Papeles del Psicólogo
ISSN: 0214-7823
papeles@correo.cop.es
Consejo General de Colegios Oficiales de Psicólogos
España

Ferrando, Pere Joan; Anguiano-Carrasco, Cristina
EL ANÁLISIS FACTORIAL COMO TÉCNICA DE INVESTIGACIÓN EN PSICOLOGÍA
Papeles del Psicólogo, vol. 31, núm. 1, enero-abril, 2010, pp. 18-33
Consejo General de Colegios Oficiales de Psicólogos
Madrid, España

Disponible en: http://www.redalyc.org/articulo.oa?id=77812441003

Cómo citar el artículo
Número completo
Más información del artículo
Página de la revista en redalyc.org

Sistema de Información Científica
Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Papeles del Psicólogo, 2010. Vol. 31(1), pp. 18-33
http://www.cop.es/papeles

EL ANÁLISIS FACTORIAL COMO TÉCNICA DE INVESTIGACIÓN EN PSICOLOGÍA

Pere Joan Ferrando y Cristina Anguiano-Carrasco
Centro de investigación para la evaluación y medida de la conducta. Universidad 'Rovira i Virgili'

El presente texto explica los principales aspectos del análisis factorial como instrumento de investigación psicológica. Se revisan en primer lugar los aspectos básicos a nivel conceptual, de modo que su lectura sea adecuada tanto para el lector principiante como para aquellos que quieran profundizar más en sus conocimientos de la técnica. Después, se discuten con cierto detalle las diferencias entre el análisis exploratorio y confirmatorio, y los procedimientos para estimar el modelo y obtener la solución transformada. Estos puntos se discuten siguiendo cada una de las etapas recomendadas en una investigación: desde el diseño y recogida de datos hasta la interpretación de la solución final. Finalmente, se explica el funcionamiento del programa Factor, un programa más completo que la mayor parte de los programas comerciales, y que además es de distribución libre. Adicionalmente, se propone al lector la realización de un ejercicio resuelto, a modo de práctica para la aplicación de lo expuesto en el texto.
Palabras clave: *Análisis factorial exploratorio, Análisis factorial confirmatorio, Componentes principales, Programa de ordenador FACTOR*

The present text explains the main aspects of factor analysis as a tool in psychological research. First, the basic issues are revised at a conceptual level, so that the review is appropriate for beginners as well as for those who want an in deep knowledge of the technique. Next, the differences between exploratory and confirmatory analysis as well as the procedures for fitting the model and transforming the initial solution are discussed in detail. These issues are discussed according to the recommended steps in a factor-analytic research: from the design and data collection to the interpretation of the final solution. Finally, the functioning of the "Factor" program is explained. Factor is more complete than most commercial programs, and is also freely distributed. Additionally, a solved exercise is proposed as a practice to apply the material discussed in the text.
Key words: *Exploratory factor analysis, Confirmatory factor analysis, Principal components, FACTOR computer program*

Nacido con el siglo XX, el análisis factorial (AF) se ha desarrollado considerablemente a lo largo de sus más de 100 años de existencia. El sencillo modelo inicial propuesto por Spearman (1904) para validar su teoría de la inteligencia ha dado lugar a una amplia familia de modelos que se utilizan no sólo en ciencias sociales, sino también en otros dominios como Biología o Economía. Dado que un tratamiento completo del AF excedería con mucho las posibilidades de este artículo, conviene delimitar primero qué temas se van a tratar.

Desde hace años el primer autor revisa trabajos empíricos en los que se emplea el AF en la investigación psicológica, y la experiencia adquirida servirá para establecer las primeras delimitaciones. En primer lugar, la mayor parte de los estudios factoriales en psicología utilizan el AF para evaluar (a) la estructura de un test a partir de las puntuaciones en sus ítems, o (b) hipótesis de tipo dimensional utilizando como medidas puntuaciones en diferentes tests. Parece razonable, por tanto, centrar la exposición en este tipo de medidas: puntuaciones en ítems o tests.

En segundo lugar, la experiencia indica que los problemas metodológicos en estos estudios son casi siempre los mismos. Un primer grupo de problemas surge en la etapa del diseño de la investigación (etapa generalmente descuidada en los estudios factoriales). Los problemas del segundo grupo se refieren a las decisiones que debe tomar el investigador en la etapa de estimación y ajuste del modelo y en la de rotación. En particular, la mayor parte de los problemas se deben al empleo injustificado del "pack" conocido como "Little Jiffy": *Componentes principales - valores propios mayores que uno - rotación Varimax*. Dedicaremos especial atención al diseño y a la estimación y ajuste del modelo.

Aún con estas delimitaciones, el tema sigue siendo demasiado amplio. En este artículo nos centraremos tan sólo en el modelo general más básico de AF: el modelo lineal, basado en correlaciones, y que analiza medidas obtenidas en un solo grupo de sujetos y en una sola ocasión. Esta limitación deja fuera temas de gran interés: los modelos extendidos de medias y covarianzas, los mode-

Correspondencia: Pere Joan Ferrando. *Universidad 'Rovira i Virgili'. Facultad de Psicologia. Carretera Valls s/n. 43007 Tarragona. España. E-mail: perejoan.ferrando@.urv.cat*

los no-lineales y sus relaciones con la teoría de respuesta al ítem, y los modelos para grupos múltiples y múltiples ocasiones. Tampoco podremos entrar en el tema general de las puntuaciones factoriales.

El enfoque del artículo es conceptual y aplicado, y se ha tratado de reducir el formalismo al máximo. Sólo se incluyen las ecuaciones básicas del modelo y se presentan en cuadros de texto aparte. Asimismo, se ha tratado de reducir al mínimo el número de referencias, y, en cambio, se ha propuesto un apartado de lecturas recomendadas. En este sentido, debemos advertir que algunos de los tópicos discutidos son controvertidos, y que las recomendaciones reflejan la posición de los autores. El apartado de lecturas puede servir para que el lector vea otras posiciones y evalúe críticamente lo que le decimos aquí.

LAS IDEAS BÁSICAS DEL ANÁLISIS FACTORIAL

El AF es un modelo estadístico que representa las relaciones entre un conjunto de variables. Plantea que estas relaciones pueden explicarse a partir de una serie de variables no observables (latentes) denominadas factores, siendo el número de factores substancialmente menor que el de variables. El modelo se obtiene directamente como extensión de algunas de las ideas básicas de los modelos de regresión lineal y de correlación parcial. Del primer modelo se derivan las ecuaciones fundamentales del AF. Del segundo se derivan las ideas clave para evaluar el ajuste del modelo a los datos.

En el modelo de regresión lineal, la puntuación en una variable criterio puede aproximarse, o explicarse en parte, mediante una combinación lineal (una suma de variables multiplicadas cada una de ellas por un peso o coeficiente) de una serie de variables predictoras o explicativas denominadas regresores. Se asume explícitamente que la combinación es tan sólo una aproximación, y que una parte de la puntuación del criterio no podrá ser predicha o explicada desde los regresores. Esta parte no explicada es el término de error (véase ec. 1 en el cuadro).

$$Y = a + b_1x_1 + b_2x_2 + \ldots + b_mx_m + e \qquad (1)$$

$$X_j = \mu_j + \lambda_{j1}f_1 + \lambda_{j2}f_2 + \ldots + \lambda_{jm}f_m + e_j \qquad (2)$$

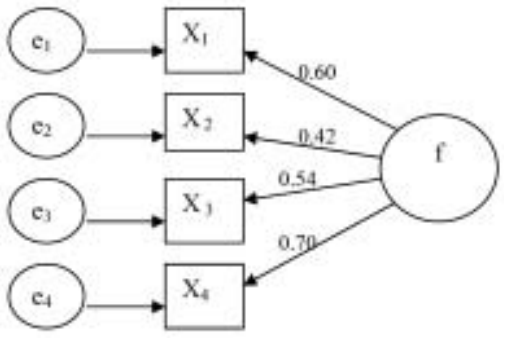
$$X_j = \mu_j + \lambda_j f + e_j \qquad (3)$$

En el AF se analiza un conjunto de variables observables (ítems, subtests o tests) cada una de las cuales puede considerarse como un criterio. Así entendido, el AF consiste en un sistema de ecuaciones de regresión como la descrita arriba (una ecuación para cada variable observable) en el que los regresores, denominados aquí factores, son comunes para un subconjunto (factores comunes) o todo el conjunto (factores generales) de variables (véase ec. 2 en el cuadro). Para cada una de estas ecuaciones la diferencia básica entre el AF y una regresión convencional es que los regresores, es decir los factores, no son observables. Esta diferencia es la que hace que el AF sea un modelo más complejo que el de regresión. Para empezar, al ser los factores no observables, carecen de una escala de medida o métrica determinada. Para resolver esta indeterminación, la práctica más simple, que seguiremos aquí, consiste en asumir que los factores están en escala típica: media cero y varianza uno. Si, además, las variables observables también están en escala típica, el modelo es más simple matemáticamente y más fácilmente interpretable.

Por analogía con el modelo de regresión, se sigue que el modelo AF más sencillo es aquel que plantea un solo factor general (ec. 3). Este modelo sería equivalente al de regresión simple, y fue el modelo AF inicial que planteó Spearman. Para empezar a fijar ideas vamos a estudiar una solución de Spearman basada en el AF de un conjunto de 4 ítems, junto con su representación en un diagrama de Wright. Para la elaboración de estos diagramas, el lector puede consultar el artículo de Ruíz, Pardo y San Martín en el presente volumen.

