

Programa informático para el análisis de ítems de pruebas dicotómicas

© Manuel Miguel Ramos-Álvarez

junio de 2005

Universidad de Jaén

mramos@ujaen.es

Resumen:

Los elementos que componen una prueba objetiva, ya sea un test un examen o algún instrumento de este tipo, frecuentemente se basan en un formato dicotómico, de manera que hay dos posibles respuestas que son de carácter excluyente. En este contexto, se pueden analizar una serie de parámetros estadísticos individuales para cada uno de los ítems que constituyen el sustrato formal sobre el que se asientan las bases científicas de cualquier instrumento o técnica de medida. El programa Analitem.xls es muy sencillo de manejar y se implementa en Excel, todo lo que requiere es incorporar filas y/o columnas para ajustarse a la muestra del investigador. El programa proporciona Índices de Dificultad, Discriminación, Homogeneidad, Fiabilidad y Validez; así como Alfa de Cronbach y Validez predictiva. También se proporciona un resumen de los conceptos fundamentales en el material que figura a continuación.

➤ [Para acceder al programa pulse aquí](#)

Índice de contenidos

1. ANÁLISIS DE DIFICULTAD.	2
1.1. RELACIÓN CON OTROS PARÁMETROS DEL TEST.	2
2. ANÁLISIS DE DISCRIMINACIÓN (IDN).	2
2.1. CÁLCULO BASADO EN LA CORRELACIÓN ÍTEM-TEST.	3
2.2. RELACIÓN CON OTROS PARÁMETROS DEL TEST.	3
3. ANÁLISIS DE FIABILIDAD Y VALIDEZ.	4
3.1. ANÁLISIS DE LA VALIDEZ (IV_j ó R_{jY})	4
3.2. ANÁLISIS DE LA FIABILIDAD (IF_j ó R_{jT})	4
3.3. RELACIÓN CON OTROS PARÁMETROS DEL TEST.	4

1. ANÁLISIS DE DIFICULTAD.

La proporción de examinados que responden correctamente a ese ítem.

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

Para elementos de elección múltiple:

$$IDd_j = \frac{A - \frac{E}{k-1}}{N}$$

Mejor llamarle índice de *facilidad*.

1.1. Relación con otros parámetros del test.

- En relación con el test global:

$$\bar{X} = \sum_j p_j \equiv \sum_j IDd_j$$

- Además, se relaciona fundamentalmente con la varianza del *del ítem y del test*.

- Para ítems (dicotómicos) se cumple:

$$S_j^2 = p_j \cdot q_j \equiv IDd_j(1 - IDd_j);$$

$$S_j = \sqrt{IDd_j(1 - IDd_j)}$$

- Para el test, se cumple:

$$S_x = \sum_j (S_j) \cdot (R_{jx}) \equiv \sum_j [IDd_j(1 - IDd_j)] \cdot [IDn_j]$$

En concreto, la Varianza será máxima para los valores intermedios de P_j ; en otras palabras la Dificultad media maximiza la Varianza del ítem. Luego, la máxima cantidad de información acerca de las diferencias entre los examinados se obtiene cuando las proporciones son intermedias ($P_j = Q_j = 0,5$).

2. ANÁLISIS DE DISCRIMINACIÓN (IDN).

- Un ítem tiene elevado poder discriminativo si contribuye a la distinción, discriminación, de los que puntúan alto en el test de los que puntúan bajo en el mismo. Es decir, si contribuye a diferenciar individuos que exhiben el constructo o rasgo de aquellos que no lo exhiben. Convencionalmente se suele denominar grupo de personas “**competentes**” a aquellas que puntúan alto en la prueba e “**incompetentes**” a los que puntúan bajo (nomenclatura un tanto desafortunada pero inevitable).
- Este análisis se puede efectuar de dos maneras, bien estimando la relación del ítem con el test o bien a través de la división del grupo en dos subgrupos, los competentes y los incompetentes, y estimando la proporción de aquellos que eligen el ítem en cada uno de los subgrupos.

2.1. Cálculo basado en la correlación ítem-test.

- Como consecuencia de lo anterior podría definirse el índice de discriminación como la **correlación** entre las puntuaciones de los sujetos en el ítem y sus puntuaciones en el test (puntuación total).
- La correlación a utilizar dependerá de los formatos de ítems y test y teniendo en cuenta que ha de descontarse la contribución que hace el ítem al test total ya que si no se estaría elevando impropia y espuriamente la estimación.
- Veamos los 4 casos más importantes:

1) **Biserial-puntual.** Ítem dicotómico y el test es cuantitativo continuo.

$$IDn_j \equiv R_{bp} = \frac{\bar{X}_A - \bar{X}_X}{S_X} \cdot \sqrt{\frac{p_j}{q_j}}$$

-
- A se refiere a los Aciertos en el test.
 - p es la proporción de los sujetos que aciertan el ítem y q la de los que lo fallan ($q = 1 - p$).
 - X se refiere al test; por tanto \bar{X}_X es la media del test y S_X su desviación típica.
-

2) **Biserial.** Ítem dicotomizado a partir de una variable continua normal y el test es cuantitativo continuo (i.e. ítems que admiten gradación de Resp. pero se dicotomizan).

$$r_b = \frac{\bar{X}_A - \bar{X}_X}{S_X} \frac{P_j}{y}$$

-
- y expresa la ordenada que corresponde al valor de la puntuación típica en la Normal que deja por debajo un área igual a "p".
-

- No emplearlo si se sospecha una distribución platycúrtica o bimodal.

