

ANÁLISIS FACTORIAL

La denominación genérica de “análisis factorial” agrupa diversos procedimientos de análisis multivariable que analizan la relación mutua (o de “interdependencia”) entre varias variables. Como el análisis de conglomerados, el análisis factorial pertenece a la clasificación de técnicas de análisis multivariable de *interdependencia*. No distingue entre variables dependientes e independientes. Su finalidad principal no es el análisis de relaciones causales, sino la agrupación de variables, en función de la variabilidad que cada variable comparte con otras variables. Concretamente, dos son sus objetivos fundamentales:

1. Analizar la correlación existente en una serie de variables, con el propósito de descubrir si comparten alguna estructura *latente* (no directamente observable).

Se busca la síntesis de la información proporcionada por “p” variables observadas (o *indicadores*), con la menor pérdida posible de información, en un número inferior de “k” variables no observadas (*factores comunes* o *componentes principales*, depende de la variedad analítica que se realice). Esta serie menor de variables *latentes* ha de caracterizarse por aglutinar variables empíricas que estén bastante correlacionadas entre sí y escasamente correlacionadas con aquellas variables empíricas que conforman otra estructura *latente* (o *dimensión* del concepto que se analice). La “no correlación” entre grupos de variables es una propiedad importante. Significa que los *indicadores* miden “dimensiones” diferentes en los datos.

La obtención de un modelo factorial se fundamenta en dos principios básicos comunes a otros procedimientos analíticos: *parsimonia* e interpretabilidad. De acuerdo con el principio de *parsimonia*, la solución factorial ha de ser sencilla, compuesta por el menor número posible de factores o componentes. A

este principio básico se suma la necesidad de que los factores extraídos sean estadísticamente significativos y susceptibles de interpretación sustantiva.

2. La obtención de *puntuaciones factoriales*, variables típicas o, en su caso, variables sucedáneas, para cada factor. Éstas actuarán en representación de los *factores* o *componentes* en análisis posteriores.

Este segundo objetivo, a diferencia del anterior, sólo se cumple cuando el plan de análisis de una investigación no concluye con la obtención de un modelo factorial. Al contrario, la obtención del modelo factorial constituye un paso previo a la aplicación de otras técnicas analíticas multivariadas, como el análisis de regresión múltiple, el análisis discriminante o el análisis de varianza, entre otros.

5.1. Orígenes del análisis factorial y su relación con otras técnicas multivariadas

Los antecedentes inmediatos del análisis factorial se remontan a una publicación de Karl Pearson de 1901 ("On lines and planes of closest fit to systems of points in space", *Phil.*, May, 2: 559-572). En esta publicación se hace la primera propuesta del procedimiento de análisis de "componentes principales". En 1904 Karl Spearman publicó un artículo con el título "General intelligence objectively determined and measured" (en la revista *American Journal of Psychology*, 15: 201-293) que trata sobre la covariación entre variables. Spearman parte de la hipótesis de que las distintas medidas de inteligencia pueden sintetizarse en un factor general (llamado "factor G"), común a todas las medidas, y un cierto número de factores específicos asociados a unas cuantas medidas: capacidad de análisis verbal, capacidad de análisis matemático y capacidad de integración espacial. Para ello observa las correlaciones entre las distintas puntuaciones de tests de varios tipos. Advierte que muchas de las correlaciones observadas pueden explicarse mediante un modelo más simple.

Pero fue en 1933, de la mano de Hotelling, cuando se describen los métodos de cálculo específicos al análisis de *componentes principales* ("Analysis of a complex of statistical variables into principal components", *Journal of Educational Psychology*, 24: 417-441; 498-520). En 1947 Thurstone (en *Múltiple factor analysis*, The University of Chicago Press) incorpora al análisis factorial el álgebra matricial para el análisis de las correlaciones. A partir de entonces se suceden las publicaciones sobre el análisis factorial en cualquiera de sus variantes. Al igual que el resto de técnicas analíticas multivariadas, su uso se amplía con la llegada de los ordenadores.

En los capítulos precedentes se ha hecho mención de la relación del análisis factorial con otras técnicas multivariadas. En especial, con el análisis de regresión múltiple, el análisis discriminante y el análisis de conglomerados. Con las dos últimas técnicas analíticas comparte un mismo propósito: la agrupación de objetos (casos o variables). Si bien se recomienda su uso complementario, con el análisis de conglomerados en una fase posterior, y con el análisis discriminante, como paso previo a su ejecución.

El uso complementario del análisis factorial con el análisis de *conglomerados* cumple una función confirmatoria. Su aplicación se dirige a la corroboración o con-

firmación de los conglomerados (o agrupaciones de objetos muy similares entre sí y distintos de otros grupos) obtenidos mediante el análisis de conglomerados. Para este propósito se emplea preferentemente la variante del análisis factorial “confirmatoria”. Esta variedad del análisis factorial se expone en el capítulo 6, por su configuración similar con el modelado de ecuaciones estructurales.

Respecto al análisis *discriminante*, ya se dijo en el capítulo anterior que es una técnica confirmatoria, no exploratoria. El análisis discriminante necesita de una configuración inicial de los datos clasificados en grupos. Esta clasificación inicial puede obtenerse mediante la realización de un análisis de conglomerados u otra técnica multivariable de interdependencia, como el análisis factorial exploratorio. Además, en la exposición del análisis discriminante continuamente se han destacado aspectos comunes en la realización de ambos procedimientos analíticos, a cuya relectura se remite (capítulo 4).

El análisis factorial también puede ser la antesala al análisis de *regresión múltiple*, en la comprobación y búsqueda de solución ante la existencia de una elevada *multicolinealidad* (correlación elevada entre una serie de variables independientes). El análisis factorial ayuda a la agrupación de variables muy correlacionadas entre sí en un número menor de factores o componentes no correlacionados. En esta situación, el análisis de regresión se realizará no con las variables independientes originales muy colineales, sino con las variables “típicas”, sucedáneos de variables o *puntuaciones factoriales* que resultan de la aplicación de un análisis factorial exploratorio. Este proceder se expone en el apartado 5.8, dedicado a las *puntuaciones factoriales*.

Por último, en el análisis de *correlación canónica* también pueden utilizarse factores o componentes principales en sustitución de las variables originales. La correlación canónica es (como se ha mencionado en el capítulo anterior) una técnica multivariable de dependencia que permite comprobar la existencia de interrelación entre una serie de variables dependientes y otra serie de variables independientes. Se busca la obtención de combinaciones lineales de cada serie de variables (dependientes e independientes) que maximicen las correlaciones entre las variables.

Cuando se realiza un análisis factorial previo a la ejecución de un análisis de correlación canónica, existen dos aproximaciones posibles:

- a) Efectuar un análisis factorial en cada una de las dos series de variables y sustituir las dos series de variables originales por los factores o componentes. Siguiendo este procedimiento se gana en simplicidad, al tener menos coeficientes canónicos que interpretar.
- b) Realizar un único análisis factorial en la serie total de variables. Esta segunda opción permite comprobar cuáles son los factores o componentes que mejor reflejan la estructura de los datos en cada una de las series de variables.

5.2. La variedad de modelos factoriales: tipologías

La denominación genérica de análisis factorial reúne una variedad de procedimientos analíticos que tienen un objetivo común: sintetizar la información contenida

en una serie de variables empíricas (u observadas) en un número inferior de variables *latentes*. El abanico de posibilidades es el siguiente:

A) *Análisis de componentes principales-análisis de factor común*

El análisis de *componentes principales* (ACP) se caracteriza por analizar la varianza total del conjunto de variables observadas. De ellas trata de determinar las dimensiones básicas (o “componentes”) que las definen. En el análisis de *factor común* (AFC) el estudio de las interrelaciones entre las variables se restringe, en cambio, a la *varianza común*. Es decir, a la búsqueda de un número reducido de “factores” que expresen lo que es “común” al conjunto de variables observadas.

En el análisis de *factor común* suelen diferenciarse distintos procedimientos de extracción de factores (que se exponen en el subapartado 5.5.1):

- Análisis de factor principal o de ejes principales.
- Máxima verosimilitud.
- Mínimos cuadrados generalizados y no ponderados.
- Factorización alfa.
- Factorización imagen.

B) *Análisis factorial exploratorio-análisis factorial confirmatorio*

La distinción entre análisis factorial “exploratorio” y “confirmatorio” depende de la finalidad del análisis y del conocimiento previo que el investigador tenga de la realidad que analice.

Cuando el investigador no parte de una configuración previa de “factores”, sino que, precisamente, realiza un análisis factorial para obtener un número mínimo de “factores” que sinteticen la información aportada por un conjunto amplio de variables, el análisis factorial es *exploratorio*. Si, por el contrario, se parte de un modelo previo sobre la estructura latente en los datos y lo que se desea es “confirmar” o negar la estructura latente hipotetizada (y no explorar las dimensiones latentes), el análisis factorial es *confirmatorio*. Su aplicación se dirige a la “confirmación”. De ahí le viene el nombre. Se quiere “confirmar” los factores hipotetizados en un modelo propuesto *a priori*, a partir de una teoría, de generalizaciones empíricas o de la estructura latente que el investigador espera encontrar en los datos.

El análisis factorial *confirmatorio* es una técnica de mayor sofisticación, que suele emplearse en fases avanzadas del proceso de investigación. Generalmente, tras la aplicación de otras técnicas estadísticas que proporcionen el modelo de partida. En cambio, el análisis factorial *exploratorio* se aplica en las primeras fases del análisis, en el estadio “exploratorio”, cuando se quiere agrupar variables muy correlacionadas, como un fin en sí mismo, o como un paso previo y necesario a la aplicación de otras técnicas analíticas multivariables. De su práctica resulta la obtención de un modelo hi-

potético de la estructura común *latente* en los datos analizados. Hasta la fecha ésta ha sido la aplicación más frecuente del análisis factorial: “explorar” las dimensiones *latentes* en los datos más que “confirmarlas”.

La explicación del análisis factorial realizada en el presente capítulo se limita exclusivamente al análisis factorial *exploratorio*. El análisis factorial *confirmatorio* se expone en el capítulo 6, dedicado al modelado de ecuaciones estructurales. La decisión de postergar la explicación de esta variedad analítica al capítulo siguiente se fundamenta en la necesidad de conocer, previamente, la técnica del modelado de ecuación estructural para comprender el análisis factorial *confirmatorio*. Como se verá en el capítulo 6, ambos procedimientos de análisis tienen elementos comunes, tanto en su formulación como en su desarrollo, razón que justifica su exposición conjunta.

C) Análisis factorial *R*-análisis factorial *Q*

En 1985 Comrey propone la distinción entre el análisis factorial tipo *R* y el análisis factorial tipo *Q*. El primero se caracteriza por tener como objetivo principal la identificación de un número reducido de dimensiones latentes en un conjunto de variables. Ésta es la modalidad de análisis factorial más usual. En cambio, el análisis factorial *Q* se ocupa de las interrelaciones entre casos, no entre variables. Ha tenido gran aplicación en psicología y demás ciencias de la conducta en la elaboración de tipologías de sujetos. Se trata de un procedimiento de clasificación similar al análisis de conglomerados. Ambos persiguen un mismo objetivo básico: la clasificación de un conjunto de individuos u objetos (países, municipios, partidos políticos, colegios...) en un número reducido de grupos mutuamente excluyentes. Los grupos han de estar formados por casos lo más similares posible entre sí y diferentes de los integrantes de los otros grupos. Pero, como se verá en la lectura comparada de ambas técnicas analíticas, existen amplias diferencias entre ellas.

