

GRADO DE DIFICULTAD Y PODER DISCRIMINATIVO DE PREGUNTAS DE ELECCION MULTIPLE EN MATERIAS DE PREGRADO DE LA CARRERA DE MEDICINA.

DEGREE OF DIFFICULTY AND DISCRIMINATIVE POWER OF MULTIPLE ELECTION QUESTIONS IN UNDERGRADUATE SUBJECTS OF THE MEDICINE MAJOR.

Dr. Carlos Laforcada Rios ¹

¹Profesor de pre y postgrado de la Facultad de Medicina de la Universidad Mayor de San Andrés, especialista en Neurología.

RECIBIDO: 08/05/17
ACEPTADO: 06/04/17

INTRODUCCIÓN

Poder establecer el grado de dificultad de un examen escrito de elección múltiple en el ámbito de pregrado basándonos en las características de sus preguntas, es una herramienta valiosa a la hora de la planificación y gestión académicas, por otra parte, la determinación del grado de dificultad de las propias preguntas es útil para mejorar sus características y optimizar el proceso de su elaboración a fin de ajustar la evaluación del aprovechamiento de los cursantes de una materia. Con el fin de disponer de un instrumento de evaluación de dicho grado de dificultad desarrollaremos una serie de consideraciones teóricas simples, apoyadas en estadística básica, referentes a las características de las preguntas en exámenes de elección múltiple.

FUNDAMENTOS

Los exámenes de elección múltiple son una colección de preguntas en las que se establece de antemano una respuesta correcta entre 5 opciones que debe ser identificada para cada pregunta, siendo las restantes 4 opciones incorrectas “distractores” que se deben descartar en el proceso. Frente a una pregunta con este formato, de la que no sepamos la respuesta, de una manera intuitiva podemos concluir que existe

una posibilidad entre 5 de elegir la respuesta correcta (proporción $1/5 = 0,2$ o dicho de otro modo el 20 % de elegir la correcta); (por el contrario si sabemos la respuesta correcta, la probabilidad de elegirla es 1 o “la certeza”).

GRADO DE DIFICULTAD DE UNA PREGUNTA POR AZAR

Lo anterior es cierto siempre que cada una de las opciones tenga la misma probabilidad de ser elegida, sin embargo, si una de ellas parece menos probable (por su formato o cualquier razón), se debe considerar un factor de ajuste que repercuta en la probabilidad de elegir la correcta, así por ejemplo, si la probabilidad de elegir una opción incorrecta entre las 5 opciones es de 0,1 (en vez de 0,2), esto aumenta la probabilidad de elegir la respuesta correcta un cierto valor; asumiendo que todas las opciones incorrectas tuvieran la probabilidad de 0,1 la probabilidad de elegir la respuesta correcta solamente por el azar se hace:

$P_{correcta} = 1 - (\text{suma de las probabilidades de las otras opciones})$

$$= 1 - (0,1 + 0,1 + 0,1 + 0,1)$$

$$= 0,6 \quad \text{A esto llamaremos}$$

el “**grado de dificultad**” (GD) de dicha pregunta. (Índice de dificultad

GRADO DE DIFICULTAD POR EFECTOS DEL DISTRACTOR

Asumiendo que un examen sea tomado a un grupo de 100 sujetos que no saben nada de un tema, el patrón de respuesta al azar seguiría una dispersión de 0,2 para cada una de las 5 opciones de respuesta, siempre que todas pudieran ser igual de posibles; el patrón de repuesta cambiaría si una o más de las opciones de respuesta fuera poco verosímil (mal distractor, que hace que casi nadie elija esta respuesta), aumentando la proporción de las otras opciones, según una secuencia aproximada de 0,2; 0,25, 0,33, 0,5 según se reduzca el efecto de los distractores hasta llegar a una proporción cercana a 1 cuando ninguno de los distractores es verosímil, como se observa en el cuadro N° 1. (donde 4d indica 4 distractores adecuados, y pA el porcentaje de respuesta A).

