EXTRACCION DEL COMPONENTE DE DIFICULTAD EN LA EVALUACION DE ESCALAS BASADAS EN ITEMS DICOTOMICOS

Pere Joan FERRANDO PIERA y Urbano Lorenzo SEVA
Departamento de Educación y Psicología. Facultad de Filosofía y Letras
Tarragona. UNIVERSIDAD DE BARCELONA

RESUMEN

Se presenta una aplicación del método propuesto por Lawley y Maxwell para extraer un componente de dificultad de una matriz de correlaciones obtenidas en reactivos dicotómicos. La comparación entre el análisis en componentes principales sobre la matriz original y sobre la matriz residual ponen de manifiesto una notable semejanza en las soluciones. Sin embargo, en el análisis sobre la matriz original emerge un componente muy similar al componente de dificultad extraido inicialmente de acuerdo al método propuesto. Los resultados sugieren la necesidad de proseguir en esta línea de investigación.

Palabras Clave: Análisis Factorial, componente de dificultad, items dicotómicos.

ABSTRACT

Extraction of a difficulty component from a set of dichotomous items. We present here an application of de method proposed by Lawley and Maxwell for the extraction of a difficulty component. This method was applied over a correlation matrix obtained from a set of dichotomous items. Comparisons between principal component analysis obtained from the original matrix and from the residual matrix seems closely similar, but in the first analisis a component emerges roughly similar to a difficulty component. Further research seems necessary to be done.

Key words: Factorial analysis, difficulty componet, dichotomous items.

La investigación acerca de los denominados "factores de dificultad" tuvo un notable interés en los años 40 y se reavivó de nuevo a finales de los 60 con el auge de la denominada actualmente "Teoría de Respuesta a los Items". La principal razón de este renovado interés es la de que, para muchos de los modelos basados en esta teoría, el supuesto de unidimensionalidad tiene una importancia crucial (Hambleton y Swaminathan, 1985). Sin embargo, en realidad, el supuesto de unidimensionalidad es básico en cualquier modelo psicométrico que utilice puntuaciones de escala basadas en combinaciones lineales de items.

El enfoque actual acerca de estos factores difiere en general del enfoque que le dieron

Psicothema, 1992, vol. 4, nº 1, pp. 269-276
ISSN 0214 - 9915 CODEN PSOTEG

los factorialistas y psicómetras en décadas pasadas. McDonald (1967,1981,1985), por ejemplo, considera zanjada ya la cuestión afirmando que estos pseudo-factores son debidos a la no linealidad en las relaciones variable –factor. La solución, en este caso es la utilización de modelos no lineales de análisis factorial, modelos que ha venido desarrollando desde hace tiempo el propio McDonald y que actualmente son ya operativos y se utilizan con buenos resultados en la investigación psicométrica (Fraser y McDonald, 1988). Por otra parte, si se cumplen los supuestos de linealidad, entonces pueden utilizarse legítimamente las medidas usuales de correlación para datos binarios (el coeficiente phi).

En definitiva, podemos decir que el núcleo del problema se desplaza desde la búsqueda del mejor coeficiente de asociación para datos dicotómicos (tema polémico en los 40) hasta la verificación del cumplimiento del requisito de linealidad.

Si bien la postura de McDonald parece tener actualmente una amplia aceptación, lo cierto es que algunos investigadores (Bernstein, 1988) siguen defendiendo la postura clásica con respecto al problema de los factores de dificultad. El trabajo que presentamos se basa también en este enfoque, por lo que trataremos de explicarlo con cierto detalle.

Los métodos de análisis factorial y el método de componentes principales normalizados toman como punto de partida la matriz de correlaciones entre variables. Como modelo de análisis todos estos métodos asumen el supuesto de normalidad multivariante y, en consecuencia, el de normalidad bivariante para todos las correlaciones entre pares de variables. En el caso de modelos lineales la medida de asociación requerida es el coeficiente de correlación de Pearson.