Ítem	f_i
X_1	0.60
X_2	0.42
X_3	0.54
X_4	0.70

Representado gráficamente sería:



La ecuación AF para el ítem 1 es:

$$X_{i1} = 0.6\, f_i + e_{i1}$$

Que se interpreta como sigue. La puntuación del individuo *i* en el ítem 1 viene en parte determinada por el efecto del factor común (el nivel de *i* en el factor común *f*) y en parte es error. En AF el término de error incluye todos aquellos efectos distintos al factor o factores comunes que influyen en la puntuación en la variable observada. De forma más sistemática, podemos distinguir tres grandes grupos de efectos o tipos de error: (a) el error de muestreo (error estadístico), (b) el error de medida (error psicométrico) y (c) el error de aproximación. Este último componente de error significa que el modelo especificado no se considera exactamente correcto ni siquiera en la población. En efecto, los modelos no son, ni pretenden serlo, exactos. Son, en el mejor de los casos, aproximaciones razonables a la realidad.

El valor 0.6 es el peso factorial y equivale a la pendiente en el modelo de regresión. Si las puntuaciones del ítem y del factor están en escala típica, este peso refleja la importancia que tiene el factor en la determinación de la puntuación en este ítem. Cuanto mayor el peso, mayor la importancia del factor, y, por tanto, menor la influencia del error. Además, en la escala típica asumida, este peso puede interpretarse como la correlación entre el factor y el ítem. Su cuadrado, que es el coeficiente de determinación, se interpreta como la proporción de varianza en las puntuaciones de este ítem que puede explicarse desde el factor. Así pues, de acuerdo con nuestra solución estimada, el ítem 1 correlacionaría 0.6 con el factor general; dicho factor explicaría el 36% de la varianza de las puntuaciones en este ítem (0.6^2=0.36) y, por tanto, el 64% restante sería varianza de error. En la terminología AF la proporción de varianza explicada recibe el nombre de "comunalidad".

Antes de seguir adelante, podría ser de interés como ejercicio que el lector interpretara la ecuación AF correspondiente a otro ítem, digamos el ítem 2. Asimismo, es muy importante tener en cuenta que las interpretaciones anteriores sólo son válidas si variables y factor están en escala típica. De no ser así, el peso no tiene una interpretación clara, ya que refleja entonces en mayor o menor grado las diferencias en la escala de medida de las variables. Al interpretar el output de un AF, es pues esencial asegurarse de que la solución que se interpreta es una solución tipificada.

Cuando se plantea más de un factor, tenemos el modelo AF múltiple, conocido también como modelo de Thurstone (1947), su principal impulsor. El modelo es el mismo para cualquier número de factores y, por simplicidad, lo explicaremos con dos. El aspecto clave aquí es la relación que se plantea entre los factores. Al igual que en regresión, el caso más simple y fácilmente interpretable es aquel en que los factores están incorrelados (conceptualmente, que son independientes entre sí). Una solución de este tipo se denomina "solución ortogonal". En una solución ortogonal los pesos factoriales siguen siendo interpretables como correlaciones variable-factor, sus cuadrados son proporciones de varianza explicada por el correspondiente factor y la suma de estos cuadrados es la comunalidad (véanse ecs. 4 y 5) o la proporción de varianza que explican conjuntamente los factores

.

$$\sigma_j^2 = \lambda_{j1}^2 + \lambda_{j2}^2 + d_{ij}^2 = 1 \qquad (4)$$

$$1 = h_j^2 + \sigma_{ij}^2 \qquad (5)$$

$$r_{jk} = \lambda_{j1}\lambda_{k1} + \lambda_{j2}\lambda_{k2} \qquad (6)$$

El caso de factores correlacionados, denominado 'solución oblicua' es el más complejo, pero posiblemente también el más realista en la práctica. El aspecto más importante a la hora de interpretar una solución de este tipo es que ahora los pesos y las correlaciones variable-factor son coeficientes distintos. Al igual que en teoría de la regresión lineal, los pesos factoriales son ahora coeficientes de regresión estandarizados y miden el efecto del factor sobre la variable de respuesta cuando los demás factores permanecen constantes. Estos pesos se presentan en la matriz denominada "patrón factorial". Por otra parte, las correlaciones variable-factor se denominan coeficientes estructurales, y se presentan en la matriz denominada "estructura factorial". Las ecuaciones correspondientes a los pesos y coeficientes estructurales se presentan en el cuadro que sigue abajo. Conceptualmente, los pesos indican hasta qué punto influye el factor en la variable, en tanto que los coeficientes estructurales indican hasta qué punto se parecen el factor y la variable. En el caso de soluciones oblicuas, en este capítulo nos centraremos sobre todo en los pesos, es decir la matriz patrón.

$$X_{ij} = \lambda_{j1} f_{i1} + \lambda_{j2} f_{i2} + \varepsilon_{ij} \qquad (7)$$

$$s_{j1} = \lambda_{j1} + \lambda_{j2}\phi_{12} \qquad (8)$$

$$\sigma_j^2 = \lambda_{j1}^2 + \lambda_{j2}^2 + 2\lambda_{j1}\lambda_{j2}\phi_{12} + \sigma_{ij}^2 \qquad (9)$$

$$r_{jk} = \lambda_{j1}\lambda_{k1} + \lambda_{j2}\lambda_{k2} + \phi_{12}\left(\lambda_{j1}\lambda_{k2} + \lambda_{j2}\lambda_{k1}\right) \qquad (10)$$

En la ecuación (8) s_{j1} es el coeficiente estructural (correlación variable-factor) y ø es la correlación entre factores.

Veamos ahora un ejemplo de solución ortogonal múltiple con 6 ítems y dos factores:

Item	F_1	F_2
X_1	0.35	**0.82**
X_2	0.31	**0.85**
X_3	0.37	**0.80**
X_4	**0.79**	0.36
X_5	**0.77**	0.32
X_6	**0.79**	0.39

La representación gráfica es:

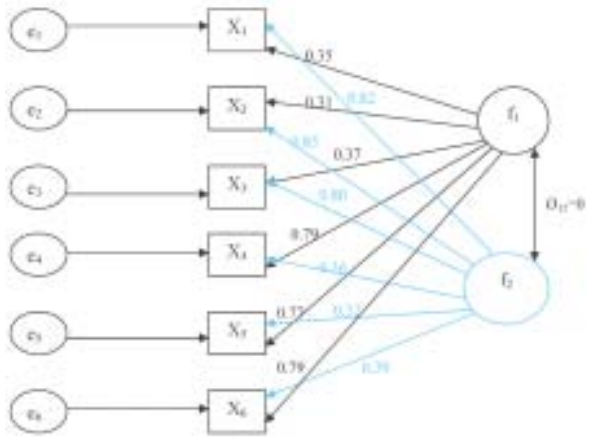


Diagrama de Wright para dos factores comunes incorrelados.

Si tomamos como ejemplo el ítem 2, obtenemos la siguiente ecuación:

$$X_{i2} = 0.31 f_1 + 0.85\, f_{i2} + e_{i2}$$

Esta ecuación nos indica que la puntuación del sujeto *i* en el ítem 2, además de por el error, viene determinada principalmente por el segundo factor (0.85), y, en menor medida, por el primero (0.31). En este caso para explicar la comunalidad (o varianza explicada) se sumarían los cuadrados de los pesos en ambos factores de modo que para el ítem 2,

$$h^2_2 = 0.31^2 + 0.85^2 = 0.82$$

obteniendo la parte de la varianza explicada por ambos factores. Restándola de uno obtenemos que la varianza de error es 0.18. En términos de proporciones, entre los dos factores explican un 82% de la varianza total (comunalidad) y el 18% restante sería error. Para el lector que lo desee sería interesante que realizara la interpretación de otro ítem, por ejemplo el 5.

A continuación mostramos una solución oblicua (patrón) obtenida con los mismos datos. La correlación estimada entre factores fue de 0.40. Con respecto a la solución ortogonal se aprecia que es más clara y simple (los pesos menores son ahora más cercanos a cero). Esto es lo que se observa generalmente al comparar ambos tipos de soluciones.

Ítem	F_1	f_2
X_1	0.23	**0.78**
X_2	0.18	**0.82**
X_3	0.25	**0.75**
X_4	**0.78**	0.18
X_5	**0.77**	0.14
X_6	**0.78**	0.21

El diagrama correspondiente sería ahora:

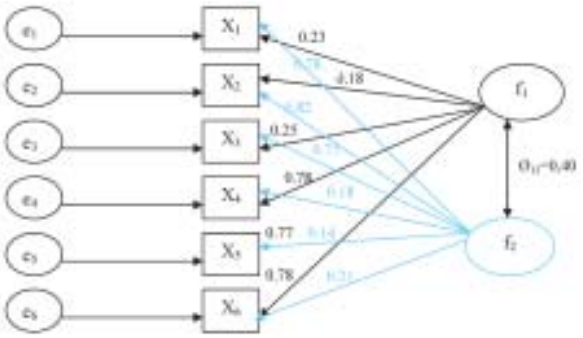


Diagrama de Wright para dos factores comunes correlados.