Cuadro N° 1

#distr	A	B	C	D	E	GD	IDD	CALIFIC.
4d	20	20	20	20	20	0,2	5	Muy difícil
3d	25	25	25	25	0	0,25	4	Difícil
2d	33	33	33	0	0	0,33	3	Media
1d	50	50	0	0	0	0,5	2	Fácil
pA	80	5	5	5	5	0,8	1	Muy fácil

El patrón de respuestas del mismo examen en un grupo de 100 sujetos que saben del tema es diferente, pues casi todos o la mayoría identificarán las respuestas correctas y evitarán los distractores dando la siguiente distribución aproximada de respuestas (Cuadro N° 2):

Cuadro N° 2

#	A	B	C	D	E	GD	IDD	CALIFIC.
pS	80	5	5	5	5	0,8	1	Muy fácil

(donde pS indica el porcentaje de respuesta entre los que saben a respuesta A)

Esto ocurrirá para todas las respuesta que conozcan, pero para aquellas que no, el patrón se hará como en el caso anterior, por azar o descarte de distractores.

GRADO DE DIFICULTAD POR EFECTOS DEL NIVEL DE CONOCIMIENTO

Como consecuencia de este análisis, si el grupo que responde por azar una pregunta (bien elaborada y con sus cuatro distractores bien formulados), tiene un promedio de 20 %

de respuestas correctas y el grupo que sabe, tiene un promedio de 80 o 90 % de respuestas correctas, podemos afirmar que el promedio de respuestas correctas cambiará de 20 % al 90 %, según el nivel de conocimiento de la pregunta que tenga un grupo cualquiera; los grupos a ser evaluados habitualmente son muy heterogéneos, con sujetos con poco conocimiento y que responden al azar, hasta sujetos con suficiente conocimiento que responden bien, por lo que el valor del promedio de respuestas correctas reflejara el nivel de conocimiento promedio del grupo a dicha pregunta, es decir, el grado de dificultad de la pregunta para todo el grupo.

Como conclusión de este análisis diremos que grado de dificultad de una pregunta, cuando es aplicada a un grupo de sujetos, **refleja tanto el nivel de conocimiento promedio del grupo, así como la efectividad de los distractores utilizados en la pregunta.**

CÁLCULO DEL GRADO DE DIFICULTAD DE UN EXAMEN

Manualmente o aprovechando las ventajas de la calificación automatizada de exámenes en bases de datos (Ej. "EXCEL"), el procedimiento para determinar el grado de dificultad de una pregunta consistirá en contabilizar todas las repuestas a las 5 opciones de una determinada pregunta en un examen de elección múltiple, identificando el número de respuestas a la opción correcta, esta cifra se divide entre número total de exámenes, con lo que obtenemos la proporción de respuestas correctas, o "grado de dificultad" de la pregunta. Aplicando el mismo procedimiento a todas las respuestas podremos calificar todas las preguntas y adjuntar el grado de dificultad de cada una. Para manejo estadístico podremos calcular el "**promedio del grado de dificultad del examen**", (**XGD**), (que por comodidad llamaremos el **promedio de dificultad** del examen) simplemente sumando los grados de dificultad de todas las preguntas y dividiendo entre el número total de preguntas, (que estadísticamente tiene la distribución normal de las variables continuas).

APLICACIONES EN EL ANÁLISIS DE LAS PREGUNTAS DE UN EXAMEN

A manera de ejemplo analizaremos las preguntas de un examen de neurofisiología de una gestión pasada,

tomado al final del curso y que tuvo un rendimiento medio-bajo con un porcentaje de aprobación del 11 % sobre algo más de 600 cursantes.

Análisis de las respuestas del examen parcial

Utilizamos el programa Microsoft-Excel 2013 para la calificación y el análisis estadístico de los resultados de las 35 preguntas, los datos recogidos permitieron calificar las preguntas de la siguiente manera (Cuadro N° 3):

Cuadro N° 3

Análisis global del grado de dificultad de 35 preguntas de elección múltiple

		Pobl. general	sólo aprobados
PROMEDIO DE DIFICULTAD		4,57	2,74
SIN 1 OUT LAYER		3,677	2,18
ESCALA	VALORES	# Prg.	# Pregs.
Muy difíciles	>5	8	4
Difíciles	4 - 4,99	8	1
Dif. media	3 - 3,99	9	1
Fáciles	2 - 2,99	8	6