En general, el análisis de conglomerados se presenta como un análisis más rígido, que no permite que un mismo caso pertenezca a más de un grupo. En el análisis factorial *Q*, por el contrario, un mismo caso puede tener un peso importante en más de un factor. Éste es uno de los inconvenientes más frecuentemente señalado en el análisis factorial *Q*. Dillon y Goldstein (1984: 43) añaden otra limitación importante: “Las correlaciones tipo *Q* eliminan diferencias atribuibles tanto a la media como a la dispersión de los individuos”. De manera que, dos individuos que muestren un mismo orden de puntuaciones de rango y espacio relacionarán bastante, sin considerar cualquier diferencia en los niveles de la media o dispersión, cuando en realidad las diferencias de medias y varianzas entre individuos adquieren especial importancia en cualquier agrupación de individuos.

A estas dos variedades del análisis factorial se suman otras, aunque de menor aplicación: el análisis factorial tipo *S*, *T*, *P* y *O*. Estas variedades difieren en la dimensión temporal que se considere, en la unidad de análisis, los índices de asociación y en la descripción de los factores en términos de variables u ocasiones (véase Dillon y Goldstein, 1984; Bisquerra, 1989).

D) *Análisis factorial métrico-análisis factorial no métrico*

La concepción tradicional del análisis factorial ha sido "métrica" debido a que en su cálculo intervienen estadísticos como la *media* y la *varianza*, que exigen que las variables sean *métricas* (de intervalo o de razón). Para variables *no métricas* existen otras técnicas multivariadas de interdependencia que permiten alcanzar objetivos similares al análisis factorial y se adecuan más a este tipo de variables. Sea el caso, por ejemplo, del análisis de correspondencias o el escalamiento multidimensional "no" métrico.

En los últimos años, sin embargo, han ido apareciendo programas, como el PRINCALS, que permiten la obtención de un modelo factorial *no métrico*. Además, las variables cualitativas pueden convertirse en variables *ficticias*, lo que permite además aplicar el llamado análisis factorial "booleano". Éste difiere del análisis factorial clásico (o *métrico*) en una serie de aspectos que, siguiendo a Bisquerra (1989: 335-336), se resumen en los siguientes:

1. En el análisis factorial *booleano* se aplica el cálculo del álgebra de Boole. De ahí le viene su denominación.
2. Las puntuaciones factoriales en el análisis factorial *métrico* suelen resultar de la combinación de las variables que presentan saturaciones o coeficientes factoriales ("factor loadings") elevados en los factores respectivos. En cambio, en el análisis factorial *booleano* cada caso tiene una puntuación de 1, si tiene una respuesta positiva (distinta de 0) para cualquiera de las variables dominantes en el factor. Cuando las saturaciones de las variables en los factores son 0, la puntuación individual también es 0.
3. En el análisis factorial *booleano* una variable puede tener saturación de 1 en más de un factor. En el análisis factorial *métrico* se intenta (siguiendo el principio de estructura simple, que lleva a la rotación de factores) que cada variable sature fundamentalmente en un factor. Lo que fuerza a que su presencia en los demás factores sea mínima.
4. Respecto a la bondad de ajuste, en el análisis factorial *booleano* ésta se comprueba comparando las respuestas binarias observadas con las estimadas a partir de la multiplicación de las saturaciones por el número de puntuaciones. Se consideran tanto las discrepancias positivas (el número de veces que la puntuación observada es 1) como las negativas (cuando la puntuación observada es 0, siendo el valor estimado 1). En el análisis factorial tradicional (*métrico*) se aplican otros criterios de bondad de ajuste que se resumen en el apartado 5.7.

5.2.1. *El análisis de componentes principales*

El análisis de componentes principales (ACP) constituye una variedad de análisis multivariable de interdependencia cuyo objetivo principal es la búsqueda de combinaciones de "p" variables observadas (o *indicadores*) en un número sustancialmente in-

ferior de “k” variables *latentes* (o no observadas). Mediante estas últimas se quiere reducir la dimensionalidad de la serie de variables originales, pero conservando la mayor parte de la información proporcionada por las variables observadas. A las variables *no observadas* se las denomina “componentes”.

En principio, pueden extraerse tantos “componentes” como variables observadas. De esta forma se lograría explicar toda la variabilidad de las variables originales; la suma de las varianzas de todos los componentes sería igual a la suma de las varianzas de las variables originales. Aunque no se habría alcanzado el objetivo reduccionista intrínseco a esta técnica analítica, al haber un componente por cada indicador. De lo que se trata es de explicar la mayor proporción de varianza total de las variables *observadas* con el menor número de *componentes* posible. Los primeros componentes suelen caracterizarse por extraer la mayor proporción de varianza de las variables originales. Los componentes se hallan dispuestos en orden decreciente. El primero se caracteriza por ser el que más cantidad de varianza de las variables originales extrae, mientras el último componente apenas explica variabilidad, siendo su tamaño considerablemente inferior. El investigador deberá, en consecuencia, decidir con cuántos “componentes” se quedará para representar la información incluida en la matriz de datos original.

Asimismo, se precisa que los *componentes* sean perpendiculares. Quiere esto decir, que no estén correlacionados entre sí. La “no correlación” entre los componentes se convierte en una propiedad importante porque supone que los componentes miden “dimensiones” diferentes en los datos. Por el contrario, las variables observadas han de estar correlacionadas entre sí para que puedan sintetizarse en un número reducido de componentes que agrupen a las variables más correlacionadas (y facilite su utilización en análisis posteriores). Si la correlación entre las variables originales fuese nula, el número de componentes coincidiría exactamente con el número de variables originales, con lo que no se habría alcanzado el objetivo principal del análisis.

Cada componente principal se expresa como función de las “p” variables observadas correlacionadas entre sí, que ponderan en dicho componente, mediante las ecuaciones siguientes:

$$CP_1 = \lambda_{11}X_1 + \lambda_{12}X_2 + \dots + \lambda_{1p}X_p$$

$$CP_2 = \lambda_{21}X_1 + \lambda_{22}X_2 + \dots + \lambda_{2p}X_p$$

.....

$$CP_k = \lambda_{k1}X_1 + \lambda_{k2}X_2 + \dots + \lambda_{kp}X_p$$

Para evitar la influencia indebida que las unidades originales de medición de las variables observadas pueden ejercer en la ponderación de los componentes, generalmente se procede a la *estandarización* de las variables originales previo a la extracción del modelo de componentes principales. Si las variables estuviesen estandarizadas, la letra “Z” sustituirá a “X” en las ecuaciones. La letra “λ” representa los pesos o coeficientes de saturación de los distintos indicadores en cada componente principal.

El modelo definido por ecuaciones puede representarse de forma gráfica. La figura 5.1 incluye la definición gráfica de un modelo de componentes principales formado por tres componentes y cinco variables originales o indicadores. Estas últimas (estandarizadas "Z" o no estandarizadas "X") se representan mediante un cuadrado, mientras que los componentes (o variables *latentes*) figuran en un círculo.

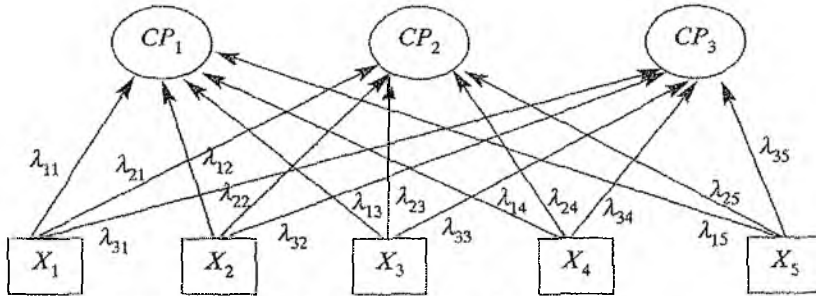


Figura 5.1. Representación de un modelo de componentes principales.

El procedimiento seguido en la extracción de un modelo de *componentes principales* se trata en el apartado 5.5, junto al modelo de *factor común*. Aunque algunos autores, como Dillon y Goldstein (1984: 24), lo consideran “un método de identificar las dimensiones factoriales de los datos y no como un modelo estadístico formal”. En su opinión, el análisis de componentes principales “esencialmente toma los datos e intenta determinar las dimensiones que definen su varianza total”. En las páginas siguientes se da al lector oportunidad de valorar dichas afirmaciones.

5.2.2. El análisis de factor común

El análisis de *factor común* (AFC) es otra técnica analítica multivariable de interdependencia igualmente aplicada a la medición de conceptos teóricos. Comparte el mismo objetivo básico del ACP: la descripción de un conjunto de “p” variables observadas (o indicadores) en términos de un número inferior de factores (o índices). Estos últimos se obtienen de la agrupación de los indicadores correlacionados entre sí, que representan una misma dimensión del concepto que se mide. Estos indicadores corresponden a las variables que presentan un coeficiente factorial más elevado en dicho factor. No obstante, existen tres diferencias básicas que distinguen a AFC de ACP:

1. En ACP se considera la varianza total de la serie de variables *observadas*. El propósito es maximizar la proporción total de la varianza explicada. En cambio, AFC está más orientado al análisis de la covarianza, no de la varianza. En esta última modalidad analítica la varianza se descompone en varianza común (o

comunalidad) y varianza específica. La *comunalidad* de cada variable (h_i^2) expresa la porción de la varianza total de la variable X_i que es compartida con las $p - 1$ variables observadas restantes. La varianza *específica* (e_i) es, por el contrario, la porción de la varianza total de la variable que no es explicada por los "factores comunes". Si la variable X_i se halla estandarizada, como es usual, su varianza total es 1: $Var X_i = 1 = h_i^2 + e_i$. De lo que se deduce que $e_i = 1 - h_i^2$.

- En suma, el objetivo fundamental del AFC es maximizar no la varianza total, sino la *comunalidad* total, a partir de un número inferior de factores latentes (que resumen la información contenida en un número superior de variables observadas). Son los factores comunes los que contribuyen a la covariación de los indicadores.
2. La diferencia básica anterior entre ACP y AFC repercute en el punto de partida de ambos análisis: la *matriz de correlación*. En ACP, la diagonal de la matriz está integrada por unos, dado que se trata de explicar la varianza total de las variables. En AFC, que analiza la covariación o varianza compartida, en la diagonal de la matriz figuran las *comunalidades*. Su valor es igual a la correlación múltiple cuadrada de cada variable con las demás variables del análisis, o la correlación absoluta más alta en una fila de la matriz de correlación (véase subapartado 5.4.3). El procedimiento seguido en la extracción de las dimensiones latentes (factores o componentes) es, no obstante, muy similar en ambas variedades analíticas de reducción de datos, como se verá en el apartado 5.5.
 3. En ACP los componentes principales se explican en función de las variables observadas, determinándose los pesos o saturaciones de cada variable en cada componente. En AFC las variables observadas o *indicadores* son, por el contrario, las que actúan a modo de variables dependientes en la ecuación lineal. Éstas se explican por variables no observadas: factores *comunes* y *únicos*. Recuérdese que en AFC se diferencia entre varianza *común* y varianza *específica*, a diferencia de ACP, donde no se hace dicha segregación de la varianza.