En este ejemplo vamos a trabajar con el ítem 4. Su ecuación básica tiene la misma forma que en el caso ortogonal:

$$X_{i4} = 0.78\, f_{i1} + 0.18\, f_{i2} + e_{i4}$$

Pero su interpretación es distinta, ya que los pesos (efecto del factor sobre la variable) y las correlaciones variable-factor son medidas distintas. Así el peso correspondiente al primer factor es 0.78. Sin embargo, la correlación entre la variable y este factor, es decir, el coeficiente estructural debe obtenerse como:

$$s_{j1} = 0.78 + 0.18 \cdot 0.40 = 0.85$$

El lector que desee practicar puede realizar los cálculos correspondientes al ítem 1.

Pasamos ahora a las aportaciones del modelo de correlación parcial. La ecuación básica se presenta en el cuadro siguiente para el modelo de Spearman y se basa en los resultados previos discutidos arriba.

$$r_{jk.f} = \frac{r_{jk} - r_{jf} r_{kf}}{\sqrt{1 - r_{kf}^2}\sqrt{1 - r_{jf}^2}} = \frac{\lambda_j \lambda_k - \lambda_j \lambda_k}{\sqrt{1 - \lambda_k^2}\sqrt{1 - \lambda_j^2}} = 0 \qquad (11)$$

La ecuación (11) indica que, si el modelo es correcto, la correlación parcial entre cualquier par de variables después de eliminar de ambas la influencia del factor general es cero. El numerador de la correlación parcial es la diferencia entre la correlación observada entre las dos variables y la correlación reproducida desde el modelo y recibe el nombre de correlación residual. Así, si el modelo es correcto, la correlación observada y la reproducida son iguales y el residual cero. Conceptualmente este resultado se interpreta como que lo único que tienen en común las variables es el factor general que miden, por lo que, al eliminar esta causa común ya no hay nada que las relacione. El caso múltiple es más complejo pero la idea esencial es la misma. Si el modelo es correcto, y las variables tienen solo *m* factores en común, entonces la correlación parcial entre cualquier par de variables tras eliminar la influencia de estos factores comunes debe ser cero. Más que un resultado, esto es un principio de importancia básica. Sugiere que la vía más directa para evaluar si el modelo AF es apropiado deberá basarse en la evaluación de los residuales tras estimar el número propuesto de factores.

Globalmente, la mayor parte de las características del modelo expuesto hasta ahora vienen dictadas por el principio de parsimonia (Carroll, 1978). La parsimonia dicta que las ecuaciones del modelo sean lineales y por tanto lo más simples posible. Este principio recomienda también establecer una distinción clara entre varianza común (comunalidad) y varianza de error. Finalmente, el principio de parsimonia sugiere que el número de factores comunes debe ser considerablemente menor que el de variables. No se ganaría nada, en efecto, interpretando una solución con tantos factores como variables. La determinación del número de factores correcto (que sea lo bastante reducido como para ser claramente interpretable y lo bastante completo como para dar cuenta de las relaciones entre variables) es, posiblemente, la decisión mas importante del AF (Thurstone, 1947).

ANÁLISIS FACTORIAL EXPLORATORIO Y ANÁLISIS FACTORIAL CONFIRMATORIO

En la literatura (e.g. Mulaik, 1972) se distinguen de forma muy marcada dos tipos de análisis factorial: el análisis factorial exploratorio (AFE) y el análisis factorial confirmatorio (AFC). En nuestra opinión esta distinción no es tan clara como se presenta en los textos y, además, plantea una serie de problemas. En primer lugar, en la distinción AFE-AFC se mezclan dos conceptos: (a) la finalidad con la que se lleva a cabo el análisis y (b) el modelo que se pone a prueba. En segundo lugar, tanto para (a) como para (b) el AFE y el AFC no son dos categorías cualitativamente distintas sino que son, más bien, los dos polos de un continuo.

Tal y como se entiende tradicionalmente, en un análisis puramente exploratorio el investigador analizaría un conjunto de datos sin tener ninguna hipótesis previa acerca de su estructura, y dejaría que fuesen los resultados del análisis los que le proporcionasen información al respecto. Por otra parte, en un AFC el investigador habría planteado una serie de hipótesis bien especificadas que pondría a prueba evaluando el ajuste de un modelo. Estas hipótesis serían de tres tipos: (a) número de factores, (b) patrón de relaciones entre las variables y los factores, y (c) relaciones entre los factores.

En las primeras tentativas para evaluar un fenómeno nuevo, una postura puramente exploratoria como la que hemos descrito arriba sería aceptable. Sin embargo no parece la más adecuada cuando analizamos un test que hemos desarrollado o adaptado nosotros mismos. En este caso, es razonable suponer que tendremos una serie de hipótesis previas acerca del número de dimensiones que pretende medir el test, de cuales son los ítems que pretenden medir una u otra dimensión, y de si dichas dimensiones son o no independientes según la teoría. Posiblemente, sin embargo, estas hipótesis no sean aún lo suficientemente fuertes como para especificar un modelo AFC tal como se plantea habitualmente. Así pues, en cuanto a la finalidad, es útil considerar que la mayor parte de las aplicaciones psicométricas del AF se encontrarán en algún punto intermedio.

En cuanto al tipo de modelo que se pone a prueba, las distinciones se refieren aquí al grado de restricción en la solución propuesta. En un AFE las restricciones impues-

tas son las mínimas que se necesitan para obtener una solución inicial, solución que luego puede ser transformada. En un AFC las restricciones son mucho más fuertes y permiten poner a prueba una solución única que no es susceptible de posterior transformación. La finalidad con que se lleva a cabo el análisis y el tipo de modelo que se pone a prueba no son conceptos independientes. Cuanto mayor información previa se tenga y más fuertes sean las hipótesis, más especificada estará la solución puesta a prueba y mayor será el número de restricciones impuestas a dicha solución. Sin embargo, aún admitiendo esta clara relación, la distinción entre AF restricto y AF no restricto nos parecería más apropiada que la de AFE y AFC para referirnos al tipo de modelo que se pone a prueba.

En un AFC, tal como se usa habitualmente, las restricciones acerca del número de factores comunes así como las relaciones entre ellos son similares a las que se plantean en un AFE. Generalmente las correlaciones entre los factores se estiman libremente. Las diferencias principales se refieren a las restricciones que se imponen al patrón factorial. La solución que se propone casi siempre es una solución denominada "de conglomerados independientes" (McDonald, 1985) que sigue el principio de estructura simple. En dicha solución cada variable tiene una carga no nula en un solo factor común siendo cero las cargas en los restantes factores. Una solución de este tipo se presenta a continuación

Ítem	f_1	f_2
X_1	0.0	*
X_2	0.0	*
X_3	0.0	*
X_4	*	0.0
X_5	*	0.0
X_6	*	0.0

donde el asterisco indica que el peso correspondiente se estima como parámetro libre. En esta solución hipotética, los tres primeros ítems serian medidas puras del segundo factor y tendrían pesos nulos en el primero. Por otra parte los tres últimos ítems serían medidas puras del primer factor. Las soluciones de este tipo son teóricamente ideales. Tienen la máxima simplicidad estructural posible y permiten interpretar el contenido de cada factor sin ambigüedades.

Veamos ahora, en contraste, una solución exploratoria obtenida a partir del análisis de estos 6 ítems. Se obtuvo de un AFE en el que se propusieron dos factores comunes. La solución inicial arbitraria se transformó a una solución oblicua ya que, en teoría, los dos factores se consideraban relacionados. Es el patrón que habíamos presentado antes

Ítem	f_1	f_2
X_1	0.23	**0.78**
X_2	0.18	**0.82**
X_3	0.25	**0.75**
X_4	**0.78**	0.18
X_5	**0.77**	0.14
X_6	**0.78**	0.21

Es desde luego bastante clara, y el ajuste del modelo bifactorial fue bueno. Parece bastante evidente que los tres primeros ítems miden principalmente f_2 y los tres últimos f_1. Sin embargo, ¿son lo bastante "limpios" estos ítems como para ajustarse bien a la hipotética solución anterior?

Si evaluamos la adecuación de los datos a la solución hipotética presentada arriba, es decir, ajustamos un AFC convencional a estos ítems, lo que estamos planteando es que cada uno de ellos es una medida pura de un solo factor y, por tanto, que los pesos menores que aparecen en la solución AFE son debidos tan sólo a error de muestreo y son, por tanto, compatibles con valores exactamente de cero en la población.

El problema con este planteamiento es que, en el mundo real, la mayor parte de los ítems (y de los tests) no son medidas factorialmente puras. Con esfuerzo y tras un proceso de selección es posible llegar a obtener algunos ítems que son medidas (casi) puras. Estos ítems se denominan "marcadores" o "indicadores" en el lenguaje del AF. Sin embargo la pretensión que todos los ítems de un test sean marcadores nos parece una hipótesis poco realista.

Si aceptamos la idea de que en la mayor parte de los AF muchos de los ítems son factorialmente complejos, deberemos concluir que la hipótesis estructural más habitual en un AFC es falsa y que, por tanto, el modelo no ajustará bien. Más en detalle, si los pesos menores (estos que están en torno a 0.20 o por debajo) son menores pero no nulos, cada vez que fijamos uno de ellos a cero

cometemos un error de especificación del modelo. Si el modelo tiene pocos ítems, como en el ejemplo, quizás aún podamos llegar a un ajuste aceptable. Sin embargo, en modelos más grandes, la acumulación de los errores llevará necesariamente a ajustes inaceptables. Este razonamiento explica dos resultados que preocupan bastante en el campo aplicado (e.g. McCrae, et al. 1996). El primero es que estructuras factoriales obtenidas mediante AFE que son claras, interpretables y replicables a través de diferentes estudios, muestran ajustes inadmisibles cuando se evalúan mediante AFC. El segundo es que es más fácil obtener malos ajustes cuando se analizan cuestionarios de tamaño realista que cuando se analizan grupos muy reducidos de ítems. El primer resultado puede provocar en el investigador una desconfianza con respecto al AF. El segundo puede llevar a prácticas poco recomendables y a la eliminación innecesaria de ítems.