El coeficiente de correlación producto–momento aplicado a datos dicotómicos (en psicometría items que se responden como si–no, verdadero–falso, acierto–error...) recibe el nombre de coeficiente phi. La expresión habitual para este coeficiente puede desarrollarse a partir de la fórmula de la correlación de Pearson, teniendo en cuenta que, en el caso de dicotomías las medias son proporciones (p) y las varianzas son los productos p q.

$$r\varnothing = \frac{pxy - px\,py}{\sqrt{(pxqx)}\,\sqrt{(pyqy)}}$$

Donde pxy es la proporción de sujetos que obtienen 1 en ambos items. Es evidente que el valor máximo que puede alcanzar pxy es el menor de los valores px o py.

Si las dos variables tienen la misma media (px = py) entonces, el valor máximo positivo que pueda alcanzar phi será:

$$\frac{px - (px)^2}{px\,(1 - px)} = 1$$

En cualquier otro caso, el valor máximo positivo será menor que uno (Guilford, 1950; Cureton, 1959). Una forma simple de ver esto es considerar, como señala Maxwell (1977) que en un rango comprendido entre 0.2 – 0.8, las varianzas no difieren demasiado, es decir, el denominador permanece relativamente constante. Respecto al numerador, el valor pxy sólo podrá valer tanto como la menor de las p, mientras que el producto será la mayor por la menor. El máximo valor en la diferencia se obtendrá cuando las dos sean iguales. Esta idea es válida para cualquier modelo de correlación; cuando el resto de los factores se mantienen constantes, el valor máximo de correlación se obtiene cuando la forma de las distribuciones es idéntica (Bernstein, 1988).

Si se analiza la matriz R (phi) en componentes principales, se espera la emergencia de los denominados "componentes de dificultad". Se denominan de este modo a los componentes espurios cuyas saturaciones son proporcionales a las medias (p) de las variables correspondientes. En el caso de que las desigualdades entre las medias sean notables (es decir, en el caso de abundancia de items muy fáciles o muy difíciles), se supone que la emergencia de estos componentes puede distorsionar la solución e imposibilitar la interpretación teórica de la dimensionalidad, ya que aparecería un mayor número de dimensiones (en términos de componentes) de las que realmente existen y, además, algunos componentes no guardarían relación con el contenido de los reactivos sino con su media.

Llegados a este punto, puede ser necesario apuntar que en este artículo abordamos el estudio de la dimensionalidad de una escala a partir del análisis en componentes principales. No es este el único enfoque posible, Hattie (1984, 1985) señala tres enfoques más (patrones de respuesta, fiabilidad, y modelos de rasgo latente); ninguno de ellos resulta totalmente satisfactorio y, de acuerdo con las revisiones del propio Hattie u otras relevantes (Hambleton y Rovinelli, 1986), debemos concluir que aún no se ha llegado a este índice satisfactorio cuya necesidad apuntaba Lord (1980).

Aunque no entra en los objetivos de este estudio hacer un análisis de estos métodos, sí puede apuntarse que las deficiencias de la utilización del análisis en componentes principales (y de los métodos factoriales en general) se centran en el hecho de que una escala puede ser unidimensional en el sentido de un único rasgo latente subyacente a los datos y, en cambio, ser multidimensional en términos de varios componentes.

Quizás la mejor explicación de esta discrepancia se encuentra en Lord y Novick:

> "..La dimensionalidad de un espacio latente completo no depende de los supuestos respecto a las distribuciones, ni de la elección del coeficiente de correlación interitem, ni de cualquier tipo de transformación de las variables latentes.....la dimensionalidad es un concepto más básico que el número de factores comunes". (Lord y Novick, 1968, p. 382).

Para concluir podríamos decir que el análisis en componentes principales nos dará una orientación acerca de la posible dimensionalidad de los items, pero que, en ningún caso nos la determinará.