3) **Coefficiente PHI.** Ítem y test son casos dicotómicos. Es una mera aplicación de Pearson, luego se puede estimar con dicha fórmula.

4) **Tetracórica.** Si las 2 están dicotomizadas a partir de variables continuas y según un modelo de la Normal. También es una aplicación de Pearson y su cálculo es complejo, luego se puede usar Pearson como aproximación.

5) **Basado en Pearson.** Si tanto el ítem como el total son variables continuas. Las escalas tipo Likert (ordinales) requieren el índice de Spearman que también se puede aproximar mediante Pearson.

2.2. Relación con otros parámetros del test.

1) **Variabilidad.**

- La variabilidad de un *test* es una medida de la capacidad discriminativa del mismo ya que por ejemplo si la desviación típica valiese cero es porque no habría discriminación alguna, pues todos los sujetos manifiestan la misma puntuación (no

se diferencian en absoluto). Además, es fácil ver que la capacidad discriminativa del test (S_x) está íntimamente relacionada con el IDn de los ítems.

$$S_x = \sum_j S_j R_{jX} = \sum_j (S_j) \cdot (IDn_j)$$

Nota: Descontar la influencia del ítem en el test global:

$$r_{j(x-j)} = \frac{r_{jx} S_x - S_j}{\sqrt{S_j^2 + S_x^2 - 2r_{jx} S_j S_x}}$$

3. ANÁLISIS DE FIABILIDAD Y VALIDEZ.

- Pretendemos medir cuánto de fiable o válido resulta cada ítem respecto al test total, lo que sería pertinente a la hora de detectar ítems concretos que pudieran estar fallando cuando la fiabilidad/validez global fuera(n) baja(s).

3.1. Análisis de la validez (IV_j ó R_{jY})

- La validez del ítem expresa la correlación del ítem con el criterio. Se calcula como el índice de discriminación pero correlacionando el componente con el criterio global en lugar del test.
- Ahora no existe el problema de tener que descontar el ítem del total ya que se trata del criterio –que incluye otros elementos distintos para medir el constructo–.

3.2. Análisis de la fiabilidad (IF_j ó R_{jY})

- Expresa la ponderación del poder discriminativo en función de la dispersión del mismo. Formalmente:

$$IF_j = R_{jY} = S_j R_{jX}$$

3.3. Relación con otros parámetros del test.

1) **Variabilidad** del test y **homogeneidad** de los ítems.

Según la formulación que veíamos más arriba para relacionar variabilidad-discriminación, quedará claro que ésta también se relaciona con la fiabilidad.

$$S_x = \sum_j S_j R_{jX} = \sum_j (S_j) \cdot (IDn_j) = \sum_j IF_j$$

Tener en cuenta:

$$S_j = \sqrt{P_j(1-P_j)} \Rightarrow S_x = \sum_j \sqrt{P_j(1-P_j)} r_{jx} \text{ Para ítems dicotómicos.}$$

- De aquí se desprende que la homogeneidad del ítem y su discriminación o fiabilidad son conceptos diferentes aunque estrechamente relacionados. O de otra manera, que el poder de discriminación del test, a través de la variabilidad, depende de la dificultad y la discriminación de los componentes y no sólo de la discriminación de los mismos.
- Es decir que depende de la variabilidad de los ítems y de su poder de discriminación o alternativamente de su dificultad (recogido en S_j) y de su poder de discriminación.

- Luego, para maximizar la variabilidad del test hay que considerar simultáneamente la interacción de los dos tipos de índice.

2) **Fiabilidad como consistencia**

$$\alpha = \frac{n}{n-1} \cdot \left(1 - \frac{\sum_j S_j^2}{\left[\sum_j S_j r_{jx} \right]^2} \right)$$

- Según esta reescritura del coeficiente alfa de Cronbach, alfa se relaciona íntimamente con la homogeneidad del ítem y con su grado de discriminación o de fiabilidad. Además lo hace de manera directa si consideramos que lo que va en el denominador a su vez lo hace restando de la unidad.

3) Validez del test.

$$\bullet \quad R_{xy} = \frac{\sum S_j r_{jy}}{\sum S_j r_{jx}}$$

$$\bullet \quad R_{xy} = \frac{\sum \sqrt{P_j(1-P_j)} r_{jy}}{\sum \sqrt{P_j(1-P_j)} r_{jx}}$$

Para ítems dicotómicos

- Aquí queda también clara la relación de validez con todos los conceptos vistos. Tiene una relación directa con los índices de validez de los ítems e inversa con los índices de fiabilidad o de discriminación de los mismos. Esto recoge una **paradoja** importante. “*Si optimizamos la fiabilidad al incrementar el poder de discriminación (ecuación de alfa), esto rebajaría la validez (ecuación que acabamos de ver)*”. Luego, se debe buscar un punto de equilibrio.