Nuestra posición puede resumirse así. En el análisis de ítems y tests, creemos que el AF debe venir guiado por la teoría previa. Esta teoría permitirá plantear hipótesis acerca del número de factores, del patrón (aproximado) que se espera encontrar, y de si los factores están o no relacionados. Sin embargo, generalmente, el conocimiento previo no será suficiente para especificar un modelo confirmatorio. Lo que proponemos es utilizar un modelo no restricto (exploratorio) pero con una finalidad confirmatoria hasta donde se pueda. Es decir, estimar una solución en la que se especifique el número de factores (o por lo menos un rango de valores) y también si estos factores son o no independientes. Además, debe tenerse una idea más o menos clara de cómo ha de ser el patrón transformado que se obtendrá. Por supuesto, si la investigación está lo bastante avanzada como para plantear una solución restricta, o todos los ítems son excepcionalmente simples, entonces el AFC es el modelo a usar.

EL DISEÑO DE UNA INVESTIGACIÓN BASADA EN EL ANÁLISIS FACTORIAL

Como en cualquier análisis estadístico, para que los resultados obtenidos mediante AF sean válidos, interpretables y generalizables, se requiere el cumplimiento de algunas condiciones básicas en el diseño de la investigación. Para ver la importancia de este punto en nuestro caso, es útil considerar el AF como un análisis a dos niveles. En el primer nivel se calculan las correlaciones entre una serie de medidas. En el segundo se analiza la estructura de dichas correlaciones. Si los resultados fallan ya en el primer nivel, nunca podrán ser correctos en el segundo. Por razones de claridad, discutiremos separadamente los dos aspectos básicos en el diseño: la muestra y las variables.

MUESTRA

En cualquier estudio factorial, y más aún en aquellos en que se desarrolla o adapta un test, debe tenerse una idea relativamente clara de la población de interés. Por tanto, el AF debería basarse una muestra representativa de esta población. Es muy habitual, sin embargo, utilizar muestras de conveniencia (generalmente estudiantes universitarios). Aparte de la no-representatividad, el problema estadístico más importante aquí es el de atenuación por restricción de rango. Si la muestra es muy homogénea en las variables a analizar (es decir, si las puntuaciones en los ítems/tests tienen poca variabilidad), las correlaciones obtenidas en el primer nivel del AF, estarán atenuadas. La matriz de correlaciones tendrá entonces mucho más "ruido" que "señal" y será difícil obtener una solución clara en el segundo nivel.

Posiblemente, el problema más discutido en AF en relación a la muestra es el de la estabilidad de la solución (¿Cuánta muestra se necesita para que una solución sea estable y generalizable?). Este es un problema complejo. La estabilidad de una solución factorial depende conjuntamente de tres factores: (a) el tamaño de muestra, (b) El grado de determinación de los factores y (c) la comunalidad de las variables. De forma que, si los factores están bien determinados y las variables tienen poco error de medida se podrán alcanzar soluciones estables con relativamente poca muestra. En este sentido queremos advertir que las "recetas" tradicionales tipo: 10 veces más sujetos que variables, etc. no tienen una base sólida.

Las medidas utilizadas habitualmente en psicología: tests y sobre todo ítems, contienen intrínsecamente mucho error de medida. Habrá que aceptar pues que las comunalidades serán generalmente bajas y, por tanto, se deberá actuar principalmente sobre los puntos (a) y (b). Con respecto al punto (b), que se discute con detalle más abajo, la idea de determinación de un factor refiere al número de variables que tienen pesos elevados en dicho factor. De otra forma, hasta qué punto el factor está bien definido y claramente medido por un buen número de indicadores. Con respecto al punto (a) es adecuado de nuevo pensar a "doble nivel". Los resultados del análisis al segundo nivel sólo podrán ser estables si lo son las correlaciones en que se basan. Y las correlaciones

tienen fluctuaciones muestrales altas. Cabe considerar entonces una muestra de 200 observaciones como un mínimo incluso en circunstancias ideales (altas comunalidades y factores bien determinados).

VARIABLES

El AF es un modelo para variables continuas e ilimitadas. Ni las puntuaciones de los ítems ni las de los test lo son. Por tanto, en la mayor parte de las aplicaciones psicológicas el AF deberá verse como un modelo aproximado cuya ventaja es la simplicidad. Es importante pues en primer lugar discutir en qué condiciones la aproximación será lo bastante buena para lo que se requiere en la práctica.

El AF funciona generalmente bien cuando se analizan puntuaciones en tests y subtests. En cuanto a los ítems, la aproximación suele ser también aceptable cuando se usan escalas de respuesta graduada (Likert) con 5 o más categorías. Finalmente, los ítems binarios y los ítems con 3 opciones y una categoría central son potencialmente los que pueden presentar más problemas. En principio recomendaríamos utilizar el formato de respuesta graduada siempre que sea posible.

Sea cual sea el tipo de respuesta, que el AF funcione bien o no depende sobre todo de la distribución de las puntuaciones. Las distribuciones simétricas no suelen dar problemas. Por otra parte los problemas más importantes suceden cuando (a) las distribuciones son marcadamente asimétricas y (b) las asimetrías van en ambas direcciones. Un ejemplo de esta situación sería el análisis de un test que contiene ítems muy fáciles e ítems muy difíciles. Las asimetrías de signo contrario dan lugar a relaciones no lineales y, por tanto, a la inadecuación del modelo AF lineal (Ferrando, 2002). Con relación a lo expuesto arriba, la magnitud del problema depende del tipo de variable a analizar. Con puntuaciones de tests es muy difícil que se produzcan relaciones no lineales. Con ítems de Likert es un problema a tener en cuenta. Finalmente es un problema muy frecuente en ítems binarios, conocido con el nombre de 'factores de dificultad' (McDonald y Alhawat, 1974).

Resultados obtenidos en simulación unidos a la propia experiencia, nos llevan a las siguientes recomendaciones. En el caso de tests y subtests el AF resulta casi siempre apropiado. En el caso de ítems de respuesta graduada, el AF se espera que funcione bien si los coeficientes de asimetría están todos en el intervalo entre -1 y +1. Finalmente, incluso los ítems binarios pueden ajustarse bien por el modelo lineal si los índices de dificultad se mueven entre 0.4 y 0.6. Cuando las variables tienen distribuciones más extremas, es necesario generalmente recurrir a enfoques no lineales que no podemos tratar aquí.

Aparte de la métrica y la distribución, hay otros factores a tener en cuenta en lo referente a las variables, sobre todo cuando se trata de ítems individuales. Como hemos dicho la fiabilidad de los ítems es intrínsecamente baja. Sin embargo, debería evitarse analizar ítems con fiabilidades excesivamente bajas, ya que dichos ítems sólo añadirían ruido a la solución factorial. Un estudio piloto convencional en el que se evalúen los índices de discriminación (correlaciones ítem-total) o las correlaciones test-retest ítem a ítem es muy recomendable. Permite eliminar aquellos ítems que sólo aportan ruido y empezar el AF desde un input más limpio.

En los cuestionarios de rendimiento típico (personalidad, motivación y actitudes), es relativamente frecuente incluir ítems redundantes: aquellos que son esencialmente la misma cuestión redactada quizás en forma ligeramente distinta. Estos ítems se utilizan para evaluar la consistencia de los sujetos o (solapadamente) para incrementar la consistencia interna del test. La presencia de ítems redundantes provoca siempre problemas en el AF. En efecto, los errores entre dos ítems redundantes no pueden ser independientes, ya que, aún después de eliminar los factores comunes, las respuestas siguen estando relacionadas debido a la semejanza de contenidos. La consecuencia es la necesidad de extraer factores adicionales definidos principalmente por parejas o tripletes de ítems redundantes. Estos factores pueden ser difíciles de identificar, sobre todo en soluciones rotadas. Un análisis de contenido previo puede eliminar redundancias y evitar estos problemas desde el principio.

Discutiremos por último el grado de determinación de los factores. Si es posible, una buena recomendación es la de utilizar marcadores o indicadores. Como hemos dicho antes los marcadores son, teóricamente, medidas puras de un factor. En forma más aplicada, Cattell (1988) las define como variables que, en estudios anteriores, han mostrado ser buenas medidas de los factores que se están evaluando. Su uso tiene principalmente dos funciones: (a) permiten identificar los factores aumentando su grado de determinación y (b) permiten relacionar los resultados del estudio con estudios anteriores. Cattell (1988) recomienda utilizar como mínimo dos marcadores por factor.

En cuanto a la relación entre el número de ítems y de factores, como sabemos, cuantos más ítems existan que midan con precisión un factor, más determinado estará dicho factor y más estable será la solución. Aunque existan recomendaciones divergentes (Cattell, 1988) nuestra opinión es que los mejores diseños en AF son aquellos en que se plantean pocos factores, se usan marcadores y se proponen un buen número de ítems para medir cada factor. Tanto si se usan marcadores como si no, para identificar claramente un factor se necesitan un mínimo de 4 variables con pesos substanciales en el mismo.