Volviendo al tema de la limitación del valor máximo de phi y su repercusión en los componentes extraidos, se han propuesto varios métodos para abordar el problema, aunque ninguno totalmente satisfactorio. La solución más antigua consiste en estimar el valor máximo que puede alcanzar el coeficiente phi dadas las proporciones de ambas variables. Este valor se denomina phi–máxima (Cureton, 1959). Una vez determinado, como indicador de correlación se utilizan los cocientes phi/phi–max. Sin embargo como señala Comrey (1985), cuando se trabaja con valores de p extremos la utilización de este indicador suele dar lugar a soluciones Heywood (comunalidades estimadas mayores a uno) por lo que suele resultar inapropiado utilizarlo, al menos en análisis factorial. En el caso del análisis en componentes, la comunalidad estimada no representa un problema ya que lo que se analiza es la varianza total (que se fija a uno en el análisis normalizado). Sin embargo, aún así puede suponerse que los valores de fuera de la diagonal estarán bastante distorsionados.

Maxwell (1977) basándose en un trabajo inicial de Lawley (1943,1944) propone la estimación de un primer componente general de dificultad. Una vez determinado dicho componente, se reproduce a partir de él la matriz de correlaciones. Se sustrae la matriz inicial de la reproducida y se lleva a cabo el análisis sobre la matriz residual. Más específicamente el método es el siguiente:

Como punto de partida se asume que los coeficientes phi de la matriz de correlaciones

R se hallan sistemáticamente afectados por las desigualdades entre las proporciones de las variables a correlacionar.

Se elabora un vector columna de valores diferenciales restando la proporción de cada reactivo de la media de todas las proporciones. Llamemos p a este vector, en el que cada elemento es: p(i) – Media p.

Si ahora elaboramos una matriz a partir del producto pp', los elementos de esta matriz estarán relacionados en forma aproximadamente proporcional con los elementos de la matriz R, de tal forma que el producto cpp', donde c es una constante de proporcionalidad, reproducirá, en forma más o menos aproximada la matriz R. Por supuesto, cuanto más determinados estén los valores de R por las diferencias de proporciones, más aproximada será la matriz reproducida.

El problema del método radica en la estimación de la constante c. Si denominamos R* a la matriz reproducida y E a la matriz residual entonces:

$$E = R - R^*; \; R^* = cpp'$$

El criterio propuesto por Maxwell es el de minimizar la traza de E'E. Este criterio supone minimizar las diferencias al cuadrado entre las diagonales de R y R* y, por tanto, maximizar la varianza de R explicada por R*. Como es sabido, este es el criterio general seguido por los métodos de condensación.

Derivando la traza de E'E con respecto a c, e igualando a 0 se obtiene:

$$c = \frac{(p' R\, p)}{(p'\, p)^2}$$

Las saturaciones de las variables en el componente de dificultad vendrán dadas por:

$$\sqrt{c * p}$$

La matriz residual E, de rango p–1 puede utilizarse ahora como punto de partida para la extracción de los restantes componentes.

En el ejemplo que presentaremos a continuación, aplicamos el método de Lawley y Maxwell a la matriz de coeficientes phi. Hemos elegido como ejemplo un grupo de reactivos pertenecientes a versión castellana de la escala de extraversión del EPQ–R (Eysenck, Eysenck y Barrett, 1985; Aguilar, Tous y Andrés, 1990). Existe abundante evidencia respecto a los cuestionarios anteriores, EPI y EPQ–A de que la escala de extraversión agrupa dos clusters de items claramente definidos. Desde el modelo de Eysenck, estos clusters corresponden a las subdimensiones "Sociabilidad" e "Impulsividad" (Rocklin y Revelle, 1981; Luengo, 1986).

En este trabajo hemos seleccionado a priori los 10 items que, desde la teoría, pretenden medir la dimensión de "Sociabilidad". Teóricamente el análisis de dichos items debe dar lugar a una clara solución unidimensional. Inicialmente, además, hemos verificado en este grupo la existencia de algunos items con medias bastante distanciadas de p=0.5 por lo que parece razonable esperar la aparición de componentes de dificultad de acuerdo al modelo expuesto.

