

Análisis de las calificaciones. Informes

1.- Estadística de ítems. Índice de dificultad (ID).

¿Qué pensaría de una pregunta que la aciertan los que peor nota han obtenido en el examen y la fallan los que mejor nota han obtenido?

Y en la situación contraria, una pregunta acertada por los que mejor nota han obtenido y fallada por los que peor nota obtuvieron.

Consideraríamos que en el primer caso dicha pregunta ha sido mal formulada, no ha sido entendida, en definitiva, es una “mala” pregunta porque discrimina equivocadamente a nuestra población evaluada, deberíamos eliminarla.

En el segundo caso es una “buena” pregunta ya que discrimina a las personas según su conocimiento de forma adecuada (correcta).

El estudio de ítems analiza la correspondencia entre la dificultad de las preguntas empleadas y la población que las ha respondido.

Es importante destacar que los resultados de este informe deben utilizarse con prudencia y diferenciándolos según el tipo de prueba realizada: examen objetivo de conocimientos, psicológico, aptitudes, etc.

- **GS** - Grupo Superior (fuerte): 27% de la población examinada con mejor nota.
- **GI** - Grupo Inferior (débil): 27% de la población examinada con peor nota.
- El resto de la población no se utiliza para este cálculo.

Este proceso permite obtener un análisis estadístico de los ítems contestados en los diferentes test de conocimiento, en concreto los índices de *dificultad* y de *discriminación*.

- **Índice de Dificultad (ID)**: Determina en qué medida una pregunta es fácil o difícil, indicando la proporción de los que la aciertan con respecto a los que lo intentan.
- **Índice de Discriminación (Di)**: Determina si la pregunta permite distinguir el grupo fuerte del débil.

Índice de Dificultad (ID) *Ver anexo.:

Se visualizarán las respuestas acertadas, las respuestas contestadas, el índice de dificultad y la evaluación del índice.

$$\text{Índice Dificultad ID} = (\text{Respuestas Acertadas} / \text{Respuestas Contestadas}) * 100$$

En otras palabras, la “Estadística de ítems” es el porcentaje de aciertos. Si el índice tiende a 100, la pregunta es muy fácil. Si tiende a 0 la pregunta es muy difícil. La evaluación del índice se efectúa a la siguiente forma:

Índice	Evaluación
Del 0 al 15 %	Muy difícil
Del 16 al 31 %	Relativamente difícil
Del 32 al 68 %	Dificultad media
Del 69 al 84 %	Relativamente fácil
Del 85 al 100 %	Muy fácil

Importante:

- Recuerde que el ID es aplicable a pruebas objetivas, careciendo de interés en pruebas de personalidad donde no existe el acierto/error.
- No se puede considerar una medida intrínseca de la pregunta ya que depende de la población sujeta a examen, por lo que con diferentes poblaciones el ID de la misma pregunta puede variar.

Índice de Discriminación (Di) *Ver anexo.:

Una pregunta (ítem) tiene capacidad de discriminación si permite diferenciar al grupo fuerte del débil o, lo que es lo mismo, si distingue al individuo que obtiene alta calificación del que no la tiene. Cuanto más alto es el índice de discriminación, el ítem diferenciará mejor a las personas con altas y bajas calificaciones.

Si todas las personas del grupo fuerte ($N^{\circ}A$) contestan correctamente un ítem y todas las personas del grupo débil ($N^{\circ}B$) contestan incorrectamente, entonces $D_i = 1$ (valor máximo de este indicador); si sucede lo contrario, $D_i = -1$ (valor máximo negativo); si ambos grupos contestan por igual, $D_i = 0$ (valor mínimo de discriminación).

La calidad y recomendación del índice se efectúa de la siguiente forma:

Índice	Calidad	Recomendación
> 0.39	Excelente	Conservar
0.30 – 0.39	Buena	Posibilidades de mejora
0.20 – 0.29	Regular	Necesidad de revisar
0.00 – 0.19	Pobre	Descartar o revisar en profundidad
< -0.01	Pésima	Descartar definitivamente

**PRECAUCIÓN: Las etiquetas utilizadas en el informe buscan simplificar la interpretación de los resultados, pero no siempre son completamente adecuadas. Por ello, ten en cuenta que el índice de dificultad no considera las omisiones, por lo que un ítem con muchas omisiones podría parecer fácil. Al revisar los resultados de discriminación, asegúrate de que ningún distractor sea seleccionado por muchas más personas que la opción correcta. Además, si el número de personas que responden al examen es pequeño, la interpretación de los resultados puede verse comprometida.*

2.- Informe de ítems detallado.