Un rápido análisis de los datos de la tabla nos muestra que el índice de dificultad de las preguntas se distribuye bastante homogéneamente entre los resultados crudos ("población general") de todos los exámenes evaluados (lo que descarta que se manipularan), pero estos mismos datos cambian radicalmente cuando solo tomamos en cuenta a los exámenes de los alumnos que aprobaron dicho examen, en estos el porcentaje de preguntas muy fáciles y fáciles es cercano al 83 %; el análisis de las preguntas con bajo índice de dificultad en la población general descubre que las preguntas repetidas² de otros exámenes previos fueron muy fáciles (conclusión: los cursantes revisan bancos de preguntas) como se evidencia en el Ej.1, (Fig.1) la pregunta # 18 del examen muestra: (la opción correcta es la "E").

Figura N° 1

18	Hertzios del rango máximo de percepción sonora del humano						
A	30 a 8.000	18		REPETIDA			
B	2.000 a 15.000	32					
C	5.000 a 20.000	59					
D	20 a 10.000	24					
E	20 a 20.000	459					
	1,289760349	592					

El uso de distractores "débiles" (poco sugestivos, fácilmente descartables) incrementa la proporción de respuestas correctas y este fue el caso de varias preguntas en este examen, también identifica cuales fueron los temas o capítulos más revisados por los cursantes.

La presencia de un distractor "fuerte" (muy parecido a la respuesta correcta) hace preguntas de dificultad media Ej.2 (fig.2). La pregunta #12 del examen muestra: (opcion correcta la "D") Lo evidente de la opción "A" enmascara las opciones "B" y "C", que al ser también correctas indican que la respuesta adecuada deberá ser "D"; estas parecen preguntas altamente discriminativas (como veremos adelante) y se recomienda que se las emplee corrientemente.

Figura N° 2

12	La estimulación del núcleo de Edinger-Wesphall en el mesencefalo provoca						
A	miosis	267		DISTRACTOR FUERTE			
B	acomodación	52					
C	convergencia	27					
D	todas	195					
E	ninguna	56					
	3,061538462	597					

La presencia de dos distractores "sugestivos" establece preguntas de dificultad media y alta con un poder discriminativo medio-alto.

Finalmente, la presencia de dos distractores "fuertes" (cada uno muy sugestivo) hace a la pregunta muy difícil (OUTLAYER). Ej.3. La pregunta # 20 del examen muestra:

Figura N° 3

20	el receptor del reflejo rotuliano se encuentra en						
A	tendón del cuadriceps	183		DOS O MÁS DISTRACTORES FUERTES			
B	piel de la rodilla	19					
C	músculo cuadriceps	17					
D	nivel L2-L4	353					
E	nervio crural	22					
	34,94117647	594					

Donde las opciones "A" y "D" identifican componentes del circuito reflejo rotuliano muy mencionados en la literatura. El análisis de otras preguntas "muy difíciles" muestra que en todas se repite la misma situación, dos distractores altamente sugestivos hacen la pregunta "outlayer"

OTRAS CONSIDERACIONES

Si analizamos el comportamiento de dos grupos de sujetos en un examen, uno que no sabe y otro que si sabe, representados por el grupo de las más bajas calificaciones y el grupo de las más altas respectivamente, veremos que el promedio del grado de dificultad de las preguntas del primer grupo tenderá a ser 0,2 (*azar – alto grado de dificultad, cuando los distractores funcionen correctamente*), mientras que en el segundo grupo este promedio tenderá a ser cercano o mayor a 0,8 (*bajo grado de dificultad*). Esto tiene consecuencias a la hora de enunciar la estrategia de análisis de las preguntas, así en el primer grupo, debemos considerar como “llamativas” y sujetas a revisión, aquellas preguntas que tengan un **grado de dificultad menor** que el promedio del grado de dificultad del examen (XGD), pues siendo respuestas por azar, un grado de dificultad menor significa un factor que distorsiona el patrón de respuesta esperado, mientras que, en el grupo de sujetos con las mejores calificaciones, debemos considerar “llamativas” las preguntas que tengan un **grado de dificultad alto** o mayor que XGD y analizar la proporción de respuestas de cada una de las otras 4 opciones de respuesta, si primara una dispersión por azar (todas más o menos la misma proporción) posiblemente la pregunta estuviera fuera del área de estudio de los alumnos o si el patrón de respuesta marcara un distractor fuerte o muy fuerte, revisar para corroborar y eventualmente modificarlo.