Como en ACP, el modelo analítico de AFC se asemeja a una ecuación de regresión múltiple. Cada variable observada X_i se expresa mediante una combinación lineal de un número pequeño de factores *comunes latentes* y un factor *único*, también *latente*. Estos últimos representan la parte de la varianza de la variable observada que "no" es explicada por los factores *comunes*. La elección de la letra "e" para denotar al factor *único* procede de su consideración como "término de error". Se asume que son independientes, lo que quiere decir que no se hallan correlacionados entre sí (unos factores *únicos* con otros *únicos*) ni tampoco con los factores *comunes*. Algebraicamente, el modelo de AFC puede representarse mediante las ecuaciones siguientes:

$$X_1 = \lambda_{11}F_1 + \lambda_{12}F_2 + \dots + \lambda_{1K}F_K + e_1$$

$$X_2 = \lambda_{21}F_1 + \lambda_{22}F_2 + \dots + \lambda_{2K}F_K + e_2$$

$$X_p = \lambda_{p1}F_1 + \lambda_{p2}F_2 + \dots + \lambda_{pK}F_K + e_p$$

Para cada variable X_p , siendo $i = (1, 2, \dots, p)$, el modelo se resume en:

$$X_i = \sum_{j=1}^k \lambda_{ij} F_j + e_i$$

Donde "k" representa el número de factores comunes y " λ_{ij} " las saturaciones o coeficientes factoriales (o "factor loadings"). Estos últimos cuantifican la relación de los factores no observables con cada indicador. En el desarrollo del modelo se intenta que estos coeficientes sean o muy elevados o muy pequeños, en los distintos factores comunes. Lo que lleva a la *rotación* del modelo factorial (como se verá en el subapartado 5.5.1). La finalidad es alcanzar un modelo factorial que sea lo más sencillo e interpretable posible, en el que cada indicador se relacione con un número mínimo de factores. Si las variables están *estandarizadas*, se representarían con la letra "Z", como es habitual.

El gráfico correspondiente a un modelo de factor común, compuesto por 3 factores comunes y 5 indicadores se incluye en la figura 5.2.

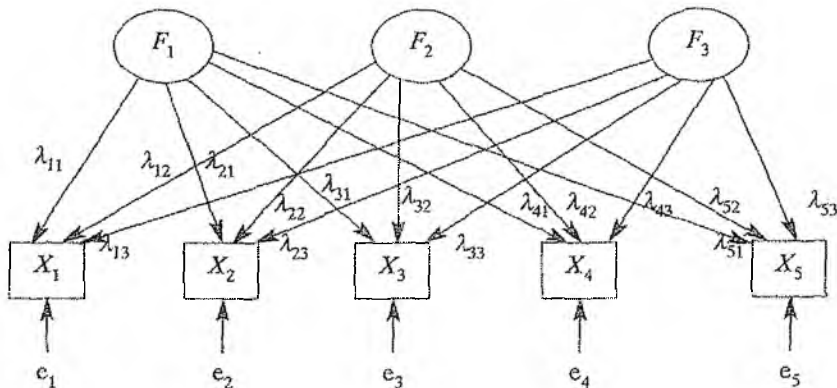


Figura 5.2. Representación de un modelo de factor común.

Si se comparan los modelos matemáticos de AFC y ACP, ilustrados en las figuras 5.1 y 5.2, se verá que ambos difieren. En ACP las variables no observadas se expresan en función de los indicadores. Lo que no equivale a expresar indicadores como función de variables latentes (*factor común y único*), como sucede en AFC.

Pese a sus diferencias, ambos modelos analíticos comparten las mismas áreas de aplicación. Incluso pueden alcanzarse resultados similares. Esto sucede cuando los valores de *comunalidad* se aproximan a 1 o existe un elevado número de variables. En caso contrario, especialmente cuando las *comunalidades* son pequeñas (próximas a 0), o varían considerablemente, los modelos ACP y AFC difieren bastante.

En paquetes estadísticos como, por ejemplo, el SPSS, ambos análisis figuran bajo el mismo encabezado de “análisis factorial”. Específicamente ACP se incluye en los procedimientos de extracción factorial. Esto puede llevar a su consideración errónea como análisis factorial (“común”).

En la decisión de qué modalidad aplicar, ACP o AFC, intervienen aspectos varios, como los objetivos de investigación y/o el conocimiento previo que se tenga de la varianza de las variables. El modelo ACP muestra mayor adecuación cuando se está interesado en predecir y determinar el número mínimo de factores necesarios para explicar la mayor proporción de varianza posible en la serie original de variables observadas. También se adecua a la situación de conocimiento previo de que la *varianza específica* representa una proporción pequeña de la varianza total de las variables. Si no se dispone de este conocimiento, y si se tiene como objetivo principal la identificación de las dimensiones latentes de las variables, la elección apropiada será efectuar un AFC.

En la práctica, ambas variedades analíticas pueden aplicarse de manera conjunta. Puede realizarse un ACP como paso previo a un AFC, con el propósito de determinar la dimensionalidad del espacio factorial común. Cuando esto sucede, el análisis de AFC parte de los *componentes principales* derivados del ACP. Éstos se utilizan como los factores originales, factores “no rotados” (Manly, 1990). Posteriormente se procederá a la rotación factorial y a la obtención del modelo AFC.

5.3. La obtención de un modelo factorial explorativo: fases principales

La consecución de un modelo factorial exploratorio supone el cumplimiento de una serie de fases principales. Éstas pueden resumirse en cinco:

1. La fase previa de *preparación de los datos para el análisis*. Comprende la comprobación de los supuestos básicos para una correcta aplicación del análisis factorial. A ello se suma una serie de decisiones clave. Desde la elección de variables hasta qué matriz emplear: la matriz de *covarianzas* o la de *correlaciones*. Esta decisión está bastante determinada por las características de las variables incluidas en el análisis. Si el investigador ha decidido proceder a la estandarización de las variables, con el objetivo de garantizar su equivalencia inicial previo a la realización de los análisis (el cálculo de los pesos o coeficientes factoriales en las estructuras latentes, sean factores comunes o componentes), la matriz elegida será la matriz de *correlaciones*. En cambio, si prefiere que las variables participen en los análisis en su unidad de medición original, suele recomendarse la *matriz de covarianzas*.
2. La *extracción de los factores o componentes iniciales*. Esta fase incluye dos decisiones clave para la obtención de los modelos factoriales. Primero, el *procedimiento* a seguir en la extracción de factores: componentes principales, ejes principales, mínimos cuadrados no ponderados, mínimos cuadrados generali-

- zados, máxima verosimilitud, alfa e imagen. Segundo, los *criterios* a adoptar sobre el número de factores a retener en el modelo factorial: el criterio de raíz latente o autovalor, el porcentaje de varianza explicada, de caída, la significatividad estadística y la interpretabilidad.
3. *La obtención de la matriz factorial y su interpretación.* El modelo factorial adquiere su forma en la matriz factorial. Ésta incluye los pesos o saturaciones de las variables observadas en cada uno de los factores o componentes no observados (estructura latente). Para facilitar su interpretación normalmente se procede a la rotación de los ejes factoriales. La *rotación* puede ser *ortogonal* (varimax, quartimax y equimax) u *oblicua* (oblimin –quartimin, covarimin–, oblimax y promax). Depende de la correlación permitida entre los factores o lo que el investigador prevea en función de las variables incluidas en el análisis.
 4. *La evaluación del modelo factorial,* desde la vertiente estadística y la lógica-sustantiva. En caso afirmativo se puede dar por concluido el análisis o proceder, si se quiere, al cálculo de las *puntuaciones factoriales*. En caso negativo, habría que volver a las fases previas del análisis para comprobar a qué se debe su no adecuación y adoptar alguna medida al respecto.
 5. *El cálculo de las puntuaciones factoriales* (para cada caso). Esta fase, a diferencia de las precedentes, no es imprescindible para la consecución de un modelo factorial exploratorio. Sólo cuando los resultados del análisis factorial van a ser objeto de análisis mediante otras técnicas de análisis multivariable (regresión múltiple, análisis discriminante, factorial confirmatorio u otro).

La figura 5.3. recoge estas cinco fases esenciales en un análisis factorial exploratorio en forma de gráfico. En las páginas siguientes se detalla cada uno de los integrantes de un análisis factorial.

5.4. Preparación de los datos para el análisis

Como en todo procedimiento analítico, la antesala del análisis factorial exploratorio es la comprobación de la pertinencia de optar por esta modalidad analítica para cubrir los objetivos de investigación. Se comprueba su adecuación tanto respecto a los objetivos de estudio como a las características de los datos que se quiere analizar. Ello lleva a verificar el cumplimiento de los supuestos básicos para la correcta realización del análisis factorial.

5.4.1. Los supuestos básicos y decisiones clave

Los supuestos básicos que garantizan el correcto desarrollo del análisis factorial exploratorio (tradicional) cabe resumirlos en cuatro:

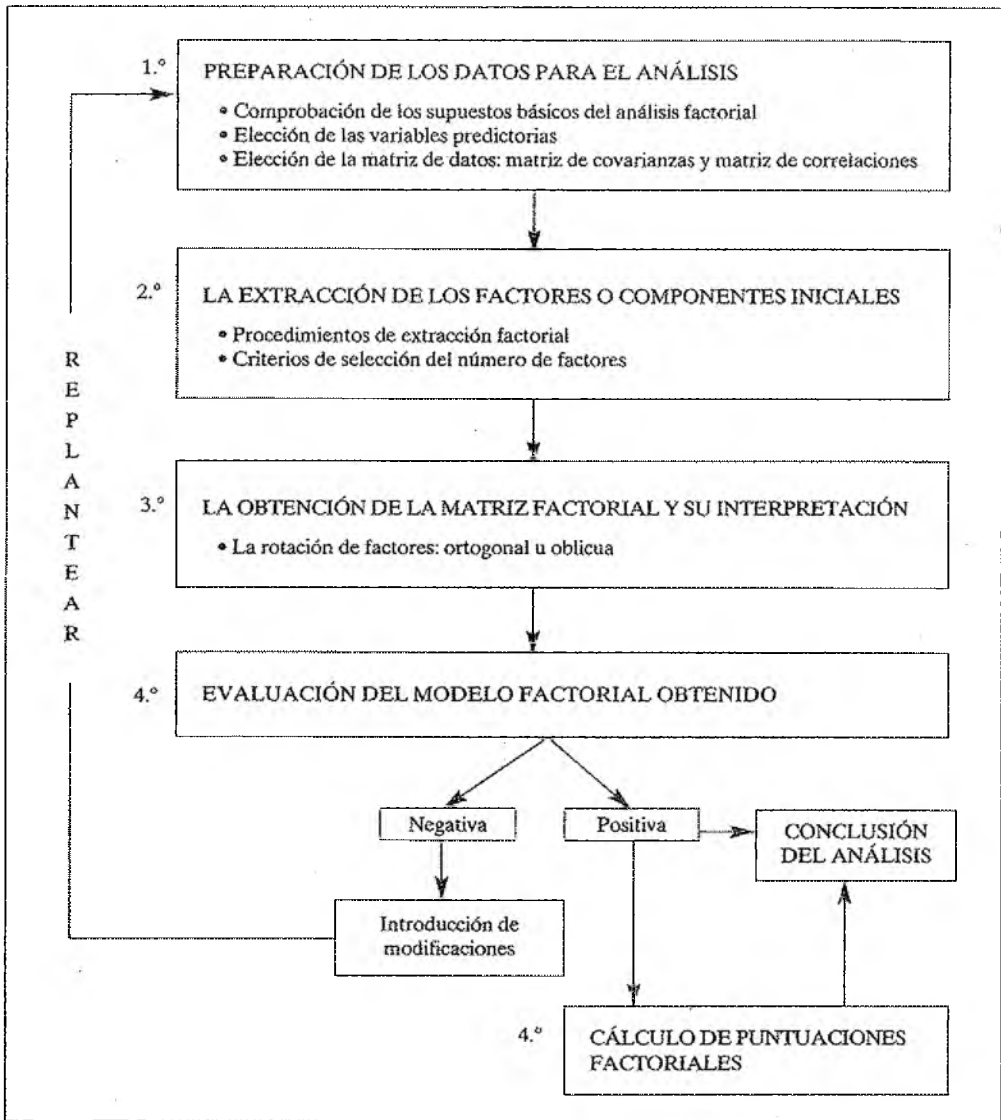


Figura 5.3. Fases esenciales de un análisis factorial exploratorio.