Dado que, como hemos apuntado antes, la solución en términos de número de componentes es tan sólo orientativa, nos hemos basado en el supuesto de que el análisis de un conjunto de reactivos aproximadamente unidimensionales dará lugar a una solución caracterizada por una primera raiz latente muy elevada siendo residuales las restantes, solución que puede ser evaluada gráficamente en el scree–plot (Cattell, 1966). No hemos considerado necesario utilizar índices más específicos.

METODO

Sujetos:

Se utilizaron datos pertenecientes al estudio psicométrico de la adaptación al castellano del cuestionario EPQ–R (Aguilar, Andrés y Tous 1990). La muestra estaba formada por 866 estudiantes de primero y segundo curso de la facultad de Psicología de la Universidad de Barcelona; 201 varones y 665 mujeres. La edad media era de 18 años.

Instrumentos:

Los paquetes estadísticos disponibles en el laboratorio no resultaban apropiados para el análisis, dado que todos ellos extraían simultáneamente las raices latentes y los autovectores asociados. Para ello se elaboró un programa en el lenguaje "MATRIX" del SPSS–X release 4.0.

Procedimiento:

La extracción del componente de dificultad se llevó a cabo en la forma que se describe en la introducción. La extracción de los sucesivos componentes a partir de la primera matriz residual se realizó mediante el método en iteraciones propuesto por Hotelling (1933) extrayendo los componentes en forma sucesiva. El criterio de convergencia se fijó en una diferencia menor a .001.

La solución obtenida utilizando este método se comparó con la solución directa aplicando el análisis en componentes principales a la matriz de correlaciones inter–item. Las comparaciones entre componentes se llevaron a cabo mediante el coeficiente de congruencia propuesto por Tucker (1951).

$$CC\,(i,j) = \frac{\Sigma\ ai\ aj}{\sqrt{lambda(i)}\ \sqrt{lambda(j)}}$$

RESULTADOS

Con extracción del Componente de dificultad				
	C.Dif.	**C1**	**C2**	**C3**
e1	–.396	.360	–.010	.670
e2	.516	.560	–.090	.300
e3	.285	.520	–.320	–.250
e4	–.123	.580	.230	–.460
e5	–.591	.520	–.130	.090
e6	.142	.400	.810	.120
e7	.193	.620	–.290	.040
e8	–.287	.570	.250	–.170
e9	–.086	.690	–.100	.000
e10	.347	.590	–.100	–.060
r.1.	1.111	3.010	1.000	.870

Sin extracción del componente de dificultad				
	C1	**C2**	**C3**	**C4**
e1	.340	.560	–.430	.050
e2	.580	–.320	–.110	.540
e3	.530	–.380	–.090	–.480
e4	.580	.080	.370	–.400
e5	.500	.410	–.300	.090
e6	.400	.260	.690	.330
e7	.630	–.340	–.190	.170
e8	.550	.400	.130	–.130
e9	.680	.060	–.120	–.130
e10	.600	–.360	.060	.040
r.1.	3.010	1.210	.980	.860

Tabla 1: Matrices de saturaciones (4 componentes)

En la inspección visual de las matrices de saturaciones puede observarse que el primer componente es similar en ambas soluciones; que el segundo componente de la matriz reducida es similar al tercero de la matriz sin reducir y que el tercero de la matriz reducida es similar al cuarto de la matriz sin reducir.

Observése tambien que el componente de dificultad guarda una notable semejanza con el segundo componente de la matriz sin reducir aunque con los signos invertidos.