Informa de qué población ha respondido a cada opción de respuesta, incluyendo aquellos que no han contestado (aparece marcado con un punto “.”) o doble respuesta (aparece un signo de interrogación “?”).

Un aspecto importante de este informe nos permite averiguar si las distintas opciones de respuesta son igualmente “consistentes”, es decir, aquella pregunta cuyas opciones de respuesta erróneas han sido respondidas por un volumen de personas semejante tiene un factor de azar correcto, es decir podríamos resumirlo en que todas las opciones de respuesta han sido formuladas con la misma probabilidad de respuesta. Es un ítem “correcto”, independientemente de su “calidad” y evaluación que se realiza en función del Grupo Superior (GS) y Grupo Inferior (GI).

3.- Distribución de notas.

La gráfica obtenida muestra cuántas personas hay en cada intervalo de calificaciones.

Debería aparecer como una campana (campana de Gauss) centrada en la mitad de la calificación máxima. Un máximo de la campana hacia la derecha (mejores calificaciones) muestra que el examen ha sido “sencillo” para esta población y, al contrario, hacia la izquierda, ha sido difícil.

La parte inferior muestra la frecuencia de personas por intervalo de calificación y su acumulado. De este modo podríamos conocer para una calificación dada, cuántas personas aprobarían y en consecuencia, cambiando el aprobado a dicha calificación, recalificar obteniendo los aprobados exactos que se desean.

Un caso particular sería cuando existe más de un máximo, es decir, no hay una campana de Gauss si no dos campanas o más superpuestas. Dos máximos expresan que hay dos poblaciones homogéneas pero diferentes entre sí. Hace años esto representaba un indicio claro de que una parte de la población examinada había tenido acceso a información diferente y realizó el examen en condiciones claramente distintas al resto.

Actualmente observamos en exámenes con grandes poblaciones la disgregación de la campana de Gauss con uno o más máximos relativos. Esto representa la rotura de nuestra sociedad, es decir, empieza a no existir un solo “edificio social” sino varios.

4.- Informe de frecuencias.

Este informe analiza a la población calificada y no tanto al examen en sí como el informe de ítems.

Curtosis: Generalmente las calificaciones más frecuentes están cerca de la media. Esto se considera una distribución normal, si las calificaciones están muy concentradas alrededor de la media (poca dispersión), mientras que otras veces tienen valores más dispersos (mayor dispersión). La curtosis nos dice si una distribución es más concentrada (con un pico) -población homogénea en respuesta al examen- o menos concentrada (más achatada).

Simetría: Nos destaca si la población inferior a la calificación con el mayor número de exámenes y la superior son semejantes o no. Un examen con una simetría alta es un examen diseñado con preguntas fáciles, difíciles e intermedias donde no se establecen “escalones u obstáculos” de conocimiento. Es decir, hay el mismo incremento de dificultad en las preguntas, y el mismo número de ellas en cada rango de dificultad. Del mismo modo hay el mismo porcentaje de examinandos en cada escalón de conocimiento.

CV: Un coeficiente de variación bajo indica que los datos están más uniformemente distribuidos alrededor del promedio, mientras que un valor alto señala una mayor dispersión.

El análisis de la distribución de notas junto con los estadísticos del análisis de frecuencias en diferentes poblaciones y en especial con la evolución temporal, es decir, en años sucesivos con la misma materia evaluada, nos permite medir y cuantificar cómo está cambiando la sociedad y en qué dirección, además de la adecuación e interés por la formación realizada.

Frecuencias: ¿Cuántas veces se repite una calificación? La frecuencia en cada nivel de calificación está aumentando o disminuyendo con respecto a los años precedentes; es decir, aumenta la población con buenas calificaciones o disminuye. ¿Mejoramos o empeoramos?

Sería un análisis directo, pero ¿qué sucedería si unas calificaciones, desordenadas, unas aumentan y otras disminuyen sin ser ni buenas ni malas? Unas buenas aumentan y otras disminuyen, unas malas calificaciones y otras disminuyen. No hay la misma población con buenas calificaciones y con malas en cada año, es decir, no hay simetría con el paso de años.

¿Qué significa? La población pierde homogeneidad, la formación y los exámenes pierden correspondencia y adecuación con la población, los recursos empleados (materiales y humanos) son incoherentes con los objetivos.

Es usted, ahora, el que debe analizar los resultados de su examen y tomar las decisiones oportunas.