LAS ETAPAS DE UN ANÁLISIS FACTORIAL

Análisis preliminares: adecuación de los datos

De acuerdo con el planteamiento a doble nivel, parece lógico que antes de emprender un AF se utilicen indicadores para evaluar si las correlaciones obtenidas en el primer nivel son adecuadas para ser analizadas factorialmente en el segundo. Estos indicadores suelen denominarse "medidas de adecuación muestral" y su uso es muy importante como una etapa previa del AF: indicará si el AF es o no el modelo apropiado para los datos. Sin embargo, esta es la etapa que más se pasa por alto en investigación aplicada.

Para empezar, es conveniente inspeccionar los estadísticos descriptivos de las variables de acuerdo con la discusión en la sección anterior. A continuación debería contrastarse el test de esfericidad de Bartlett (1950). Dicho test pone a prueba la hipótesis nula de que la matriz de correlación poblacional es identidad, es decir, que las variables están incorreladas en la población. Si no puede rechazarse dicha hipótesis, habrá que aceptar que la matriz de correlación solo contiene "ruido". Es importante tener en cuenta que, aún así, si se factoriza dicha matriz se obtendrán factores. Sin embargo dichos factores serán totalmente espurios. En este sentido, es útil considerar el test de Bartlett como una prueba de seguridad y una condición necesaria. En la mayor parte de los AF se rechazará la hipótesis nula y, por tanto, se admitirá que existe alguna relación entre las variables. Sin embargo esto puede no ser suficiente. El modelo AF, como hemos visto, asume además que la relación es substancial. Si la relación es tan difusa que se necesitan prácticamente tantos factores como variables para explicarla, entonces no vale la pena llevar a cabo el análisis.

Supuesto que se cumpla la condición necesaria, en tercer lugar se evaluaría el grado de relación conjunta entre las variables. La medida más habitual es el KMO de Kaiser, (1970) que evalúa hasta que punto las puntuaciones en cada una de las variables son predecibles desde las demás. El rango de valores del KMO es de 0 a 1, y, cuanto más alto el valor, más substancialmente relacionadas entre ellas estarán las variables. Como valor de referencia, Kaiser (1970) sugiere que la matriz de correlación será apropiada para factorizar si el KMO es igual o superior a 0.80.

Estimación del modelo

Como hemos avanzado antes, esta es la etapa crucial del AF. En ella se estima una solución inicial y, sobre todo, se determina la dimensionalidad de los datos, es decir el número de factores más apropiado. La etapa de estimación debe guiarse por el principio de parsimonia. Se trata de determinar la solución más simple (es decir el menor número de factores) compatible con residuales suficientemente cercanos a cero.

El procedimiento de estimación implementado por defecto en los programas estadísticos suele ser el análisis en componentes principales (ACP). El ACP, sin embargo, no es un procedimiento para estimar el modelo factorial. Es un método para reducir el número de variables. En esencia, el AF es un modelo basado en el principio de que las variables tienen error de medida, distingue claramente entre varianza común (comunalidad) y varianza de error, y pretende reproducir tan sólo la varianza común, que es la que interviene en las correlaciones entre las variables. El ACP, en cambio, no hace esta distinción, sólo considera la varianza total y es esta varianza total la que pretende reproducir.

Los defensores del ACP argumentan que es más simple, que está mejor determinado y que produce virtualmente los mismos resultados que el AF (e.g. Velicer, 1990). Sin embargo esto último es una verdad a medias. Teóricamente, y desde el punto de vista del AF, el ACP podría considerarse como el caso extremo del modelo factorial en el que todas las variables a analizar están libres de error (es decir, varianza común y varianza total coinciden). En la práctica, el ACP y el AF llevan a resultados similares cuando: (a) el número de variables a analizar es grande (digamos más de 30) y (b) las variables tienen poco error y, por tanto, una elevada comunalidad (Mulaik, 1972). Un principio básico en Psicometría, sin embargo, es que las puntuaciones en los tests tienen error de medida (y las de los ítems mucho más). No parece, por tanto, muy razonable utilizar una técnica que no asume este principio.

El problema de usar ACP cuando el modelo correcto es el AF se ilustra con una pequeña simulación. Se generó

una matriz de correlación a partir de la siguiente solución factorial verdadera

0.50
0.50
0.50
0.50
0.50
0.50
0.50
0.50
0.50
0.50

A continuación la matriz de correlación se analizó mediante un método propiamente factorial y mediante ACP. La solución factorial directa (se especificaron dos factores) fue:

0.50	0.00
0.50	0.00
0.50	0.00
0.50	0.00
0.50	0.00
0.50	0.00
0.50	0.00
0.50	0.00
0.50	0.00
0.50	0.00

que recupera exactamente la solución verdadera. En cambio la solución ACP fue:

0.57	0	-0.82	0	0	0	0	0	0	0
0.57	0.03	0.09	-0.05	-0.78	0.22	0.03	-0.02	-0.01	-0.02
0.57	-0.04	0.09	-0.64	0.01	-0.47	-0.12	0	-0.1	-0.04
0.57	0.4	0.09	0.24	0.01	-0.34	0.44	0	0	0.34
0.57	0.15	0.09	0.18	0.12	-0.02	0.13	-0.15	-0.06	-0.74
0.57	0.06	0.09	0.18	0.12	0.13	-0.33	0.33	-0.59	0.1
0.57	0.09	0.09	0.24	0.01	-0.14	-0.55	0	0.51	0.05
0.57	0.03	0.09	-0.26	0.23	0.41	0.24	0.45	0.32	0
0.57	-0.73	0.09	0.24	0.01	-0.14	0.19	0.01	0.01	0.05
0.57	0	0.09	-0.13	0.23	0.35	-0.02	-0.63	-0.06	0.24

que muestra los dos problemas típicos del ACP: estimaciones sesgadas hacia arriba de los pesos en el factor de contenido y sobreestimación de la dimensionalidad. El primer componente es un estimador sesgado del único factor 'real' (cuyas cargas 'verdaderas' son todas de 0.50). Por otra parte, en los sucesivos componentes algunas de las variables tienen pesos por encima de 0.20-0.30. A este respecto es importante decir que, en la práctica, suele recomendarse interpretar tan sólo los pesos que estén por encima de estos valores mínimos (Catell, 1988, McDonald, 1985). McDonald (1985) propone un criterio heurístico más restrictivo en el que sólo se interpretarían aquellos factores que tuviesen al menos tres variables con pesos superiores a 0.30. Aún siguiendo este criterio más restrictivo, los resultados de la solución ACP llevarían a interpretar como factores 4 componentes que reflejan únicamente error. Si, además, esta solución hubiese sido posteriormente rotada la interpretación hubiera sido, posiblemente, totalmente errónea.

Existen varios métodos recomendables para estimar el modelo AF. Por limitaciones de espacio, discutiremos tan sólo los dos más utilizados: Mínimos Cuadrados Ordinarios (MCO) y Máxima Verosimilitud (MV). Sin embargo, hay otros métodos muy interesantes cuyo posterior estudio recomendamos al lector interesado. En particular vale la pena revisar el AF de rango mínimo (Shapiro y ten Berge, 2002).

El AF por MCO no es, propiamente, un método de estimación, sino un conjunto de métodos basados en un criterio general común. Para el número especificado de factores, los estimadores MCO son aquellos que hacen mínima la suma de cuadrados de las diferencias entre las correlaciones observadas y las reproducidas desde el modelo. Conceptualmente pues, los métodos MCO determinan la solución que hace que los residuales sean lo más cercanos a 0 posible. Esta es, como sabemos, la idea básica del ajuste en AF. Aunque el criterio es muy claro y directo, los métodos MCO son, en principio, puramente descriptivos. Como veremos, sin embargo, esto no es necesariamente una limitación.

Los principales métodos basados en el criterio MCO son: (a) AF de ejes principales, (b) MINRES de Harman (Harman y Jones, 1966), (c) ULS de Jöreskog (1977), y (d) Residual Mínimo de Comrey (1962). Para el mismo número de factores, las soluciones obtenidas con cualquiera de ellos son virtualmente idénticas. Sin embargo, recomendaríamos especialmente usar MINRES o ULS por dos razones: (a) no requieren la estimación inicial de las comunalidades y (b) son muy eficientes en términos de computación.

En contraste con los métodos MCO, el método MV (Lawley y Maxwell, 1971) es estadístico (inferencial). Su principal ventaja es que permite contrastar, de forma rigurosa, el ajuste del modelo a los datos mediante un índice referido a la distribución chi-cuadrado (χ^2). Esta ventaja, sin embargo, debe ser matizada. En primer lugar, la inferencia en AF MV se basa en el supuesto de que las variables que se analizan son continuas, métricas y distribuidas según la ley normal multivariante. En el caso de ítems y tests este supuesto nunca se cumple. En segundo lugar se asume que el modelo propuesto en

m factores ajusta perfectamente en la población y por tanto, que todo el error es error muestral (esta es la hipótesis nula del test de bondad de ajuste). Sin embargo, como hemos visto antes, los modelos se proponen tan sólo como razonables aproximaciones, y se acepta que parte del error será error de aproximación. Así pues, el método contrasta una hipótesis nula que ya sabemos desde el principio que es falsa y que se rechazará siempre en cuanto se tenga la potencia suficiente. Para agravar el tema, la potencia generalmente es muy alta, ya que en AF se trabaja habitualmente con muestras grandes. En suma, aún con distribuciones 'razonables' el uso del AF MV basado en el test de bondad de ajuste llevará casi siempre a la necesidad de estimar más factores de los que son substantivamente interpretables. Este fenómeno se denomina "sobrefactorización".