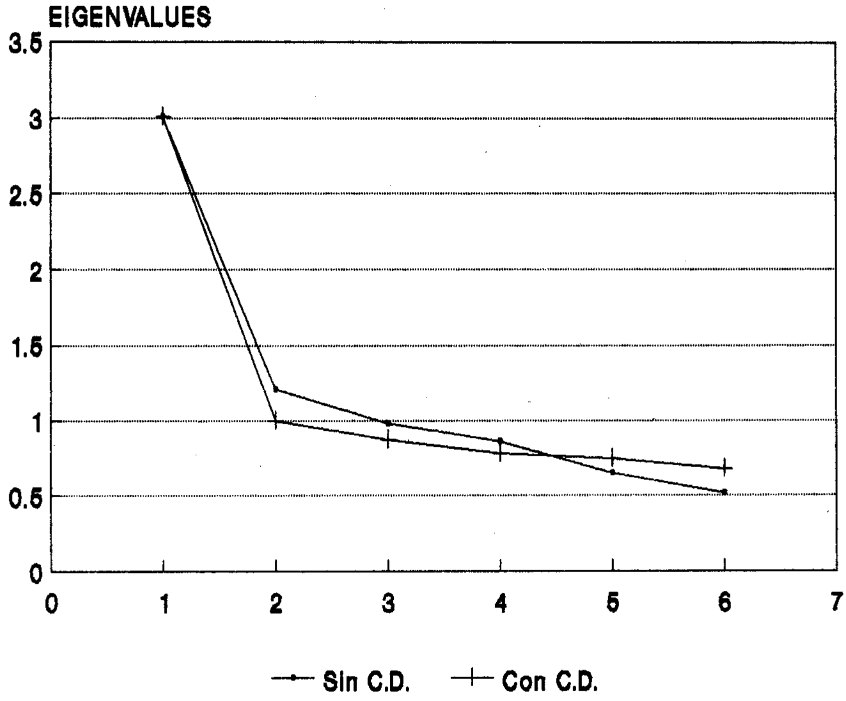

Figura 1: Scree–plots

Si bien en ambos casos las soluciones parecen indicar unidimensionalidad, la solución proporcionada por el análisis tras la extracción del componente de dificultad se ajusta mejor al criterio propuesto. La caida desde la primera raiz es mayor y los valores de las restantes raíces son mas similares entre sí.

Comparación del componente de dificultad con el segundo componente de la matriz original

$$CC = 0.70$$

Obsérvese que los coeficientes de congruencia apoyan con bastante claridad las observaciones realizadas a partir de la inspección de las matrices de saturaciones.

	CD1	C1	CD2	C2	CD3	C3
CD1						
C1	.99**					
CD2	–.45	–.45`				
C2	–.51	–.62	.49			
CD3	–.55	–.55	–.03	.33		
C3	–.09	–.05	.83*	.00	–.48	
C4	–.22	–.17	.20	.00	.66	–.01

CD(i) = Componente con extracción del componente de dificultad.
C(i) = Componente sin extracción del componente de dificultad.
* sig < .01 ** sig < .001

Tabla II: Comparación de las soluciones: Matriz de coeficientes de congruencia.

DISCUSION

En general los resultados del análisis coinciden en señalar que el método de Lawley y Maxwell parece resultar apropiado en este caso. En primer lugar la solución para evaluar la dimensionalidad tiende a ser un poco más clara. Sin embargo, en nuestra opinión, el resultado más interesante es el de la similitud entre uno de los componentes extraido de la matriz original y el componente de dificultad extraido según el método propuesto. Esta similitud, unida a la semejanza que presentan los restantes componentes, permite suponer que el análisis aplicado rutinariamente a la matriz original hubiese dado lugar a un componente espurio, concretamente el segundo componente, difícilmente interpretable desde un modelo teórico.

A pesar de que los resultados obtenidos en este caso parecen interesantes, no podemos evaluar la utilidad general del método a partir de este trabajo por una serie de razones. En primer lugar, en este caso los datos se ajustaban bien a los supuestos del modelo, es decir, se disponía de un grupo reducido de items que habían sido ya seleccionados anteriormente como unidimensionales mediante análisis factorial y que, además presentaban en algunos casos notables variaciones en los valores p. Habría que ver si el método resulta apropiado (o útil) en caso de que los valores medios estén más igualados.