UN PASO ADELANTE – EVALUACIÓN DEL PODER DISCRIMINATIVO DE LAS PREGUNTAS

Otra característica muy interesante a la hora de evaluar las preguntas de un examen, es la de establecer la capacidad de una pregunta para diferenciar correctamente a aquellos cursantes que estudiaron de los que no lo hicieron, lo llamaremos el “poder discriminativo” de las preguntas y que afecta en última instancia a la validez del examen.

LOS EXÁMENES DE ELECCIÓN MÚLTIPLE COMO UN TEST DIAGNÓSTICO

Si asumimos que finalmente un examen es un test diagnóstico diseñado para evaluar el comportamiento (grado de conocimiento) de los cursantes que rindieron una prueba y que cada

pregunta es un elemento de dicho test, entonces podremos aplicar los conceptos y procedimientos de evaluación de los test diagnósticos a las preguntas de dicho examen; en la evaluación de un test diagnóstico se comparan los resultados de la prueba con aquellos de un “gold standard” (valor de referencia o situación verdadera) a fin de calcular la capacidad del test para identificar los casos verdaderos, que se indican como la Sensibilidad y Especificidad del test diagnóstico. En el caso concreto que nos sirve de ejemplo, el “gold standard” lo construimos con los exámenes de los 40 alumnos que mejores notas sacaron en la prueba, asumiendo que son los que estudiaron y los pareamos con aquellos 150 que peores notas sacaron en la prueba “los que no estudiaron” y frente a estos evaluamos cada pregunta del examen para determinar su capacidad de diferenciar a los que realmente estudiaron de los que no lo hicieron. Los datos de las 10 primeras preguntas se muestran en el Cuadro N° 4 adjunta:

Cuadro N° 4

	40 MEJORES					150 PEORES					proporc. En tabla					
	A	B	C	D	E	A	B	C	D	E	SEN	a		ESP		VP+
1	4	3	1	25	7	21	16	13	70	28	0,63	0,467	0,375	0,533	0,57	0,59
2	28	0	2	1	8	32	10	24	30	52	0,7	0,213	0,3	0,787	0,77	0,72
3	0	2	1	8	28	23	23	62	19	0,7	0,127	0,3	0,873	0,85	0,74	
4	2	0	34	0	4	27	17	60	11	33	0,85	0,4	0,15	0,6	0,68	0,8
5	33	4	2	1	0	35	73	18	8	16	0,83	0,233	0,175	0,767	0,78	0,81
6	4	31	0	4	1	67	29	10	32	12	0,78	0,193	0,225	0,807	0,8	0,78
7	6	27	4	0	3	21	31	39	22	37	0,68	0,207	0,325	0,793	0,77	0,71
8	8	0	28	4	0	35	22	23	49	19	0,7	0,153	0,3	0,847	0,82	0,74
9	0	0	4	31	5	23	31	9	35	52	0,78	0,233	0,225	0,767	0,77	0,77
10	0	1	28	2	9	18	25	32	37	36	0,7	0,213	0,3	0,787	0,77	0,72

En la que: el número de respuestas de cada pregunta se inscriben en las columnas A,B,C,D,E tanto de los 40 mejores, como de los 150 peores exámenes, se destaca en color amarillo las respuestas correctas; la proporción de respuestas correctas entre los 40 mejores se inscribe en la columna “a”; en la “c” la proporción agrupada del resto de las respuestas erróneas de este grupo. En la “b” la proporción de respuestas correctas entre los 150 peores y en la “d” la proporción agrupada del resto de respuestas erróneas del grupo; si combinamos estos datos y los agrupamos en el formato de una tabla de contingencia de 2 x 2 como se observa en el Cuadro N° 5, resulta fácil calcular la Sensibilidad (A/A+C) y Especificidad (D/D+B) de cada pregunta del examen en el arreglo de hoja de “Excel” La

columna "a" que es la proporción de sujetos que respondieron adecuadamente del total de 40 sujetos, correspondería a la SENSIBILIDAD de la pregunta y la columna "d" que es la proporción de sujetos que no respondieron adecuadamente a la pregunta entre los 150 sujetos, correspondería a su ESPECIFICIDAD, en las columnas VP+ y VP- se inscriben sus valores predictivos positivos (A/A+B) y negativos (D/C+D) respectivamente, su utilidad queda a ser determinada.