A pesar de los problemas discutidos arriba, existen razones para recomendar el uso del AF MV. En primer lugar, y aunque es poco conocido, la solución MV puede obtenerse también sin hacer supuestos inferenciales. Es la que hace mínimas las correlaciones parciales entre las variables tras eliminar de ellas la influencia de los factores (Mulaik, 1972). Esencialmente es el mismo criterio básico que el del AF por MCO. Los métodos MCO minimizan las correlaciones residuales. MV minimiza las correlaciones parciales (la correlación residual es el numerador de la correlación parcial). Por esta razón, en la práctica, las soluciones MCO y MV suelen ser muy similares. En segundo lugar, aunque el test χ^2 de bondad de ajuste evalúa una hipótesis falsa, existen indicadores de bondad de ajuste derivados de este test que evalúan el error de aproximación y el grado de ajuste del modelo. Como veremos en la siguiente sección estos indicadores son muy recomendables.

En una situación en que (a) las variables tienen distribuciones aceptables, (b) la solución esta bien determinada, y (c) el modelo propuesto es razonablemente correcto, las soluciones MCO y MV serán prácticamente idénticas. En este caso, el uso de MV tiene la ventaja de que permite obtener indicadores adicionales muy útiles en la evaluación del ajuste. En caso de distribuciones extremas, soluciones débiles o poco claras y notable error de aproximación la opción MV dará problemas. La convergencia del método es muy delicada y puede dar lugar a estimaciones inaceptables. Además, los indicadores adicionales no serán de fiar. En estos casos los métodos MCO son claramente superiores. De acuerdo con los estudios de simulación, son métodos muy robustos (convergen casi siempre) y, al no hacer distinciones entre las fuentes de error, suelen recuperar mejor la solución correcta que el método MV (MacCallum y Tucker 1991).

Evaluación del ajuste

Para decidir si un modelo con *m* factores resulta apropiado, debe evaluarse el grado de ajuste del modelo a los datos. Existen una variedad de criterios y procedimientos para llevar a cabo esta evaluación. En nuestra opinión, algunos son considerablemente mejores que otros.

Empezaremos con los criterios y procedimientos que no recomendamos. Posiblemente, el criterio más utilizado en Psicología, y el que suelen aplicar por defecto los programas comerciales es la regla de Kaiser: el número de factores relevantes es el número de valores propios mayores de 1 que tiene la matriz de correlación original. Este criterio presenta varios problemas, siendo el primero de ellos la falta de una justificación clara. Tiene varias (que no veremos aquí) pero ninguna convincente. El segundo problema es que se basa en la lógica del ACP no del AF. En efecto, los valores propios de la matriz sin reducir (con unos en la diagonal principal) equivalen a las proporciones de varianza total explicadas por los correspondientes componentes principales. Sin embargo, como hemos visto, la varianza que interesa realmente en el AF es la común, no la total. En tercer lugar, el número de factores determinado mediante esta regla está relacionado con el número de variables que se analizan. Más en detalle, si se analizan *n* variables, el criterio de Kaiser indicará un número de factores comprendido entre n/5 y n/3. Así, con 30 variables, el criterio indicará entre 6 y 10 factores. Sin embargo, si hemos diseñado una escala de 30 ítems para medir una sola dimensión, el número de factores esperado es 1, no de 6 a 10.

El test de sedimentación (Scree-test; Cattell, 1988) es un procedimiento gráfico ampliamente utilizado. En un gráfico bivariado, se representan puntos cuyas coordenadas son los valores propios de la matriz de correlación original (es decir, las proporciones de varianza total explicada) en el eje de ordenadas, y el número de componentes en el de abcisas. En una solución típica, el gráfico que une los puntos es una función decreciente, similar en forma a la ladera de una colina de residuos. A partir de cierto punto la función se hace prácticamente horizontal y es este punto el que, según Cattell, indica el número más adecuado de factores. La lógica es que, a partir de este número, los sucesivos factores son triviales y sólo explican varianza residual. Aunque la lógica es más convincente que la de la regla de Kaiser, el procedimiento tiene, en nuestra opinión, dos problemas. En primer lu-

gar la decisión se apoya en una inspección visual y, por tanto, tiene un fuerte componente de subjetividad. En segundo lugar, se basa en la lógica ACP y no distingue entre varianza común y de error. Ahora bien, si en lugar de los valores propios de la matriz sin reducir (proporciones de varianza total), se representasen los de la matriz reducida (es decir, las comunalidades) entonces el test sería útil como procedimiento auxiliar.

Existen dos criterios muy en boga actualmente, que son el MAP de Velicer (1976) y el análisis paralelo (AP; Horn, 1965). En nuestra opinión son de utilidad como criterios auxiliares, pero adolecen del mismo problema básico que tienen los criterios anteriores basados en la lógica ACP: no distinguen entre varianza común y varianza de error.

En el criterio MAP, se lleva a cabo un ACP en forma secuencial y en cada etapa se calcula la raíz media cuadrática de las correlaciones parciales que resultan si se elimina el componente correspondiente y los anteriores. Aunque las correlaciones residuales disminuyen siempre a medida que se estiman más componentes, las correlaciones parciales no lo hacen. De hecho, la función que relaciona la raíz media cuadrática de las parciales con el número de componentes suele tener forma de U. El mínimo de la función indicaría el número de componentes a retener.

El AP puede entenderse como una combinación del criterio de Kaiser y del scree test. En su forma más básica, consiste en generar una matriz de correlación aleatoria a partir de unos datos de la misma dimensión que los empíricos (sujetos y variables): teóricamente una matriz así debería tener todos los valores propios cercanos a 1. El método consiste en comparar los valores propios de la matriz empírica con los de la matriz generada al azar. Gráficamente la comparación sería como un doble scree test en el que se representan simultáneamente la curva correspondiente a los datos empíricos y la correspondiente a los datos al azar. La primera, como sabemos, se esperaría que mostrase un fuerte descenso seguido de estabilización. La segunda debería mostrar una tendencia mucho más plana (en una matriz muy grande sería una recta horizontal que cortaría a la ordenada en 1). El punto de intersección entre las dos curvas indicaría el número de factores a retener. Visto así el AP comparte muchas de las críticas del scree-test, pero tiene la ventaja de que el criterio para determinar el número de factores es mucho más objetivo.

Pasamos ahora a discutir los criterios y procedimientos que consideramos más recomendables. Empezaremos por aquellos de tipo general que pueden aplicarse cualquiera que sea el método de estimación. Después discutiremos aquellos específicamente relacionados con el AF MV.

Como bien sabemos ya, si el número de factores propuesto es apropiado, entonces las correlaciones residuales entre las variables tras eliminar la influencia de los factores deben ser todas prácticamente cero. De acuerdo con este principio, los criterios más claros para evaluar el ajuste de un modelo en *m* factores serán aquellos que más directamente se relacionan con la evaluación de las correlaciones residuales.

En problemas pequeños, la inspección visual de la matriz de residuales puede dar ya una idea importante del grado de ajuste. Sin embargo, habitualmente el AF trabaja con un número substancial de variables, y la inspección global de la matriz residual no resulta práctica. En este caso, deberá condensarse la información mediante estadísticos descriptivos.

Para empezar es conveniente inspeccionar la distribución de frecuencias de los residuales. Si el número de factores propuesto es adecuado, esta distribución será simétrica, aproximadamente normal y centrada en torno a una media de cero. Distribuciones asimétricas, descentradas o con las colas muy pesadas, indican que aún queda covariación sistemática por explicar y, por tanto, que es necesario estimar más factores.

La raíz media cuadrática residual (RMCR) es una medida descriptiva que indica la magnitud media de las correlaciones residuales. Si la media de éstas últimas es cero, entonces la RMCR coincide con la desviación típica de los residuales. Harman (1976), propone un valor de referencia de 0.05 o menor para considerar que el ajuste del modelo era aceptable. Este criterio es puramente empírico, pero funciona generalmente bien en la práctica. Mejor fundamentado es el criterio propuesto inicialmente por Kelley (1935). El error típico de un coeficiente de correlación de cero en la población es, aproximadamente, $1/\sqrt{N}$ donde *N* es el tamaño de muestra. Este sería, por tanto, el valor de referencia. Así, en una muestra de 300 sujetos, $1/\sqrt{N}=0.058$. Si la RMCR se mueve en torno a este valor, o es menor, cabe interpretar que los valores residuales observados no difieren significativamente de cero y, por tanto, que no quedan ya relaciones sistemáticas por explicar.

El índice gamma o GFI propuesto inicialmente por Tanaka y Huba (1985) es una medida de bondad de ajuste normada entre 0 y 1 que puede utilizarse con la mayoría de los procedimientos de estimación. Puede interpretarse como un coeficiente de determinación multivariado que indica la proporción de covariación entre las variables explicada por el modelo propuesto. De acuerdo con los criterios actuales (véase el artículo de Ruíz, Pardo y

San Martín en este volumen) para considerar un ajuste como bueno, el GFI debería estar por encima de 0.95.