A) Tamaño muestral elevado

El análisis factorial, como técnica de análisis multivariable, exige que el tamaño muestral sea elevado. Las muestras pequeñas están más predispuestas a estimaciones de los coeficientes de correlación in fiables. Pero, ¿cuál es el tamaño mínimo reco-

mendado? Comrey (1973a) propone una escala muestral guía que va desde 50 casos (que se considera un tamaño muestral "muy pobre") hasta 1.000 (que representa un tamaño muestral "excelente"). Entre ambos extremos se sitúan los tamaños muestrales de 100 ("pobres"), 200 ("justo"), 300 ("bueno") y 500 ("muy bueno"). En suma, entre 200 y 300 casos se sitúa el tamaño mínimo recomendado para un desarrollo adecuado del análisis factorial. Pero, como sucede en otros procedimientos analíticos expuestos, lo más preciso es considerar el tamaño muestral en relación con el número de variables a analizar. Tabachnick y Fidell (1989) proponen, como regla, que exista al menos 5 casos por variable. Éste sería el mínimo. Cuanto más se supere esta proporción, mejor porque ayuda a la obtención de estimaciones muestrales estables.

B) *Normalidad multivariable*

Todas las variables observadas y sus combinaciones lineales han de estar distribuidas normalmente. Aunque ACP y AFC se emplean con una finalidad eminentemente descriptiva (sintetizar las relaciones entre un conjunto amplio de variables), lo que no demanda el cumplimiento obligatorio del supuesto de *normalidad*, su existencia favorece la obtención de un modelo factorial más preciso. Máxime cuando se recurre a la inferencia estadística para la determinación del número de factores a retener. El recurso a procedimientos de extracción habituales en AFC, como los llamados "máxima verosimilitud" (ML) o "mínimos cuadrados", exige el cumplimiento del supuesto de *normalidad multivariable*. En ACP, este segundo supuesto no es un requisito básico para su ejecución, aunque la *asimetría* severa puede distorsionar los resultados.

En el subapartado 1.1.6 se resumen los procedimientos más seguidos en la detección del supuesto de *normalidad*, junto a los remedios más aplicados ante su incumplimiento. La salida de la *normalidad* se relaciona con estimaciones sesgadas de estadísticos como la media, la desviación típica, la correlación o la covarianza, por ejemplo. De ahí la exigencia de introducir alguna transformación en las variables que se distancian de la *normalidad* para evitar su incidencia negativa en los resultados del análisis. Para no redundar en temas ya tratados, se remite a la relectura del susodicho subapartado.

C) *Linealidad*

La *normalidad multivariable* lleva a un tercer supuesto básico: la *linealidad*. Las relaciones entre pares de variables han de ser lineales. En caso contrario, se estaría igualmente ante un análisis sesgado.

Asimismo, se asume que las relaciones entre las variables observadas y los factores latentes, representadas en un sistema de ecuaciones, ha de ser también lineal. El cumplimiento de este supuesto suele realizarse con la ayuda de los gráficos referidos en el subapartado 1.1.4, donde también se indican remedios ante la "no linealidad". Pero, tén-

gase presente que este supuesto, al igual que el anterior, sólo se exige para la realización de un análisis factorial *métrico*. En los *no métricos* dicho supuesto se relaja.

D) *Correlación entre las variables*

A diferencia de otros análisis multivariantes, como regresión o discriminante, en los análisis factoriales la *multicolinealidad* no es un problema. Al contrario, se demanda la existencia de correlación entre las variables. Los análisis son pertinentes sólo cuando existe correlación entre las variables: al menos $\geq ,30$. Si de la inspección de la matriz de *correlación* R se observa que ninguna o muy pocas correlaciones superan el valor 0,30, se debería desconsiderar la aplicación de un análisis factorial. Las variables apenas están correlacionadas. Por lo que, no tiene sentido la búsqueda de estructuras *latentes* (llámense factores comunes o componentes principales), que agrupan a variables observadas (o indicadores), correlacionadas entre sí, que expresan una misma dimensión del concepto que se mida. En cambio, a medida que aumenta la correlación entre los indicadores (desde 0,30 hasta 1,0), se incrementa la probabilidad de que su contenido (varianza) pueda sintetizarse en un número bastante inferior de factores o componentes. Si bien, hay que advertir que la existencia de correlación elevada entre pares de variables, por separado, no siempre garantiza la existencia de factores, como se verá en el subapartado 5.4.3, dedicado a la matriz de correlación.

- Además de comprobar estos cuatro supuestos mínimos, antes de proceder a la ejecución del análisis factorial, se deberán adoptar algunas decisiones importantes. Entre ellas la referida al tratamiento de los *casos sin respuesta*: si incluirlos o eliminarlos del análisis. Sobre este particular, se actúa siguiendo las pautas resumidas en capítulos precedentes. En especial, el subapartado 1.3.1, donde se ofrece una información más detallada del tratamiento de los *casos sin respuesta*.

Otra decisión clave concierne a la *estandarización* de las variables observadas. Antes de comenzar los análisis, ha de decidirse si las variables van a analizarse en su unidad original de medición o, por el contrario, va a procederse a su *estandarización*. Como ya se ha mencionado con anterioridad, la *estandarización* es una buena opción porque favorece la comparabilidad de las variables. En el análisis factorial, en concreto, es la opción más practicada. Previo a la realización del análisis, se estandarizan las variables observadas, dividiendo cada una de ellas por su desviación típica estimada. De esta forma se evita la incidencia desigual de la unidad de medición de las variables.

Las variables muy heterogéneas (con elevada varianza) logran mayores pesos en el factor o componente que las homogéneas (aquellas con escasa variabilidad). La cuantía de la varianza depende de la unidad de medida de la variable. Si se compara, por ejemplo, la incidencia de la variable "ingresos" (medida en pesetas) con la variable "edad" (medida en años), lo normal es que la varianza de la primera variable supere a la segunda, repercutiendo en el modelo factorial; e incluso la varianza de la misma variable "ingresos" será mayor, si se halla medida en pesetas que cuando está en dólares o en euros.

Con la *estandarización* de las variables, que quedan transformadas a unidades de desviación típica, se las sitúa en un plano de igualdad ante los análisis. De esta forma se posibilita el tratamiento conjunto y comparativo de la incidencia de variables de distinto grado de heterogeneidad.

La *estandarización* tiene su repercusión inmediata en la matriz "input" (de entrada) a elegir para la obtención de un modelo factorial. Si el investigador decide que las variables se analicen en su unidad de medición original, la matriz será de *covarianzas*. La matriz de *correlación* supone que las variables estén estandarizadas.

5.4.2. La matriz de covarianzas

Cuando las variables observadas se encuentran en métricas similares, con varianzas apenas divergentes en cuantía, una decisión apropiada es realizar ACP o AFC a partir de la matriz de *covarianzas*. La pertinencia de esta matriz se cuestiona, cuando se analizan variables de varianzas muy diferentes. Su elección llevaría a un modelo factorial, cuyos primeros factores o componentes estarían integrados precisamente por aquellas variables de mayor varianza, en menoscabo de las variables de menor varianza, lo que puede desvirtuar la realidad que se analiza. Este hecho ha influido en el menor uso de la matriz de *covarianzas*, incluso en modelos analíticos como el de factor común, desarrollado en términos de varianzas-covarianzas. La mayoría de los autores (Kim y Mueller, 1978a, 1978b; Dillon y Goldstein, 1984; Dunteman, 1989) aconsejan la *estandarización* y el empleo consiguiente de la matriz de *correlación*, tanto en ACP como en AFC.

5.4.3. La matriz de correlación

Cuando se procede a la *estandarización* de las variables originales, la matriz de datos brutos ($N \times p$; de casos \times variables) se transforma en una matriz de *correlaciones* R ($p \times p$), integrada por las distintas correlaciones entre cada par de variables observadas.

En ACP, la diagonal principal de la matriz está integrada por unos, al analizarse toda la variabilidad de las variables. En AFC, como únicamente se analiza la varianza *común*, la porción de varianza que cada variable comparte con el resto de variables, las *comunalidades* de cada variable forman la diagonal principal de la matriz de correlaciones. Ésta es la primera diferencia importante entre ambas modalidades analíticas.

En AFC la matriz de *correlaciones* suele referirse como matriz de *correlación reducida*, denotándose por " \bar{R} " o " R^* ", en vez de R . Adopta la forma siguiente, con las *comunalidades* en la diagonal principal.

$$\bar{R} = \begin{bmatrix} h_1^2 & r_{12} & \dots & r_{1p} \\ r_{12} & h_2^2 & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ r_{1p} & r_{2p} & \dots & h_p^2 \end{bmatrix}$$

Las *comunalidades* h_i^2 (para X_i , siendo $i = 1, 2, 3, \dots, p$) son las correlaciones múltiples cuadradas de cada variable con el resto de las variables en el análisis. Su valor absoluto es el más elevado en cada fila de la matriz de correlaciones. Los " r_{ii} " son los coeficientes de correlación producto-momento de Pearson usuales, que se calculan para cada par de indicadores (subapartado 1.3.2). La matriz de correlaciones en ACP sería igual salvo en la diagonal, donde figura 1 en vez de " h_i^2 ".

Calculada la matriz de *correlación*, se procede, en primer lugar, a su inspección visual. La finalidad es comprobar si las variables se hallan relacionadas, y en qué grado. El investigador espera encontrar correlaciones elevadas entre las variables, dado que el objetivo principal del análisis es la agrupación de variables que comparten una misma estructura latente. La correlación mínima propuesta normalmente es 0,30. Si la mayoría de las correlaciones en la matriz no excede este valor, se debería reconsiderar la pertinencia del análisis factorial. Con correlaciones pequeñas (inferiores en magnitud a 0,30) es improbable que las variables compartan suficiente varianza común para constituir un factor. Por el contrario, cuando las correlaciones exceden el valor mínimo de 0,30 (mejor cuanto más próximas se sitúen del valor máximo de 1,0) puede, a partir de la matriz de *correlaciones*, preverse que las variables correlacionadas entre sí pueden componer una misma dimensión latente.

No obstante, hay que enfatizar que se está en el plano de la conjetura, no en la plena certeza. Correlaciones bivariadas elevadas no siempre garantizan la existencia de factores. "Es posible que las correlaciones sean entre sólo dos variables y que no reflejen procesos latentes que estén simultáneamente afectando varias variables" (Tabachnick y Fidell, 1989: 604). De ahí la recomendación de examinar matrices de correlaciones parciales.

En las matrices de *correlaciones parciales*, las correlaciones se ajustan a pares para los efectos de las demás variables. Si las variables comparten factores comunes, los coeficientes de *correlación parcial* entre pares de variables deberían ser pequeños (próximos a 0,0), al haberse eliminado los efectos lineales de las demás variables o, lo que es igual, al haberse controlado la varianza común. Estos coeficientes de *correlación parcial* se toman como indicadores de la fuerza de la relación entre las variables. Pero, a diferencia de los coeficientes de *correlación no parciales*, ahora interesan coeficientes muy bajos porque denotan la presencia de factores. Se consideran estimaciones de correlación entre factores únicos y, como es de esperar, la varianza única se desea que sea escasa para que pueda procederse a la agrupación de variables (correlacionadas) de una misma dirección latente.

Para comprobar el grado de intercorrelación entre las variables y la presencia de una estructura común latente, existen unos estadísticos que se suman a los coeficientes de correlación. Los de uso más común son:

A) El determinante de la matriz de correlación

Un valor del determinante de la matriz de correlación próximo a "0,0" expresa la existencia de intercorrelaciones muy elevadas entre las variables. En esta situación, el

análisis factorial es pertinente, al poderse obtener combinaciones lineales de variables correlacionadas.