UN ESTIMADOR INTERESANTE: EL ODDS RATIO ACADÉMICO

El indicador que nos pareció más interesante para determinar el poder discriminativo de una pregunta es el equivalente a la **RAZÓN DE PROBABILIDADES**. *Odds ratio* (OR), el que calcularemos en la tabla de contingencia (cuadro N° 5 6) como sigue: $(A \cdot D / B \cdot C)$

		Estudiaron	No estudiaron	totales
Preg. #	Resp. Correcta	Proporción de resp. Correcta entre los que estudiaron A	Proporción de resp. Correcta entre los que NO estud. B	Total de resp. correctas
	Resp. errónea	Proporción de resp. Errónea entre los que estudiaron C	Proporción de resp. Errónea entre los que NO estudiaron. D	Total de resp. erróneas
Totales		40	150	Total

Los resultados obtenidos en el ejercicio de prueba dan los siguientes datos (Cuadro N° 6)::

	40 MEJORES EXAMENES					150 PEORES					proporc. En tabla				
	A	B	C	D	E	A	B	C	D	E	SEN	ESP			
1	0,1	0,08	0,03	0,63	0,18	0,14	0,11	0,09	0,47	0,19	0,63	0,47	0,38	0,53	1,9
2	0,7	0	0,05	0,03	0,2	0,21	0,07	0,16	0,2	0,35	0,7	0,21	0,3	0,79	8,6
3	0	0,05	0,03	0,2	0,7	0,15	0,15	0,15	0,41	0,13	0,7	0,13	0,3	0,87	16,1
4	0,05	0	0,85	0	0,1	0,18	0,11	0,4	0,07	0,22	0,85	0,4	0,15	0,6	8,5
5	0,83	0,1	0,05	0,03	0	0,23	0,49	0,12	0,05	0,11	0,83	0,23	0,18	0,77	15,5
6	0,1	0,78	0	0,1	0,03	0,45	0,19	0,07	0,21	0,08	0,78	0,19	0,23	0,81	14,4
7	0,15	0,68	0,1	0	0,08	0,14	0,21	0,26	0,15	0,25	0,68	0,21	0,33	0,79	7,97
8	0,2	0	0,7	0,1	0	0,23	0,15	0,15	0,33	0,13	0,7	0,15	0,3	0,85	12,9
9	0	0	0,1	0,78	0,13	0,15	0,21	0,06	0,23	0,35	0,78	0,23	0,23	0,77	11,3
10	0	0,03	0,7	0,05	0,23	0,12	0,17	0,21	0,25	0,24	0,7	0,21	0,3	0,79	8,6

Nos parece que el OR en esta aplicación, expresa la proporción de veces que un suceso ocurra frente a la que no ocurra, en este caso la proporción de veces que los sujetos que estudiaron, respondan

correctamente la pregunta, frente a los que no estudiaron y responden erróneamente. De tal manera que por ejemplo, un OR de 3 debemos interpretarlo como una pregunta que es respondida correctamente 3:1 veces más por los que estudiaron que por los que no estudiaron. Un OR de 1 expresa que ambos grupos respondieron igual, vale decir que la pregunta no discrimina entre los unos y los otros; por lo tanto, el poder de discriminación de una pregunta será tanto mayor cuanto mayor sea su OR.

Así pues, como conclusión final de este análisis, proponemos incluir el grado de dificultad y el poder de discriminación como los dos índices calificadoros de cada pregunta de elección múltiple, lo que adquiere importancia a la hora de modificarlas para incluirlas en bancos de preguntas para subsecuentes aplicaciones o para realizar evaluaciones piloto entre grupos especiales a fin de corregirlas, ajustarlas y perfeccionarlas para aplicaciones posteriores.