Discutiremos por último dos indicadores que se utilizan en el caso de estimación por MV. Son, por tanto, inferenciales, pero aquí recomendamos su uso de acuerdo con una lógica descriptiva. El primero de ellos es el coeficiente TLI-NNFI, propuesto inicialmente por Tucker y Lewis (1973) precisamente para el modelo AF. Es un índice relativo y mide la mejora de ajuste que produce el modelo propuesto con respecto al modelo nulo en 0 factores, en relación a la mejora esperada por un modelo que ajustara bien. Sus valores se mueven entre 0 y 1 (aunque no esté estrictamente normado) y Tucker y Lewis recomendaron interpretarlo como un coeficiente de fiabilidad. En este sentido, y aunque los criterios actuales son más rigurosos (véase Ruíz, Pardo y San Martín en este volumen), nuestra opinión es que valores por encima de 0.85-0.90 empezarían a considerarse aceptables.

El índice RMSEA (Steiger y Lind, 1980), muy en boga actualmente, estima el error de aproximación del modelo propuesto. Más específicamente, estima la discrepancia que habría entre la matriz de correlación poblacional y la matriz reproducida por el modelo propuesto, también en la población. Conceptualmente, el RMSEA se basa en el enfoque, discutido antes, de que los modelos son sólo aproximaciones y estima hasta qué punto el modelo puesto a prueba es una aproximación razonable. El RMSEA es un índice relativo a los grados de libertad (complejidad) del modelo y, por tanto, puede penalizar a los modelos menos parsimoniosos. Como referencia, valores por debajo de 0.05 podrían considerarse como indicadores de buen ajuste, en tanto que valores entre 0.05-0.08 indicarían un ajuste admisible.

Cerraremos esta sección con tres observaciones. En primer lugar, no recomendamos tomar una decisión tan importante como la del número apropiado de factores basándose en un solo indicador o criterio. Es conveniente utilizar múltiples indicadores que nos proporcionen más información y, por tanto, elementos de juicio. En segundo lugar, en situaciones reales el proceso que lleva a la determinación del número de factores no es tan lineal como lo describimos aquí por razones didácticas. En efecto, una vez obtenida la solución transformada podemos encontrarnos con que uno o más factores sean muy débiles, estén pobremente identificados o reflejen contenidos triviales (por ejemplo, que no lleguen al mínimo de 3 variables con pesos superiores a 0.3 antes comentado). Este resultado podría llevar a la reconsideración del número de factores y a una nueva inspección de la solución. Por último, en muchos casos la información teórica previa no indica un número claro de factores sino que es más difusa. Puede indicar un rango plausible de factores o quizás varias soluciones alternativas plausibles. Es conveniente en este caso examinar las diferencias entre los valores de los indicadores de ajuste correspondientes a las diferentes soluciones evaluadas.

Obtención de la solución transformada (rotación)

En el análisis factorial no restricto, la solución inicial estándar que se obtiene mediante la aplicación de los métodos de MCO o MV se denomina “forma canónica”. Para el número de factores especificado, esta solución tiene la propiedad de que los factores sucesivos explican el máximo posible de la varianza común. En algunos casos la solución canónica es directamente interpretable y no requiere ninguna transformación posterior. El caso más claro es cuando se analiza un conjunto de ítems que pretenden medir una sola dimensión. Dado que el primer factor explica el máximo posible de la varianza común, si el conjunto es esencialmente unidimensional, entonces los factores que siguen al primero deberán ser residuales. Un ejemplo extremo es la solución directa en dos factores que hemos usado arriba para comparar el AF con ACP. Es una solución canónica en la que el primer factor explica ya toda la varianza común y al segundo no le queda nada por explicar. Este grupo de variables sería pues perfectamente unidimensional.

Cuando se espera encontrar una solución en múltiples factores, sin embargo, la solución canónica inicial es arbitraria. Debe ser entonces transformada o rotada hasta obtener una solución interpretable de acuerdo con la teoría previa.

La principal decisión que debe tomar el investigador en esta etapa es si utilizará una rotación oblicua o una ortogonal. Este es un tema controvertido. Los autores que defienden las soluciones ortogonales, consideran que son más simples y fáciles de interpretar. Además creen que son también más estables en estudios de replicación. La base estadística de este argumento es que, si los factores son independientes en la población, no lo serán exactamente en las muestras y, por tanto, si se utilizan soluciones oblicuas, las correlaciones entre factores reflejarán tan sólo error muestral Por otra parte, los autores que defienden las soluciones oblicuas consideran que la mayor parte de constructos psicológicos están relacionados y que exigir factores incorrelados es imponer artificialmente una solución que no es correcta tan sólo porque es más sencilla (e.g. Mulaik, 1972). En definitiva, que ha de ser la teoría la que guíe esta decisión.

Los autores nos situamos esencialmente en la segunda postura. Sin embargo, también creemos que debe tenerse en cuenta (como siempre en AF) el criterio de parsimonia. Si la teoría no permite hipótesis fuertes al respecto, parece razonable empezar con una solución oblicua. Si las correlaciones estimadas entre factores son substanciales, esta es la solución a mantener. Sin embargo, si las correlaciones entre factores son consistentemente bajas (digamos por debajo de 0.30 o 0.20), entonces podría hacerse un segundo análisis especificando una solución ortogonal. Si las dos soluciones fuesen similares, sería preferible aceptar provisionalmente la solución ortogonal.

Una vez decidido el tipo general de rotación, la importancia de la decisión respecto al método específico a utilizar depende de la solidez del diseño. Si las variables son 'limpias', se utilizan marcadores y los factores están bien determinados, entonces los diferentes métodos deben llevar esencialmente a la misma solución. No es una mala estrategia probar diferentes métodos y evaluar si convergen aproximadamente a una solución común. En soluciones complejas, ruidosas y pobremente determinadas el empleo de diferentes métodos de rotación puede llevar a soluciones muy dispares. Este no es un problema de los métodos, es un problema de diseño.

Los métodos analíticos de rotación ortogonal son generalmente 'cuárticos' en el sentido de que se basan en las potencias cuartas de las los pesos factoriales (conceptualmente, varianzas de los cuadrados de los pesos). Existen dos métodos generales de este tipo (véase e.g. Mulaik, 1972): Quartimax y Varimax. Dado el patrón inicial no rotado (variables ´ factores), la transformación Quartimax es la que maximiza la varianza de los cuadrados de los pesos por filas. En cambio Varimax maximiza la varianza por columnas. La solución Quartimax es pues compatible con una columna del patrón en el que la mayor parte de los pesos son elevados y tiende a dar soluciones con un factor general. En cambio Varimax tiende a dar soluciones múltiples en las que no hay un factor dominante. Existe un tercer método, Equamax que combina ambos criterios y lleva por tanto a soluciones intermedias. Los tres métodos funcionan bien, y tal vez la elección de uno de ellos debería venir guiada por lo esperado desde la teoría (existencia o no de un factor general). Nuestra consideración positiva de los métodos (Varimax en particular) no es incompatible con la crítica hecha al principio del capítulo. Nuestra crítica inicial se refiere al uso indiscriminado de Varimax por defecto aun cuando la teoría indique claramente que los factores están relacionados o cuando se espera encontrar un factor general.

La experiencia con los métodos anteriores basada en estudios de simulación sugiere que pueden llevar a resultados erróneos o inestables (sobre todo Varimax) cuando una elevada proporción de las variables a analizar son factorialmente complejas. Como hemos dicho arriba, este no es un problema del método de rotación sino de un mal diseño. Para minimizar el problema se han propuesto versiones ponderadas (weighted) de los tres métodos (Quartimax, Varimax y Equamax) en los que se asigna mayor peso a los ítems inicialmente evaluados como más simples.

Los métodos analíticos tradicionales de rotación oblicua son una extensión de los métodos cuárticos ortogonales, con el problema adicional de determinar el grado de relación (oblicuidad) entre los factores. En la práctica, el método más utilizado con diferencia es Oblimin (véase e.g. Mulaik, 1972) cuyo criterio puede ser considerado como una combinación de los criterios Quartimax y Varimax extendida al caso oblicuo. El criterio Oblimin incluye un parámetro (delta, que puede tomar valores entre 0 y 1) y que pondera la maximización de la simplicidad por filas o por columnas. Indirectamente el parámetro delta controla también el grado de oblicuidad permitido en la solución. Los programas factoriales utilizan por defecto el valor delta=0 siguiendo la recomendación de Jennrich (1979). Este valor suele llevar a una buena convergencia y a una solución simple e interpretable. En contrapartida, para conseguir esta simplicidad, los factores resultantes suelen estar muy relacionados entre ellos. Browne (comunicación personal) recomienda utilizar delta=0.5.

Una alternativa interesante a Oblimin son ciertos métodos analíticos que incorporan la idea de rotación sobre una matriz diana que se construye analíticamente. Esencialmente la idea es la siguiente. En primer lugar, a partir de una solución ortogonal se construye una matriz hipótesis o diana. Dicha matriz es una modificación de la solución ortogonal lo más cercana posible a la estructura simple; así, y principalmente, los pesos muy bajos en la solución ortogonal se hipotetizan como ceros en la matriz diana. En segundo lugar se determina la solución transformada oblicua que mejor se aproxima a la matriz diana. El método inicial que incorporaba estas ideas es Promax (Hendrickson y White, 1964). Lorenzo-Seva (1999) ha propuesto recientemente un método mucho más refinado denominado Promin. Con respecto a los métodos previos, Promin incorpora mejoras en todas las etapas: solución ortogonal de partida, determinación de la matriz diana y procedimientos y criterios de aproximación. Es, por tanto, el método más recomendable dentro de esta familia.