En segundo lugar, el método tiene, a nuestro juicio, una debilidad importante y es la de que, los valores de p dependen de las características de la muestra a la que se aplica la escala. Esto quiere decir que el componente de dificultad no es estable sino que fluctuará en análisis llevados a cabo en distintas muestras en función de la proporción de sujetos que respondan al item en un determinado sentido.

Por último, debería estudiarse la utilidad del método en relación a otras propuestas alternativas. Concretamente, nos referimos al análisis en componentes principales de la matriz de correlaciones tetracóricas inter–item. Actualmente estamos trabajando también en esta línea por lo que creemos que sería interesante comparar las soluciones obtenidas en ambos casos. Por todo ello, consideramos que los resultados únicamente indican que el método puede resultar interesante a pesar de su antiguedad y que merece la pena seguir investigando en este sentido.

REFERENCIAS

Aguilar, A, Tous, J.M. y Andrés, A. (1990) Adaptación y estudio psicométrico del EPQ–R *Anuario de Psicología* 46, 3, 101–119

Bernstein, I.H. (1988) *Applied Multivariate analysis*. Heidelberg: Springer–Verlag.

Cattell, R.B. (1966) The scree–test of the numbers of significant factors *Multivariate behavioral research* 1, 140–161

Comrey, A.L. (1985) *Manual de análisis factorial*. Madrid: Cátedra

Cureton, E.E. (1959) Note on ∅/∅ max: *Psychometrika,* 24, 1, 89–91

Eysenck, S.B.G, Eysenck, H.J. y Barret, P. (1985) A revised version of the Psychoticism scale. *Personality and individual differences* , 6, 1, 21–30

Fraser, C. y McDonald, R.P. (1988) NOHARM: Least squares item factor analysis. *Multivariate behavioral research* 23, 267 – 269

Guilford, J.P. (1950). *Fundamental statistics in psychology and education.* New York: McGraw–Hill

Hambleton, R.K. y Swaminathan, H. (1985) *Item response theory: principles and applications*. Boston: Kluwer–Nijhoff

Hambleton, R.K. y Rovinelli, R.J. (1986) Assesing the dimensionality of a set of tests items *dings of the royal society of Edinburgh* 61, 273–287.

Hattie, J. (1984) An empirical study of various indices for determinig unidimensionality. *Multivariete behavioral research* 49–78.

Hattie, J. (1985) Methodology review: assesing unidimensionality of tests and items. *Applied psychological measurement*, 9, 2, 139–164.

Lawley, D.N. (1943) On problems connected with item selection and test construction. Proceedings of the royal society of Edinburgh, 61, 273–287.

Lawley, D.N. (1944) The factorial analysis of multiple items tests. *Proceedings of the royal society of Edinburgh*, 62, 74–82

Lord, F.M. (1980) *Applications of item response theory to practical testing problems. Hillsdale:* L.E.A.

Lord, F.M. y Novick, M.R. (1968) *Statistical theories of mental tests scores*. Massachusetts:. Addison–Wesley

Luengo, M.A. (1986) La dimensión de extraversión en el EPI y el EPQ: un estudio comparativo *Revista de Psicología General y Aplicada*, 41, 3, 463–487

Maxwell, A.E. (1977) *Multivariate analysis in behavioural research*. London:. Chapman and Hall

McDonald, R.P. (1967) Nonlinear factor analysis *Psychometric Monograph* No 15

McDonald, R.P. (1981) The dimensionality of tests and items. *British Journal of mathematical and statistical psychology*, 34, 100–117

McDonald, R.P. (1985) *Factor analysis and related methods*. Hillsdale: L.E.A.

Rocklin, T. y Revelle, W. (1981) The measurement of extraversion: A comparison of the Eysenck personality inventory and the Eysenck personality questionnaire *British Journal of social psichology*, 20, 279–28

Tucker, L.R. (1951) A method of syntesis of factor analysis studies. *Personel research section* report No. 984 Washington, D.C. Department of the Army