DISCUSIÓN

Los procedimientos de valoración del grado de dificultad de preguntas en el campo de la medicina y afines, son variados ^{1,2}, se insertan en la fase 4 de análisis de resultados de las pruebas tipo test ³ y a menudo utilizan indicadores de proporción de respuestas, agrupados o no en rangos de valores ⁴, nuestro procedimiento para determinar el grado de dificultad nos parece fiable y de amplia aplicación en diferentes asignaturas, fácil de realizar, sobre todo usando técnicas informáticas y permite extraer conclusiones directas respecto a la formulación de preguntas y sus distractores además de otros ítems. Estas técnicas actualmente son recomendables para evaluar exámenes ya sea para comparar la homogeneidad, poder discriminativo, consistencia y otras variables de los exámenes, su estabilidad a lo largo del tiempo o para comparar poblaciones diferentes. Por otra parte el cálculo del odds ratio diagnóstico en el campo médico ya se ha aplicado en muchas oportunidades ^{5,6} estableciéndose sus ventajas y limitaciones, su aplicación en el campo académico nos parece interesante como un indicador directo y simple del poder discriminativo de una pregunta entre aquellos que estudiaron de aquellos que no lo hicieron, con índices matemáticos conocidos

y ampliamente utilizados y que podríamos desarrollar más aún sobre todo en lo que a su utilidad y significancia se refiere. Finalmente remarquemos que esta modalidad de análisis es incipiente, y su aplicación por nuestra comunidad docente, pudiera derivar en ventajas que aún nos son insospechadas.

ANEXO 1

EL PROBLEMA DE CLASIFICAR LOS DATOS – USO DEL ÍNDICE DE DIFICULTAD

El uso numérico de grado de dificultad en comunicaciones y publicaciones es contraintuitivo, pues un grado de dificultad de 0,6 sugiere un alto grado de dificultad, cuando en realidad está indicando lo contrario; para evitar este inconveniente sugerimos emplear el “**ÍNDICE DE DIFICULTAD**” (IDD) que es la “certeza” (1) de la respuesta correcta, dividida entre el grado de dificultad (GD).

$$(\text{ÍNDICE DE DIFICULTAD}) = 1 / p.\text{correcta}$$

$$(\text{IDD}) = 1 / 0,6$$

(redondeando) (Que tiene aplicación directa a la hora de clasificar una pregunta, como veremos luego).

Si todas las opciones de respuesta tuvieran la misma probabilidad de ser elegidas (0,2) el GD de la pregunta sería el calculado antes:

$$\begin{aligned} \text{GD} &= 1 - (\text{suma de las probabilidades de las otras opciones}) \\ &= 1 - (0,2 + 0,2 + 0,2 + 0,2) \\ &= 1 - 0,8 \\ &= 0,2 \end{aligned}$$

Entonces el índice de dificultad de dicha pregunta sería:

$$\begin{aligned} \text{IDD} &= 1 / \text{GD} \\ &= 1 / 0,2 \\ &= 5 \end{aligned}$$

CLASIFICACIÓN DE LAS PREGUNTAS SEGÚN EL ÍNDICE DE DIFICULTAD

Para fines prácticos, proponemos clasificar el

índice de dificultad (IDD) de las preguntas de elección múltiple en 6 rangos cualitativos en el siguiente cuadro:

Rango de I.D.D.	CLASIFICACION
1 – 1,99	MUY FACIL
2 – 2,99	FACIL
3 – 3,99	MEDIA DIF.
4 – 4,99	DIFICIL
5 – 5,99	MUY DIFICIL
>6	OUTLAYER (fuera de rango)

Lo cual tiene claras ventajas a la hora de comunicar la información y para realizar consideraciones generales.

ANEXOS 2

CÁLCULO DEL GRADO DE DIFICULTAD DE LAS PREGUNTAS DE UN EXAMEN

Presentamos los datos de la corrección automatizada de un examen de fisiología de una gestión anterior, y el proceso al que fueron sometidos para obtener los grados de dificultad utilizando el programa “EXCEL”.