SOFTWARE: EL PROGRAMA FACTOR[1]

Si bien es cierto que el AFE es una técnica *clásica* de análisis de datos, hoy en día la investigación estadística sobre el propio análisis continúa siendo muy activa. Así, en los años recientes se han publicado en revistas especializadas diversos avances relacionados con los métodos utilizados en AFE. Sin embargo, los autores de los programas informáticos de análisis de datos más populares (no hace falta citar nombres) no parecen interesados en implementar estos nuevos avances. Por esta razón, los investigadores de las propias universidades se han preocupado por desarrollar programas específicos que recojan tanto los métodos clásicos como las nuevas aportaciones. Un ejemplo de este tipo de programas es Factor (Lorenzo-Seva y Ferrando, 2006). Se trata de un programa fácil de usar (basado en los típicos menús de Windows) que tiene como finalidad el cálculo de AFE.

Factor implementa procedimientos e índices clásicos, así como algunas de las aportaciones metodológicas más recientes. Esto incluye todos los discutidos en este capítulo y además otros de indudable interés y que debieran ser objeto de estudio futuro para los lectores interesados en el AF. Así por ejemplo Factor permite trabajar con correlaciones policóricas cuando se sospecha que el modelo lineal puede ser inadecuado. Buenos ejemplos de la metodología propuesta recientemente que se han implementado en Factor son: (1) Minimum Rank Factor Analysis, que es el único procedimiento de extracción de factores que permite estudiar la proporción de varianza explicada por cada factor; y (2) el método de rotación Simplimax, que ha demostrado ser el método de rotación más eficiente de todos los propuestos hasta el momento. Muchos de estos métodos no están disponibles en ningún programa comercial.

Finalmente, cabe destacar que Factor es un programa que se distribuye gratuitamente como software de libre disposición. Se puede obtener en Internet, en la página: http://psico.fcep.urv.es/utilitats/factor/. En la misma página se ofrece un breve manual de utilización, así como una versión de demostración para ejemplificar su utilización. Hasta la fecha, y desde que se puso a disposición de los investigadores en el año 2006, se ha utilizado por investigadores de las más diversas procedencias en 29 artículos internacionales aparecidos en revistas recogidas en el ISI.

EJEMPLO ILUSTRATIVO

A continuación se propone al lector la realización de un ejercicio práctico para aplicar el material expuesto. Los datos se pueden encontrar en la siguiente dirección: http://psico.fcep.urv.cat/ejemplopapeles. Se trata de un test de 14 ítems que mide dos tipos de ansiedad. Se propone al lector que mediante el programa factor, realice el análisis factorial del test para determinar la estructura y propiedades del mismo. En la misma web se encontrará la resolución de este ejemplo, aunque es recomendable que el lector trabaje sobre el ejemplo antes de leer la resolución.

BIBLIOGRAFÍA RECOMENDADA

- **(Manuales de tipo general)** los de Harman (1976) y Mulaik (1972), ambos en la lista de referencias, tratan bastante a fondo los principales aspectos del AF, incluyendo los que no hemos podido desarrollar. Su nivel técnico es bastante más elevado que el que hemos utilizado aquí.

  El manual de McDonald (1985, en lista de referencias) es mucho más personal que los dos anteriores. Se muestra muy crítico con el enfoque tradicional del AF y refleja las fuertes posiciones del autor. Muy recomendable.

  Por último, el siguiente texto del autor:

  Ferrando, P.J. (1993) Introducción al análisis factorial. Barcelona: PPU.

  Puede usarse como una ampliación más técnica de los aspectos básicos desarrollados en las primeras secciones.
- **La problemática AFE vs AFC en el análisis de ítems se discute a fondo en**:

  Ferrando, P.J. y Lorenzo-Seva, U. (2000).

  "Unrestricted versus restricted factor analysis of multidimensional test items: some aspects of the problem and some suggestions", *Psicológica*. 21, 301-323.
- **La siguiente tesis proporciona una buena revisión de los principales criterios y métodos de rotación**

  Lorenzo-Seva, U. (1997). Elaboración y evaluación de métodos gráficos en análisis factorial. Universidad de Barcelona.
- **Por último, el desarrollo paso a paso de un AF mediante un programa informático se describe en:**

  Ferrando, P.J. y Lorenzo, U. (1998)

  "Análisis factorial".

  En: Renom, J. (Coord.) 'Tratamiento informatizado de datos'. Cap. 5 pp 101-126. Barcelona. Masson.

[1] El autor de esta sección es Urbano Lorenzo-Seva

REFERENCIAS

Bartlett, M.S. (1950). Tests of significance in factor analysis. British Journal of Mathematical and Statistical Psychology, 3, 77-85.

Carroll, J.B. (1978). How shall we study individual differences in cognitive abilities?-Methodological and theoretical perspectives. Intelligence, 2, 87-115.

Cattell, R.B. (1988). The meaning and strategic use of factor analysis. In J.R. Nesselroade and R.B. Cattell (eds.) Handbook of multivariate experimental psychology (pp 131-203). New York: Plenum Press.

Comrey. A.L. (1962). The minimum residual method of factor analysis. Psychological Reports, 11, 15-18.

Ferrando, P.J. (1993) Introducción al análisis factorial. Barcelona: PPU.

Ferrando, P.J. y Lorenzo-Seva, U. (1998). Análisis factorial. En: Renom, J. (Coord.) Tratamiento informatizado de datos. Cap. 5 (pp 101-126). Barcelona: Masson.

Ferrando, P.J. y Lorenzo-Seva, U. (2000). Unrestricted versus restricted factor analysis of multidimensional test items: some aspects of the problem and some suggestions. Psicológica, 21, 301-323.

Ferrando, P.J. (2002). Theoretical and empirical comparisons between two models for continuous item responses. Multivariate Behavioral Research, 37, 521-542.

Harman, H.H. y Jones, W.H. (1966). Factor analysis by minimizing residuals (minres). Psychometrika, 31, 351-368.

Harman, H.H. (1976). Modern factor analysis. Chicago: Univ. of Chicago press.

Hendrickson, A.E. y White, P.O. (1964). Promax: A quick method for rotation to oblique simple structure. British Journal of Statistical Psychology. 17, 65-70.

Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. Psychometrika, 30, 179-185.

Jöreskog, K.G. (1977). Factor analysis by least-squares and maximum-likelihood methods. En: K. Enslein, A. Ralston, & H.S. Wilf (Eds.): Statistical methods for digital computers. (pp 125-153). New York: Wiley.

Kaiser, H.F. (1970). A second generation little jiffy. Psychometrika, 35, 401-415.

Kelley, T.L. (1935). Essential Traits of Mental Life. Harvard Studies in Education, 26 (pp. 146). Cambridge: Harvard University press.

Lawley, D.N. y Maxwell, A.E. (1971). Factor analysis as a statistical method. London: Butterworths.

Lorenzo-Seva, U. (1997). Elaboración y Evaluación de Métodos Gráficos en Análisis Factorial. Tesis doctoral. Universidad de Barcelona.

Lorenzo-Seva, U. (1999). Promin: A Method for Oblique Factor Rotation. Multivariate Behavioral Research, 34, 347-365.

Lorenzo-Seva, U. y Ferrando P.J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. Behavior Reserch Methods, 38, 88-91.

MacCallum, R.C. y Tucker, L.R. (1991). Representing sources of error in the common factor model: implications for theory and practice. Psychological Bulletin, 109, 502 511.

McCrae, R.R., Zonderman, A.B., Costa, P.T., Bond, M.H. y Paunonen, S.V. (1996). Evaluating replicability of factors in the revised NEO personality inventory: confirmatory factor analysis versus Procustres rotation. Journal of Personality and Social Psychology, 70, 552-566.

McDonald, R.P. (1985). Factor analysis and related methods. Hillsdale: LEA

McDonald, R.P. and Ahlawat, K.S. (1974). Difficulty factors in binary data. British Journal of Mathematical and Statistical Psychology, 27, 82-99.

Mulaik, S.A. (1972). The foundations of factor analysis. New York: McGraw-Hill.

Shapiro, A. y ten Berge, J.M.F. (2002). Statistical inference of minimum rank factor análisis. Psychometrika, 67, 79-94.

Spearman, Ch. (1904). General intelligence; objectively determined and measured. American Journal of Psychology, 115, 201292.

Steiger, J.H. y Lind, J. (1980). Statistically based tests for the number of common factors. Comunicación presentada en el meeting anual de la Psychometric Society. Iowa City, Mayo de 1980.

Tanaka, J.S. y Huba, G.J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. British Journal of Mathematical and Statistical Psychology, 38, 197-201.

Thurstone, L.L. (1947). Multiple factor analysis. Chicago: University of Chicago press.

Tucker, L.R. y Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38, 1-10.

Velicer, W.F. (1976). Determining the number of components from the matrix of partial correlations. Psychometrika 41, 321–337.

Velicer, W.F. y Jackson, D.N. (1990). Component analysis versus common factor analysis: Some further observations. Multivariate Behavioral Research. 25, 97-114.