B) La prueba de la esfericidad de Barlett

En 1950 Barlett propone una prueba que ayuda a determinar si existe relación significativa entre las variables analizadas. Esta prueba es de utilidad en el análisis factorial, al igual que en el análisis de la varianza. Mediante ella se comprueba la correspondencia de la matriz de correlación con la matriz *identidad*. Recuérdese que por *matriz identidad* se entiende aquella matriz cuya diagonal principal (que expresa la correlación de cada variable consigo misma) está formada por unos, mientras el resto de los términos de la matriz son ceros. El determinante de esta matriz es igual a 1 ($|R| = 1$). Adviértase que el determinante de la matriz de correlación se representa entre dos barras).

$$\text{Matriz identidad} = I = R = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

En la *prueba de esfericidad de Barlett* la hipótesis *nula* (aquella que se espera rechazar) se formula en los términos siguientes: $H_0 : |R| = 1; R = I$. Significa que la *matriz de correlación* se corresponde con la *matriz identidad*. Las variables no están correlacionadas. El gráfico que corresponde a esta situación es el de una nube de puntos cuya forma se ajusta a una *esfera*. La hipótesis *alternativa* es la opuesta: $H_1 : |R| \neq 1; R \neq I$. Su aceptación supone que la matriz de *correlación* difiere de la *matriz identidad*. Existe *correlación* (de mayor o menor grado) entre las variables. El determinante de la matriz difiere de 1. Todo lo cual indica que puede realizarse un análisis factorial.

En la prueba de esfericidad de Barlett se utiliza el *determinante* de la matriz de correlación de la muestra observada como estimador del determinante de la matriz de correlación de la población a la que pertenece la muestra analizada. Su aplicación exige el cumplimiento del supuesto de *normalidad multivariable* ya referido en el subapartado 5.4.1.

Para el contraste de hipótesis se recurre a la distribución *chi-cuadrado*, con $\frac{1}{2}(p^2 - p)$ grados de libertad, siendo "p" el número de variables en la matriz de correlación.

$$B = \chi^2 = \left[N - 1 - \frac{1}{6(2p + 5)} \right] \ln|R|$$

Donde: " $\ln|R|$ " es el logaritmo neperiano del determinante de la matriz de correlación.
"N" el número de individuos en la muestra.

Como en todo contraste de hipótesis, la obtención de un valor χ^2 elevado, superior a su correspondiente valor *teórico* (definido por los grados de libertad y el nivel de significación elegido), supone el rechazo de la hipótesis *nula*. La matriz de *correlación* no se corresponde con la matriz *identidad*. El análisis factorial es pertinente. En caso contrario, habría que optar por otra técnica de análisis.

C) Índice KMO (Kaiser-Meyer-Olkin)

Kaiser propone en 1970 (en "A second-generation little jiffy", *Psychometrika*, 35: 401-415) un índice que compara las correlaciones observadas con sus correspondientes correlaciones parciales. La finalidad es determinar la propiedad de hacer un análisis factorial. El índice, cuyas siglas recoge la primera letra del nombre de sus promotores (KMO: Kaiser, Meyer y Olkin) se define en su aplicación a variables múltiples mediante la siguiente ecuación:

$$KMO = \frac{\sum_{i=1}^N \sum_{j=1}^N r_{ij}^2}{\sum_{i=1}^N \sum_{j=1}^N r_{ij}^2 + \sum_{i=1}^N \sum_{j=1}^N s_{ij}^2}$$

Donde: " r_{ij} " es el coeficiente de correlación simple entre las variables i y j .

" s_{ij} " es el coeficiente de correlación parcial entre las variables i y j .

De los sumatorios se excluyen los coeficientes de correlación (simple o parcial) de una variable consigo misma.

El rango de valores va de 0,0 a 1,0. Interesan valores elevados (próximos a 1,0) porque indican la existencia de intercorrelación entre las variables. La suma de los coeficientes de correlación es relativamente grande comparada con la suma de los coeficientes de correlación parcial. El análisis factorial es posible. Valores de KMO inferiores a 0,50 suponen, por el contrario, la no adecuación del análisis factorial, al haber poca correlación entre las variables. Se obtendría un modelo factorial con tantos factores o componentes como número de variables observadas. Con lo que no se alcanzaría el objetivo fundamental del análisis factorial: la síntesis de una serie de variables empíricas en un número inferior de factores o componentes.

Para mayor exactitud, Kaiser propone en 1974 (en "An index of factorial simplicity", *Psychometrika*, 39: 31-36) la siguiente interpretación de valores KMO: 0,90, maravillosos o muy bueno; 0,80, meritorio; 0,70, medio o normal; 0,60, mediocre; 0,50, despreciable o bajo; y un valor $< 0,50$, totalmente inaceptable.

La mayoría de los programas estadísticos ofrecen el índice KMO global (para el conjunto de variables), calculado mediante la fórmula anterior. Pero también se ofrece el valor KMO para variables individuales. Este segundo índice KMO individual se calcula mediante la siguiente ecuación:

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{(i,j)} r_{ij}^2 + \sum_{i \neq j} S_{ij}^2}$$

Es la suma de los cuadrados de los coeficientes de correlación entre una variable concreta y todas las demás dividido por la suma de este valor y el sumatorio de los coeficientes de correlación parcial cuadrado. Pero, a diferencia del índice KMO global, en el individual no se considera la correlación de la variable consigo misma. De ahí el término “ $i \neq j$ ”. En el programa SPSS los valores KMO individuales se obtienen de la matriz de *correlación anti-imagen*. Figuran en la diagonal principal de la matriz. Toda variable que presente un valor KMO inferior a 0,50 debería considerarse candidato a ser eliminado del análisis factorial. La eliminación de una variable supone un nuevo cálculo del índice KMO, debido a que éste se ve afectado por la eliminación de la variable. Los valores KMO individuales, para las variables no eliminadas, cambian al igual que el índice KMO global, cuando se elimina una variable del análisis. Véanse los datos del ejemplo siguiente.

D) Medida de adecuación de la muestra (MSA)

Esta medida de adecuación muestral (MSA_i ; “Measure of Sampling Adequacy”) se calcula para cada variable individual, de forma similar al índice KMO individual. Se compara, para cada variable por separado, un coeficiente de correlación simple con el correspondiente coeficiente de correlación parcial. Para cada variable la fórmula es la siguiente:

$$MSA_i = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{(i,j)} r_{ij}^2 + \sum_{i \neq j} S_{ij}^2}$$

Como en el índice KMO, en MSA_i se excluyen de los sumatorios los casos en que $i = j$. Asimismo, interesan valores próximos a 1,0. Indican que la variable es adecuada para su inclusión en un análisis factorial. En cambio, valores próximos a 0,0 aconsejan la desestimación de la variable para el análisis.

E) Correlación anti-imagen (AIC)

Es el negativo del coeficiente de correlación parcial. De modo que interesan valores AIC bajos para que pueda realizarse un análisis factorial. Recuérdese que las *correlaciones parciales* miden la correlación entre dos variables cuando se ha eliminado la

influencia del resto de las variables. Sus valores son altos (próximos a 1,0), cuando las variables comparten variabilidad.

En la matriz de correlación *anti-imagen* se observan las distintas correlaciones entre las variables (fuera de la diagonal principal). Cuando estas correlaciones se aproximan a 0,0, se deduce que puede aplicarse un análisis factorial. La existencia de una elevada proporción de coeficientes elevados (próximos a 1,0) desaconsejan su aplicación.

F) El coeficiente de correlación múltiple cuadrado (R_i^2)

Una última medida del grado de relación entre las variables lo proporciona el *coeficiente de correlación múltiple cuadrado* (R_i^2). Las correlaciones múltiples cuadradas han de ser elevadas, de lo contrario, se plantearía la desestimación del análisis de las variables con bajos coeficientes de correlación múltiple cuadrado. Antes de proceder a la desestimación de estas variables se aconseja comprobar sus correspondientes valores de *comunalidad* y coeficientes factoriales. Es decir, no quedarse únicamente por lo dicho por esta medida de correlación.

Los valores de los *coeficientes de correlación múltiple cuadrado* coinciden con las *comunalidades* "iniciales" de cada variable, salvo cuando las dimensiones latentes se extraen mediante *componentes principales*. En ACP las *comunalidades iniciales* son 1,0, al no diferenciarse a priori entre *varianza común* y *específica* (o *única*). A ello se suma que en la estimación de las *comunalidades iniciales* se consideran "todos los componentes posibles". Éstos igualan al número de variables empíricas. En la estimación de las *comunalidades posteriores* ya no sucede así. Éste es un aspecto clave que distingue al ACP de AFC. En AFC las estimaciones iniciales de *comunalidad* sí reflejan *varianza común*: la proporción de la varianza total de X_i que es explicada por la regresión en las "p - 1" variables restantes. Dicho con otras palabras, con *correlaciones múltiples cuadradas* de cada variable observada X_p que actúa a modo de variable dependiente, siendo las "p - 1" variables restantes las independientes.

Diversos programas estadísticos, como SPSS, muestran estos coeficientes de correlación múltiple cuadrado para cada variable en una columna de la tabla de *comunalidad* o incluida en la tabla de estadísticos iniciales. La *comunalidad* se comentan, además, en el apartado 5.6.

EJEMPLO DE COMPROBACIÓN DE LA PERTINENCIA DEL ANÁLISIS FACTORIAL

Las mismas 14 variables utilizadas en el análisis de conglomerados *jerárquico de vinculación simple* (capítulo 3) vuelven a analizarse mediante distintos procedimientos factoriales, en busca de su validación. Si bien, se parte de que puede haber discrepancias entre los distintos modelos, principalmente, debido a que el análisis de conglomerados realizado es "je-

rárquico", lo que lleva a la agrupación de las variables por "etapas", en función de la correlación existente entre ellas. En el análisis factorial, por el contrario, las relaciones entre todas las variables se analizan simultáneamente. No obstante, a lo largo del presente capítulo podrá comprobarse la similitud de la agrupación de las variables obtenida en los distintos procedimientos analíticos, pudiéndose reafirmar el modelo de agrupación de las variables que resulta del análisis de conglomerados. Para su constatación, se recomienda comparar los resultados de los distintos análisis: conglomerados (capítulo 3) y factorial.

Como las variables analizadas son las mismas que se han ido observando mediante distintos procedimientos analíticos, no se va a repetir información ya proporcionada en los capítulos precedentes (es el caso, por ejemplo, de la *matriz de correlaciones*, ya expuesta en el capítulo 1). Los datos que van a comentarse en el presente capítulo se limitan a lo específico del análisis factorial, como lo correspondiente a la comprobación del grado de intercorrelación entre las variables y la presencia de una estructura latente común que permita la obtención de un modelo factorial.

En el subapartado 1.3.2 figura la *matriz de correlaciones* de las 14 variables de interés, a las que se añade una variable ahora excluida para que los resultados puedan compararse con los obtenidos en el análisis de conglomerados. Se trata de la variable X_5 ("simpatía hacia latinoamericano"), que fue eliminada para evitar problemas de *colinealidad* en la aplicación posterior del análisis discriminante.

En la *matriz de correlaciones* puede comprobarse que, aunque las correlaciones no son en general muy elevadas —la más alta se da entre las variables X_{13} ("vecino marroquí") y X_{10} ("casar con marroquí"): $r = ,573$; seguida de la correlación entre X_{15} ("simpatía marroquí") y X_{10} ("casar con marroquí"): $r = ,476$ —, son muchas las que superan el referente mínimo de $\pm,30$, que indica la posibilidad de encontrar una estructura latente en los datos, que permita la síntesis de las 14 variables observadas en un número inferior de variables *latentes* (llámense factores comunes o componentes). Aunque, hay que advertir que la existencia de correlaciones elevadas entre pares de variables (como se registra en la matriz de correlaciones) no siempre garantiza la existencia de dimensiones latentes en los datos. Para constatarlo hay que proceder a la realización del análisis factorial, que exige que previamente se haya comprobado la existencia de interrelaciones mínimas entre las variables que permita su realización. A tal fin, se calcula:

A) El determinante de la matriz de correlaciones

El valor obtenido es ,07667, un valor próximo a 0,0, que indica la existencia de intercorrelaciones muy elevadas entre las variables. Como el valor en realidad obtenido no es exactamente 0, puede afirmarse que sí existen intercorrelaciones, aunque no demasiado elevadas entre las variables. Ello permite la realización del análisis factorial. Existe suficiente *varianza común* de las variables que ayude a su agrupación u obtención de combinaciones lineales de variables "correlacionadas".