	A	B	C	D	E	F	G	H	I	J	K	L	AJ	AK
1	A1	datos crudos de la correccion automatica del examen												
2		# preg. →	1	2	3	4	5	6	7	8	9	10	34	35
3		resp. Correcta →	D	A	E	C	A	B	B	C	D	C	C	A
4	sujeto	identif. ↓	R	R	R	R	R	R	R	R	R	R	R	R
5	1	4884445	D	D	E	E	A	B	A	C	D	C	C	A
6	2	8369541	A	A	E	C	B	B	B	C	D	C	C	A
7	3	12923581	D	A	D	C	A	B	B	C	D	C	C	A
8	4	8325842	D	A	E	C	A	B	B	C	D	C	C	A
9	5	10017661	C	A	E	C	A	D	B	C	D	C	C	A
10	6	9102613	A	A	C	C	A	B	A	C	D	C	C	A
11	7	11100982	E	C	E	C	A	D	B	C	D	C	C	A
12	8	6146780	E	E	E	C	A	B	B	C	C	E	C	A
13	9	10908184	A	A	E	C	A	D	B	C	D	C	C	A
14	10	8360500	B	A	E	C	A	B	E	C	D	C	C	A

En el cuadro A1 incluimos los datos de la hoja que contienen las 35 preguntas del examen, (sólo se muestran las preguntas 1 a 10 y 34, 35, también se muestran los datos de respuestas únicamente de 10 sujetos, para facilitar el análisis).

La opción de respuesta correcta está consignada en las casillas al lado de "resp. Correcta", los datos deben ser previamente depurados, asegurando una sola letra de respuesta por cada pregunta para todos los sujetos y casillas en blanco para respuestas en blanco y nulas.

A2	# RESPUESTAS										#	PROPORCION DE RESPUESTAS					
	R					exam	en recuadro las correctas					GD	indice ↓				
	A	B	C	D	E		A	B	C	D				E			
1	66	42	46	343	96	593	0,11	0,07	0,08	0,58	0,16	0,58	1,72				
2	207	29	103	79	178	596	0,35	0,05	0,17	0,13	0,3	0,35	2,86				
3	53	92	68	198	186	597	0,09	0,16	0,11	0,33	0,31	0,31	3,23				
4	86	52	322	25	111	596	0,15	0,09	0,54	0,04	0,19	0,54	1,85				
5	268	201	58	21	46	594	0,45	0,34	0,1	0,04	0,08	0,45	2,22				
35	49	6	4	3	3	65	0,08	0,01	0,01	0,01	0,01	0,08	12,5				

en el cuadro A2 que muestra los datos del procesamiento, se muestra el número de respuestas a cada opción (A,B,C,D,E) del total de sujetos para la pregunta 1,2,3,4,5,35; la respuesta correcta se destaca con un recuadro y color, su grado de dificultad (GD) en la columna adyacente, lo mismo que el índice de dificultad (índice).

El procedimiento de análisis se consigna

En el cuadro A3 se muestran las formulas y valores de cada casilla para obtener dichos datos.

