	1,00	0,00	0,88	0,23
Itém 4 Grp2	0,01	0,99	0,00		
Itém 4 Grp3	0,01	0,99	0,00		
Itém 4 Grp4	0,05	0,95	0,00		
Itém 4 Grp5	0,23	0,77	0,00		

Fuente: Libro de Excel con el programa Letrap en el computador del autor.

COMENTARIOS FINALES

Lertap es una herramienta bastante útil, que ayuda a realizar, de forma rápida, una gran cantidad de análisis, presentando resultados de manera organizada y sencilla. El procedimiento aquí mostrado es un análisis de ítems básico; empero, el software no se limita solo a estos, puesto que incluye múltiples opciones para la graficación de los datos y resultados,

análisis de puntuaciones por grupos, así como análisis de DIF. Aunque aquí se ahondó en el análisis de ítems de ejecución máxima, el software también realiza análisis de ítems de ejecución típica Likert.

Como toda herramienta informática, es una forma de ahorrar tiempo y esfuerzo, haciendo cálculos estadísticos; no obstante, los criterios de calidad de los ítems, como puntos mínimos en los índices de dificultad y discriminación, o criterios de análisis, como número de particiones de la muestra, son responsabilidad y decisión del investigador, aun cuando el software pueda hacer comentarios sobre resultados no esperados.

A pesar de que en esta didáctica, se intentó mostrar una pequeña introducción sobre el análisis de ítems, se recomienda al investigador y usuario del software realizar una muy consciente revisión teórica del tema, para poder aprovechar esta herramienta, tomando las decisiones correctas para el análisis de cada prueba psicológica en particular.

REFERENCIAS

Attorresi, H. F., Lozzia, G. S., Abal, F. J., Galibert, M. S. y Aguerri, M. (2009). Teoría de respuesta al ítem. Conceptos básicos y aplicaciones para la medición de constructos psicológicos. *Revista Argentina de Clínica Psicológica*, XVIII (2), 179-188.

Curtin University of Technology. (2001). *Item analysis for tests and surveys using Lertap* 5. Recuperado de: https://bit.ly/2QCo5LT

Herrera, A. N., Sánchez, N. y Jiménez, H. (2001). De la teoría clásica de los test a la teoría de respuesta al ítem. En: E. Ardila, R. Sánchez. *Aula Psicológica* 3, (pp. 293-332). Bogotá: Universidad del Bosque

Hogan, T. P. (2004). *Pruebas psicológicas. Una introducción práctica.* México: El Manual Moderno S. A.