B) La prueba de la esfericidad de Bartlett

Mediante ella se comprueba la H_0 de que la matriz de correlaciones es una matriz *identidad* (en tal caso, las variables no estarían correlacionadas) frente a la H_1 que niega la existencia de dicha correspondencia (lo que favorece la realización de un análisis factorial).

La correcta realización de esta prueba exige el cumplimiento del supuesto de *normalidad multivariable*.

El valor χ^2 aproximado obtenido es 3.270,734, con 91 grados de libertad. Recuérdese que los grados de libertad son iguales a $1/2 (p^2 - p) = 1/2 (14^2 - 14) = 91$. Al ser el nivel de significación asociado pequeño (el valor χ^2 empírico supera bastante al correspondiente valor teórico), se rechaza la H_0 incluso a un nivel de significación de ,001. La *matriz de correlación* no coincide con la *matriz identidad*. El análisis factorial es posible al haber correlación suficiente entre las variables.

C) El índice KMO (Kaiser-Meyer-Olkin)

Este índice global compara las magnitudes de los coeficientes de correlación observados con los coeficientes de correlación parcial. Al ser el valor obtenido ,821 (valor que Kaiser calificaría de "meritorio"), significa que puede realizarse un análisis factorial porque las correlaciones entre pares de variables pueden explicarse por otras variables.

Respecto a los valores *KMO individuales*, éstos figuran en la diagonal principal de la *matriz de correlación anti-imagen* respectiva (tabla A).

Como todas las variables presentan un valor KMO superior a ,50 (el valor KMO más bajo es ,545, que corresponde a la variable "sexo", y el más alto es ,895, en la variable "inmigrante delincuente"), no es necesario eliminar ninguna variable antes de realizar el análisis factorial. Todas las variables muestran ser idóneas para participar en el análisis.

Además, en la *matriz de correlación anti-imagen* obsérvese que los coeficientes obtenidos son, en su mayoría, valores próximos a 0,0. Lo que nuevamente indica la adecuación de los datos para la realización del análisis factorial. Recuérdese que la *correlación anti-imagen* (AIC) se define como el negativo del coeficiente de correlación parcial (que mide la correlación entre dos variables, una vez eliminada la influencia del resto de las variables). Cuanto más bajo sea su valor, más varianza comparten las variables.

D) Coeficientes de correlación múltiple cuadrados

Los coeficientes de correlación múltiple cuadrados se obtienen para el análisis de factor común, al diferenciarse entre *varianza común* y *específica*. En la salida SPSS (versión 10.0) figuran en la tabla de *comunalidades*, en la columna correspondiente a las *comunalidades iniciales*. La tabla B corresponde al método de factorización de *ejes principales*. De ella sólo interesa ahora las *comunalidades "iniciales"*, que son independientes del método de extracción de factores comunes utilizado.

Las *comunalidades iniciales* se corresponde con valores R_i^2 (coeficientes de correlación múltiple cuadrados) que se obtienen de regresionar las demás variables respecto de la que se calcula su *comunalidad inicial*. Es como si se realizase un análisis de regresión tomando la variable en cuestión como la dependiente y el resto como las independientes. En ACP la *comunalidad inicial* no es de interés, al ser siempre 1,0, debido a que no se diferencia entre *varianza "común"* y *"específica"*, como se hace en AFC.

En AFC, la *comunalidad inicial* dice la proporción de *varianza* de la variable que es explicada por las demás variables que participan del análisis. Interesan valores elevados, próximos a 1,0, porque indican que la variable es "adecuada" para incluirse en el análisis factorial. Por el contrario, valores próximos a 0,0 pueden llevar a considerar su eliminación del análisis.

Tabla A. Matriz de correlación anti-imagen

	X ₁₅	X ₁	X ₂	X ₃	X ₄	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄
X ₁₅	,873 ^a	-,107	-,008	,052	-,009	,119	-,089	,054	,044	,269	-,031	,029	,125	,056
X ₁		,873 ^a	,107	,033	-,000	,159	-,076	,194	-,037	,017	-,032	-,057	,038	,092
X ₂			,793 ^a	,048	-,091	-,053	,012	-,021	-,113	-,032	-,012	-,104	-,002	-,050
X ₃				,545 ^a	,017	,038	,027	-,031	-,018	,075	-,089	-,086	-,030	-,063
X ₄					,750 ^a	,031	-,041	,000	,000	-,077	,336	,116	-,023	-,096
X ₆						,878 ^a	,098	-,043	-,047	-,025	,093	,110	,002	-,151
X ₇							,865 ^a	,226	,047	,045	-,063	-,013	,055	,078
X ₈								,864 ^a	-,125	-,058	-,010	,057	-,042	-,118
X ₉									,884 ^a	-,041	,045	,007	-,118	-,033
X ₁₀										,776 ^a	-,009	-,066	-,436	-,082
X ₁₁											,719 ^a	-,361	-,016	,052
X ₁₂												,708 ^a	,005	-,012
X ₁₃													,790 ^a	,006
X ₁₄														,895 ^a

^a Medida de adecuación muestral

• X₁₅: "simpatía hacia norteafricano (marroquí...)" (P201); X₁: "leyes inmigración" (P16); X₂: "ideología política" (P39); X₃: "sexo" (P41); X₄: "edad" (P42); X₆: "número de inmigrantes" (P11); X₇: "regularizar a inmigrantes" (P19); X₈: "entrada inmigrantes" (P21); X₉: "partido racista" (P37); X₁₀: "casar con marroquí" (P306); X₁₁: "estudios" (P43a); X₁₂: "ingresos" (P52); X₁₃: "vecino marroquí" (P506); X₁₄: "inmigrante delincuente" (P2904).

sis porque no comparte suficiente varianza con otras variables que permita su agrupación en una misma dimensión *latente*.

Tabla B. Comunidades

	Inicial	Extracción
simpatía marroquí	,316	,380
leyes inmigración	,251	,332
ideología política	8,422E-02	,335
sexo	3,300E-02	3,429E-02
edad	,234	,335
n.º inmigrantes	,246	,304
regularizar inmigrante	,232	,312
entrada inmigrante	,286	,393
partido racista	,140	,155
casar con marroquí	,429	,687
estudios	,351	,611
ingresos	,264	,400
vecino marroquí	,368	,483
inmigrante delincuente	,221	,270

Método de extracción: Factorización de ejes principales.

En la tabla B puede observarse que hay tres variables ("sexo", "ideología política" y "partido racista"), en especial las dos primeras, cuya variabilidad apenas se ve afectada por el resto de variables incluidas en el análisis (el 3%, el 8% y el 14% de su varianza, respectivamente, es explicada por las otras variables). Puede considerarse su eliminación, aunque antes con-

viene comprobar la solución factorial que resulta de su inclusión en el análisis. En cambio, las variables "casar con marroquí" ($F^2 = ,429$), "vecino marroquí" ($F^2 = ,368$) y "estudios" ($F^2 = ,351$) son las tres variables cuya variabilidad puede prevenirse más por el conjunto de las variables consideradas.

En general, las *comunalidades iniciales* son bajas. Ninguna variable llega a sobrepasar, e incluso obtener, la mitad de su varianza determinada por el resto de variables. Además, se advierte que, como las *comunalidades iniciales* se distancian del valor ideal de 1,0, se prevé que la solución de *factor común* (en este caso de *factorización de ejes principales*) diferirá de la alcanzada mediante ACP. Las *comunalidades* de "extracción" se explican en el apartado 5.6.

5.5. La extracción de factores comunes o componentes principales

Una vez comprobado que el análisis factorial puede llevarse a efecto con las variables de interés y que se cumplen los supuestos necesarios para su correcta realización, la segunda fase del proceso de análisis corresponde a la extracción de factores o componentes, depende de la modalidad de análisis factorial que se elija. Esta fase incluye también decisiones clave. En este caso referidas al método de extracción de factores a seguir, además de los criterios a seguir en la determinación del número de factores que compondrán el modelo factorial. Ambos aspectos son relevantes en la resolución del análisis y merecen un tratamiento detallado.

5.5.1. Procedimientos de extracción factorial

Los distintos programas estadísticos ofrecen una amplia variedad de métodos para la extracción de factores. Los más habituales son los siguientes:

- a) Componentes principales
- b) Ejes principales o factor principal
- c) Máxima verosimilitud
- d) Mínimos cuadrados no ponderados o generalizados
- e) Factorización alfa
- f) Factorización imagen

Exceptuando el primer procedimiento (que supone la realización de un ACP), los cinco restantes se aplican en la consecución de un modelo analítico AFC. Dillon y Goldstein (1984: 73) destacan dos procedimientos, el llamado "factor principal", junto al de "máxima verosimilitud", como "los más ampliamente utilizados y estudiados. El primer método es el más antiguo de los dos y frecuentemente se confunde con el análisis de componentes principales. El segundo método es el único para la extracción factorial que corrientemente proporciona una base estadística razonable para comprobar la adecuación del modelo analítico de factor común básico". Para poder valorar estas afirmaciones, se remite al lector a

la lectura de los rasgos definitorios de cada uno de los procedimientos de extracción factorial que a continuación se detallan.

A) *Componentes principales*

A las características definitorias de ACP resumidas en el subapartado 5.2.1 hay que añadir algunas consideraciones relacionadas con el procedimiento de extracción de componentes principales.

ACP, al igual que AFC, trata de la combinación lineal de una serie de variables empíricas correlacionadas en una serie de variables latentes (no observadas) no correlacionadas, que reciben el nombre de *componentes principales*. Pero ambos modelos analíticos difieren en el estudio que hacen de la varianza de las variables. En ACP se analiza toda la varianza, no sólo la compartida como hace AFC. De los *componentes* se pide que extraigan la mayor proporción de varianza de la serie original de datos. El primer componente principal está integrado por aquella combinación lineal de variables que explica la mayor cantidad de varianza en la muestra analizada. El segundo componente reúne la combinación lineal de variables que explica la mayor proporción de varianza *residual* (la que queda sin explicar por el primer componente principal), con la condición de que no esté correlacionada con la primera combinación de variables que forman el primer componente. Los componentes posteriores también han de ser *ortogonales* (no correlacionados entre sí) y secuencialmente explican cada vez menos proporción de varianza residual.

En principio, en ACP pueden extractarse tantos componentes como variables observadas. Pero, si se quiere que el modelo analítico cumpla los objetivos de *parsimonia* y simplicidad, deberá contener un número de componentes inferior al de variables observadas. Su número lo decide el porcentaje de varianza de las variables empíricas que logren explicar, además de otras consideraciones resumidas en el subapartado 5.5.2.

Por último, añadir que ACP puede aplicarse de forma aislada o en conjunción con AFC. Varios autores (como Tabachnick y Fidell, 1989; o Afifi y Clark, 1990) recomiendan la aplicación de ACP como paso previo a AFC. Este proceder favorece que el investigador disponga de información previa sobre varianza compartida, número de factores y su naturaleza, que será de interés para indagaciones posteriores.

B) *Ejes principales o factor principal*

“En términos de métodos de extracción de factores iniciales, la solución de factor principal iterativa es el método empleado con más frecuencia por los científicos sociales, los principales usuarios del análisis factorial” (Afifi y Clark, 1990: 408). De hecho, “la mayoría de los tratamientos de análisis factorial identifican el modelo de factor común mediante un procedimiento de factorización de eje principal” (Kim y Mueller, 1978b: 21).