A3	# RESPUESTAS										#	PROPORCION DE RESPUESTAS					
	R					exam	en recuadro las correctas					GD	indice ☒				
	A	B	C	D	E		A	B	C	D				E			
1	=BDCONTARA(\$C\$5:\$C\$606;1;BH3:BH4)	=BDCONTARA(\$C\$5:\$C\$606;1;BI3:BI4)	=BDCONTARA(\$C\$5:\$C\$606;1;BJ3:BJ4)	=BDCONTARA(\$C\$5:\$C\$606;1;BK3:BK4)	=BDCONTARA(\$C\$5:\$C\$606;1;BL3:BL4)	=SUMA(BH5:BL5)	=(BH5/\$AX\$5)	=(BI5/\$AX\$5)	=(BJ5/\$AX\$5)	=(BK5/\$AX\$4)	=(BL5/\$AX\$5)	=REDOND(EAR(BQ5;2))	=1/BS5				
2	=BDCONTARA(\$D\$5:\$D\$606;1;BH3:BH4)	=BDCONTARA(\$D\$5:\$D\$606;1;BI3:BI4)	=BDCONTARA(\$D\$5:\$D\$606;1;BJ3:BJ4)	=BDCONTARA(\$D\$5:\$D\$606;1;BK3:BK4)	=BDCONTARA(\$D\$5:\$D\$606;1;BL3:BL4)	=SUMA(BH6:BL6)	=(BH6/\$AX\$6)	=(BI6/\$AX\$6)	=(BJ6/\$AX\$6)	=(BK6/\$AX\$6)	=(BL6/\$AX\$6)	=REDOND(EAR(BN6;2))	=1/BS6				
3	=BDCONTARA(\$E\$5:\$E\$606;1;BH3:BH4)	=BDCONTARA(\$E\$5:\$E\$606;1;BI3:BI4)	=BDCONTARA(\$E\$5:\$E\$606;1;BJ3:BJ4)	=BDCONTARA(\$E\$5:\$E\$606;1;BK3:BK4)	=BDCONTARA(\$E\$5:\$E\$606;1;BL3:BL4)	=SUMA(BH7:BL7)	=(BH7/\$AX\$7)	=(BI7/\$AX\$7)	=(BJ7/\$AX\$7)	=(BK7/\$AX\$7)	=(BL7/\$AX\$7)	=REDOND(EAR(BR7;2))	=1/BS7				
4	=BDCONTARA(\$F\$5:\$F\$606;1;BH3:BH4)	=BDCONTARA(\$F\$5:\$F\$606;1;BI3:BI4)	=BDCONTARA(\$F\$5:\$F\$606;1;BJ3:BJ4)	=BDCONTARA(\$F\$5:\$F\$606;1;BK3:BK4)	=BDCONTARA(\$F\$5:\$F\$606;1;BL3:BL4)	=SUMA(BH8:BL8)	=(BH8/\$AX\$8)	=(BI8/\$AX\$8)	=(BJ8/\$AX\$8)	=(BK8/\$AX\$8)	=(BL8/\$AX\$8)	=REDOND(EAR(BP8;2))	=1/BS8				
5	=BDCONTARA(\$G\$5:\$G\$606;1;BH3:BH4)	=BDCONTARA(\$G\$5:\$G\$606;1;BI3:BI4)	=BDCONTARA(\$G\$5:\$G\$606;1;BJ3:BJ4)	=BDCONTARA(\$G\$5:\$G\$606;1;BK3:BK4)	=BDCONTARA(\$G\$5:\$G\$606;1;BL3:BL4)	=SUMA(BH9:BL9)	=(BH9/\$AX\$9)	=(BI9/\$AX\$9)	=(BJ9/\$AX\$9)	=(BK9/\$AX\$9)	=(BL9/\$AX\$9)	=REDOND(EAR(BN9;2))	=1/BS9				

REFERENCIAS

1. César Orrego E., Gonzalo Pacheco L., Miguel Inga A., Raymundo Casas N. Evaluación del índice de dificultad de las preguntas de la prueba general de admisión 2014-II a la Universidad Nacional Mayor de San Marcos. El caso de la asociaciones lineales de un agrupamiento de carreras profesionales REVISTA IIPSI VOL. 17 - N.º 1 – 2014.
2. Lewis R. Aiken Test psicológicos y evaluación 11 ed. PEARSON - PRENTICE HALL EDUCACION Cap. 4 pags. 62 – 83. Mexico 2003 ISBN 970-26-0431-1
3. Torrado M. Pruebas de Rendimiento Universidad de Barcelona <http://diposit.ub.edu/dspace/bitstream/2445/21285/1/Pruebas%20de%20rendimiento.pdf> Pag. 38 – 50
4. Dificultad del examen MIR –Asturias curso-mir.com/Como-ha-sido-el-ultimo-examen-MIR/dificultad-del-examen.html
5. Pita Fernandez S.; Pertegas Diaz,S.; Pruebas diagnósticas: sensibilidad y especificidad. Fistera.com. Unidad de epidemiología clínica y Bioestadística. Complejo Hospitalario Universitario de Coruña. Cad Aten Primaria 2003, actualizada el 7/12/2010.
6. Sebastian Bravo–Grau, Juan Pablo Cruz Estudios de exactitud diagnostica: herramienta para su interpretación Revista chilena de radiología Vol 21 # 4, 2015; 158-164. En “<http://www.scielo.cl/pdf/rchradiol/v21n4/art07.pdf>”