A diferencia de ACP, ahora sólo se analiza la varianza común o compartida, lo que incide en el empleo de la matriz de *correlación reducida*. La diagonal principal de es-

ta matriz se caracteriza por estar compuesta de las estimaciones iniciales de la *comunalidad* de cada variable. Estas *comunalidades* se derivan mediante un procedimiento iterativo con las correlaciones múltiples cuadradas de cada variable con todas las demás variables empíricas, empleadas como valores de partida de la interacción. Es decir, como algoritmo en la estimación de los factores comunes.

A partir de la matriz de *correlación reducida* se procede a la extracción de factores. La condición que se impone es que sea máxima la contribución de cada factor a la *comunalidad* total. Primero se elige el eje sobre el cual la variabilidad de las proyecciones de los datos es máxima. Después, se escoge el segundo eje en el que la variabilidad residual (o restante) de la proyección sea máxima. Y así, sucesivamente.

“Una regla a considerar es extraer factores hasta que la suma de los autovalores se aproxime a la comunalidad total” (Dillon y Goldstein, 1984: 74). La mayoría de los programas estadísticos generan coeficientes factoriales (λ_p) para cada factor de manera que la contribución de cada factor a la *comunalidad* total sea la mayor posible.

En suma, este segundo método de extracción de factores difiere de componentes principales en que factoriza la *comunalidad* total que, obviamente, es inferior a la varianza total. Por esta razón, no sorprende que el valor de la *comunalidad* total en una salida de ACP sea superior al obtenido en la extracción de *factor principal*. Asimismo, se observa que los coeficientes factoriales (o “factor loadings”) calculados mediante este segundo procedimiento dependen, a diferencia de componentes principales, del número de factores extraídos.

Además, como el análisis factorial de *ejes principales* parte de la matriz de *correlación reducida* (o ajustada), las *comunalidades* que componen su diagonal principal pueden ser considerablemente inferiores a 1,0. En consecuencia, cabe la posibilidad de obtener *autovalores* negativos, a diferencia de ACP. Ante estos *autovalores* negativos el proceder habitual es incluirlos en los análisis, junto a los factores a ellos relacionados.

Las diferencias destacadas entre ACP y el análisis de *factor principal* se agudizan cuanto más se distancien los valores de *comunalidad* de 1,0. Si las *comunalidades* se aproximan en magnitud a la unidad, la distancia entre ambos modelos se minimiza. Lo mismo acontece cuando el número de factores extractados mediante el procedimiento de *eje principal* se aproxima al número de variables empíricas (“p”). En ambas situaciones se obtienen resultados similares realizando ACP y un análisis de *factor principal*.

C) *Máxima verosimilitud*

Este tercer procedimiento de extracción factorial tiene su desarrollo inicial en los años cuarenta, de la mano de Lawley. De los dos anteriores difiere en dos aspectos fundamentales:

1. Requiere que las “p” variables empíricas cumplan el supuesto de *normalidad*. Ni ACP ni el método de *factor principal* precisan, para su ejecución, de este supuesto. Hecho que incide en que el procedimiento de *máxima verosimilitud* se

convierta en una opción deseable, sólo cuando se esté ante distribuciones normales.

2. El cumplimiento del supuesto de *normalidad* permite la inferencia estadística: las pruebas de hipótesis y la estimación de los parámetros poblacionales (intervalos de confianza), a partir de las estimaciones muestrales. Ésta es su principal ventaja frente a los demás procedimientos de extracción factorial.

Para ello se aplica una prueba de *chi-cuadrado* propuesto por Barlett en 1954, con " $\frac{1}{2} [(p - k)^2 - p - k]$ " grados de libertad (siendo "p" el número de variables empíricas y "k" el número de factores). Mediante dicha prueba se comprueba si el modelo factorial obtenido en la muestra logra la configuración exacta de los parámetros poblacionales. Es decir, si puede generalizarse a la población a la que pertenece la muestra. Para este propósito se comparan las correlaciones calculadas en la muestra con las reproducidas a partir de los coeficientes factoriales de cada variable empírica en cada factor del modelo factorial.

La hipótesis *nula* afirma que se ha logrado extraer toda la varianza poblacional mediante los factores obtenidos en la muestra. La hipótesis *alternativa* es la negación de la anterior. Si a un nivel de probabilidad específico (dígase del usual $\alpha = ,05$ o el más restrictivo $\alpha = ,01$), el valor χ^2 empírico resulta significativo (superior al correspondiente valor χ^2 teórico), se rechaza la hipótesis *nula*. Lo cual significa que las predicciones a partir del modelo con los factores extractados no son buenas estimaciones de las correlaciones observadas entre las variables. Se debería incrementar el número de factores en el modelo para lograr reproducir adecuadamente las correlaciones entre las variables originales. Esto suele ocurrir cuando en la matriz *residual* aún queda varianza significativa no explicada por algún factor.

En este método factorial es habitual incrementar el número de factores hasta alcanzar un buen ajuste del modelo a los datos, a un nivel de significación $\geq ,05$. Este proceder se convierte en su peor crítica: el confiar únicamente en la prueba de significatividad puede llevar a la consecución de un modelo de factor común con más factores de los deseables. Esto es más probable que suceda en tamaños muestrales elevados, que suelen ir acompañados de valores χ^2 significativos. En palabras de Kim y Mueller (1978b: 42), "la aplicación del método ha mostrado que para muestras grandes con muchas variables el número de factores retenidos tiende a ser mucho mayor que el número de factores que el investigador desea aceptar". Ante ello se recomienda (Harman, 1976) no confiar exclusivamente en el test χ^2 . Sólo debería considerarse el número de factores obtenidos mediante la prueba χ^2 como referente del número máximo de factores a extraer. Únicamente se retendrán aquellos factores que sean interpretables teórica y sustantivamente. Preferiblemente, después de la *rotación* de la matriz factorial, como se verá en el subapartado 5.6.1.

El objetivo principal del método de *máxima verosimilitud* es encontrar aquella configuración exacta de los parámetros poblacionales que maximice:

1. La probabilidad de producir la matriz de correlación observada.
2. La correlación canónica entre los k factores y las p variables empíricas.
3. El determinante de la matriz de correlación parcial.

Dependiendo de qué aspecto se prime, se estará ante una u otra variedad del procedimiento de *máxima verosimilitud*: la *factorización canónica de Rao* o actuaciones dirigidas a incrementar los *determinantes* de la matriz de *correlación parcial residual*.

De nuevo, la consecución de un modelo factorial resulta del empleo de un algoritmo *iterativo*. Además, las correlaciones se ponderan por el inverso de la varianza *única* de las variables, obteniéndose un modelo factorial distinto del alcanzado mediante otros procedimientos.

Aunque también se consideran las correlaciones múltiples cuadradas como estimaciones de las *comunalidades* iniciales, dichas estimaciones de comunalidad suelen ser inferiores a las alcanzadas mediante otros métodos factoriales. Las varianzas *únicas* (tratadas como varianzas de “cuasi-error”) serán más elevadas. Los factores comunes suelen explicar una proporción inferior de la varianza entre todas las variables empíricas. Y, por último, los pesos o coeficientes factoriales serán, en general, más elevados en aquellos indicadores (o variables empíricas) de mayor *comunalidad* y, por tanto, de menor varianza *única*.

En resumen, el análisis factorial de *máxima verosimilitud* permite, frente a otros procedimientos factoriales, comprobar la bondad de ajuste de un modelo factorial concreto a una matriz muestral de varianza o de correlación. En su contra está la necesidad de especificar el número de factores a extraer. Este hecho incide en su mayor adecuación a la variedad factorial *confirmatoria*, aunque no restringe su aplicación a la *exploratoria*.

D) *Mínimos cuadrados no ponderados y generalizados*

Los procedimientos factoriales de *mínimos cuadrados* tuvieron un desarrollo inicial en los años sesenta. Primero, de la mano de Comrey, en un artículo publicado en 1962 (“The minimum residual method of factor analysis”, *Psychological Reports*) y, posteriormente, con la aportación de Harman y Jones de 1966 (“The factor analysis by minimizing residuals (Minres)”, *Psychometrika*, 31: 351-368).

Su actuación se dirige a *minimizar* la correlación residual (de ahí le viene su denominación de Minres: residuo mínimo), después de haber extractado un número determinado de factores y de asegurar el grado de ajuste entre las correlaciones cuadradas observadas en la muestra y las reproducidas mediante el modelo factorial. Se analizan las diferencias entre ambas matrices, o se estudia la que mejor reproduzca la matriz de *covarianzas* original (si el modelo de factor común se ha obtenido a partir de la matriz de varianzas-covarianzas).

Los análisis comienzan con la hipótesis de un único factor común para, a continuación, ir aumentando el número de factores hipotéticos hasta que se alcance una solución satisfactoria: cuando los factores logren explicar la mayoría de las correlaciones

observadas. También se comprueba la bondad de ajuste mediante la prueba χ^2 , como en el procedimiento de *máxima verosimilitud*.

La extracción factorial de *mínimos cuadrados* incluye dos variedades básicas:

1. *Mínimos cuadrados no ponderados*. Consiste en la obtención de una matriz factorial para un número específico de factores, de manera que sea mínima la suma de las diferencias al cuadrado entre los elementos (excluyendo los que componen la diagonal principal de la matriz) de las matrices de correlación observada y la reproducida a partir de los factores extractados. Mediante este procedimiento factorial suelen alcanzarse resultados similares al análisis de *factor o ejes principales*, cuando se tienen las mismas *comunalidades*. Compruébese en los ejemplos del subapartado 5.6.2.

Como en los análisis de *factor principal* y de *máxima verosimilitud*, en la factorización de *mínimos cuadrados no ponderados* las estimaciones iniciales de las *comunalidades* coinciden con las correlaciones múltiples cuadradas de cada variable con las demás variables observadas. Después, se procede a su reestimación, considerando el modelo factorial obtenido en las *fases iniciales* del análisis. Por lo que, nuevamente, se está ante una estimación iterativa de las *comunalidades*. Se calculan los pesos o *coeficientes factoriales* (λ_{ij}) de cada variable empírica en cada factor común para, a partir de estos valores, calcular nuevas *comunalidades* que reemplacen a las iniciales. El proceso de estimación iterativa de las *comunalidades* prosigue hasta que no se observen variaciones en los valores de las *comunalidades* en las dos últimas iteraciones (de la última respecto a las calculadas en el paso precedente).

2. *Mínimos cuadrados generalizados*. Aunque comparte con el procedimiento anterior el mismo criterio genérico de extracción factorial, difiere de él precisamente porque *pondera* las correlaciones inversamente por la *varianza única* de las variables. De manera que, a las correlaciones de las variables con elevada *varianza única* se les da un peso factorial inferior al obtenido por las correlaciones de las variables con baja *varianza única*.

E) *Factorización alfa*

“Puede verse o como una variante del modelo de factor común o como una estrategia alternativa” (Kim y Mueller, 1978b: 11). Su desarrollo principal lo alcanza con Kaiser y Caffrey en 1965 (“Alpha factor analysis”, *Psychometrika*), en indagaciones en el campo de la psicometría. En concreto, en el interés por descubrir factores comunes consistentes cuando se toman muestras repetidas de una misma población (de variables).

A diferencia de los demás procedimientos de extracción factorial, en la *factorización alfa* las variables se consideran una muestra del universo de variables. En los demás métodos factoriales se analiza una muestra de individuos, no de variables, al asumirse que éstas constituyen el universo. Por esta razón, en la *factorización alfa* no se

aplica ninguna prueba de significatividad al modo usual, como se hace en los métodos de *máxima verosimilitud* y de *mínimos cuadrados*. Se considera que el análisis incluye la población de individuos y no una muestra de ellos. La lógica es, por tanto, opuesta al resto de los métodos factoriales.

En la *factorización alfa* se busca maximizar la generalidad de los factores. Los pesos o coeficientes factoriales se determinan de manera que los factores comunes extraídos tengan correlaciones máximas con los factores comunes correspondientes, que se asume existen en el universo.

Con el método de *máxima verosimilitud* comparte la particularidad de que las estimaciones factoriales son independientes de la escala de medición. Pero, difiere de él (al igual que de otros procedimientos factoriales) en que proporciona un coeficiente factorial (λ_{ij}) más elevado a las variables con menor *comunalidad*.

Las *comunalidades* para cada variable no resultan de la suma de los coeficientes factoriales al cuadrado (λ_{ij}^2) en cada factor común. Igualmente, los *autovalores* no se obtienen de la suma de los coeficientes factoriales al cuadrado de todas las variables en cada factor común. En la *factorización alfa* las *comunalidades* se estiman mediante procedimientos iterativos que maximizan el coeficiente *alfa* para los factores.

El *coeficiente alfa* es una medida derivada en psicometría para comprobar la fiabilidad (o capacidad de obtener resultados consistentes en una medición) de una puntuación tomada en una variedad de situaciones. Mide la consistencia interna de las variables empíricas (de forma global e individual), en la configuración de una misma dimensión de un concepto teórico. Se calcula a partir de la matriz de varianzas-covarianzas. Su rango de valores posibles va de 0,00 (infiabilidad) y 1,00 (fiabilidad perfecta). Un valor α igual o superior a 0,80 indica que la medida es fiable.

F) Factorización imagen

Recibe este nombre (“imagen”) por la peculiaridad del procedimiento de análisis. En él se distribuye, entre los factores comunes, la varianza de cada variable observada que es “reflejada” por el resto de las variables consideradas. La parte *común* de cada variable (aquella que puede predecirse mediante la combinación lineal de las demás variables en la serie) se llama la “imagen de la variable”. Su cuadrado equivale a la *comunalidad* de una variable como se define en AFC: la correlación múltiple cuadrada entre una variable y el resto. En cambio, la parte *única* de la variable, la no predecible mediante la combinación lineal de las demás variables, recibe el nombre de “anti-imagen”. Su cuadrado equivale a la varianza *única*.

En este procedimiento factorial el principio que rige la extracción de factores es el de la *covarianza* de las estimaciones de cada variable regresionada en las demás variables; o sea, como si se efectuase un análisis de regresión lineal de dicha variable considerando el resto de las variables. En consecuencia, interesa la parte de la varianza de cada variable que es compartida con las otras variables. De ahí su clasificación dentro de los procedimientos de AFC.

Con ACP comparte la particularidad de ofrecer una solución matemáticamente única, al existir valores fijos en la diagonal principal de la matriz de correlación. Pero, como en todo AFC, la diagonal de la matriz se halla integrada por las *comunalidades* de cada variable. Éstas se obtienen, igualmente, de las correlaciones múltiples cuadradas de cada variable con el resto. Si bien difiere de los otros procedimientos de factor común en que los valores de las *comunalidades* no cambian; es decir, no son reestimados *iterativamente*, sino que permanecen fijos. Éste es el nexo que le une con ACP: el ofrecer soluciones matemáticamente únicas.

En la *factorización imagen* tampoco se aplica ninguna prueba de significatividad. Se considera que se abarca al universo de variables y a la población de individuos. Aunque sus estimaciones logran mayor consistencia conforme aumenta el número de casos analizados y se incrementa el número de variables.

- Mediante cualquiera de los seis procedimientos de extracción factorial referidos se pueden, en general, alcanzar resultados similares cuando:

- a) Aumenta el número de casos y de variables en el análisis.
- b) Las *comunalidades* se aproximan a 1,00.

En ambas situaciones la elección de uno u otro procedimiento factorial adquiere menor importancia. De manera especial –siguiendo la propuesta de Gorsuch (1974)–, cuando el número de variables supera en cuantía a 30 y las *comunalidades* son mayores de 0,40.

5.5.2. *Criterios de selección del número de factores*

Otra decisión clave en el análisis factorial concierne al número de factores comunes o componentes principales a retener en el modelo factorial. Del modelo se quiere que explique la mayor proporción de *varianza* (ACP) o de *covarianza* (AFC) de los datos analizados. Este objetivo indudablemente se alcanza cuando el número de factores iguala al de variables observadas. Aunque, procediendo de esta manera no se logra la finalidad básica del análisis: la reducción de la información contenida en una serie amplia de variables en un número inferior de *factores* o *componentes*.

De un modelo factorial se pide que sea *simple* y *parsimonioso*. Ha de contener un número bajo de factores, pero suficiente para explicar las correlaciones habidas entre las variables observadas. Autores como Frei y Ruloff (1989: 307) proponen que el número de factores no sea “en ningún caso superior a la mitad del número de indicadores”.

Pero de los factores también se exige que sean significativos e interpretables. Todos estos aspectos han de considerarse en la decisión del número de factores que compondrán el modelo factorial. Concretamente, los criterios más seguidos pueden resumirse en cinco:

- a) Autovalores (o varianza total explicada por cada factor)
- b) Porcentaje de varianza total atribuible a cada factor
- c) El gráfico de sedimentación (o "scree test")
- d) Significatividad
- e) Interpretabilidad

Estos cinco criterios suelen valorarse conjuntamente, y no de forma aislada. La decisión sobre el número de factores normalmente se fundamenta en la convergencia de la mayoría de ellos. En los paquetes estadísticos al uso esta información se da en una tabla inicial, posterior a las matrices de partida (de *correlaciones* o de *varianzas-covarianzas*), y anterior a la matriz factorial. En el programa SPSS la tabla es como se ilustra en el ejemplo. En versiones anteriores del SPSS, en la tabla aparecían las *comunalidades* iniciales para cada variable (1,000 en ACP y las correlaciones múltiples al cuadrado de cada variable respecto a las demás en AFC). En las últimas versiones de SPSS (versión 7.5 y posteriores) los valores de *comunalidad* se recogen en una tabla propia. Ésta incluye las *comunalidades* iniciales y las resultantes de la extracción factorial. También se ofrece una tabla de *estadísticos descriptivos univariados* que muestra el número de observaciones válidas, junto a la *media* y *desviación típica* de cada una de las variables observadas. Esta última información no es específica al análisis factorial, sino compartida con otras técnicas analíticas como el análisis de regresión. Por esta razón se decide no volver a extraerla en el texto, sino remitir al lector interesado al subapartado 1.3.2 para su observación.

Una tabla que sí es específica del análisis factorial es la relativa a la *varianza total explicada* por cada factor o componente. Incluye los *autovalores* iniciales y las sumas de las saturaciones al cuadrado de la extracción previo y posterior a la *rotación* de los ejes factoriales. De cada uno de ellos se ofrece el porcentaje de la varianza y el acumulado para cada factor o componente. La información relativa a los *autovalores* iniciales protagoniza la decisión del número de factores o componentes a extraer, como se expone seguidamente.

A) Autovalores

El criterio más seguido en la determinación del número de factores responde al "autovalor" de cada factor o componente. Por *autovalor* ("eigenvalue") se entiende la cantidad de varianza explicada por cada factor común o componente principal. También se le llama "raíz característica o latente". Su tamaño describe la dispersión de los datos en un espacio multivariable que incluye un eje para cada variable observada. Su suma equivale, en ACP, al número de variables empíricas y su producto al determinante de la matriz de correlación. En el ejemplo siguiente puede verse cómo se disponen (en la tabla de varianza total) en orden decreciente: de mayor a menor tamaño. Al primer factor le corresponde el *autovalor* de mayor cuantía y al último factor el de menor.

El empleo de los *autovalores* como criterio de extracción del número de factores fue propuesto por Kaiser en varios artículos de 1958 ("The varimax criterion for

analytic rotation in factor analysis", *Psychometrika*, 23) y 1959 ("The application of electronic computers to factor analysis", *American Psychological Association*). Su aplicación varía si el modelo básico es ACP o AFC.

En ACP, el punto de corte se fija en 1,00. Todo componente que presente un *autovalor* $\geq 1,00$ formará parte del modelo factorial. De él se excluirán aquellos componentes cuyo *autovalor* sea inferior a 1,00, porque ni siquiera logran explicar la *varianza* de una variable. Recuérdese que en ACP se analiza la *varianza* total de una variable, sin diferenciar entre *varianza común* y *específica*. Las variables suelen estar estandarizadas, siendo su *varianza* total igual a 1. La suma de los *autovalores* de todas las variables observadas será, por tanto, igual al número total de variables y su promedio igual a 1,00. De ahí que sea éste el valor de referencia adoptado. Además, cada *autovalor* puede expresarse como una proporción de esta suma. Según estimaciones de Tabachnick y Fidell (1989: 635), "el número de componentes con *autovalores* mayores de 1 suele estar en algún lugar entre el número de variables dividido por 3 y el número de variables dividido por 5 (por ejemplo, 20 variables deberían producir entre 7 y 4 componentes con *autovalores* mayores de 1)".

En AFC interesa sólo la *varianza común* de las variables, hecho que revierte en la determinación del punto de corte que decide qué factores comunes formarán el modelo factorial. En AFC cada *autovalor* indica la proporción de *varianza* extractada por el factor común en relación con la *varianza "común"* total de las variables. Su suma expresa la *varianza común* total de las variables. Ésta, obviamente, no será igual al número total de variables, a diferencia de ACP. Su valor será tanto más bajo cuanto menos *varianza* compartan las variables analizadas. Asimismo, el punto de corte no será 1,00, sino un valor inferior. El interés no está en explicar la *varianza* total de una variable, sino sólo la compartida con otras variables. El punto de corte se obtiene del promedio de las estimaciones de las *comunalidades* iniciales de todas las variables.

También puede seguirse la recomendación de Harman (1976) de parar la extracción del número de factores antes de que la suma acumulada de los *autovalores* exceda la suma de las *comunalidades* estimadas. Estas *comunalidades* figuran en la diagonal principal de la matriz de *correlación reducida* \bar{R} . Además se encuentran en la tabla de *comunalidades*, que incluye las "iniciales" y las resultantes de la "extracción" factorial.

El criterio del *autovalor* se considera "sencillo" (Kim y Mueller, 1987b: 43) y suele funcionar bien. Normalmente proporciona resultados consistentes con las expectativas del investigador. Pese a ello, existen voces críticas a la aplicación única de este criterio. Afifi y Clark (1990) afirman que los *autovalores* están sujetos a grandes variaciones muestrales, al ser estimaciones de *varianza* de cada factor o componente. Por su parte, Duntzman (1989: 22) alerta —en referencia a ACP, pero también extensible a AFC— del peligro de "desconsiderar componentes principales que, aunque pequeños, pueden ser importantes", si se aplica rígidamente el criterio de Kaiser. Esto sucede cuando una o más variables no quedan suficientemente bien representadas por los componentes o factores con *autovalores* más elevados y sí, en cambio, por aquellos más bajos.

En la literatura también existen propuestas de cuándo este criterio es más operativo y fiable. Para Hair *et al.* (1992: 237), cuando el número de variables se halla comprendido entre 20 y 50. “En casos donde el número de variables es inferior a 20 existe la tendencia de que mediante este procedimiento se extraiga un número conservador de factores; cuando se hallan implicadas más de 50 variables no es inusual que se extraigan demasiados factores”. Para Tabachnick y Fidell (1989: 635) el criterio de *autovalor* es “razonable, si el número de variables es 40 o menos, y si el tamaño muestral es grande”. Cuando coinciden ambos supuestos, el número de factores determinado siguiendo este criterio será el correcto. En caso contrario, se estará o sobrestimando, o subestimando, el número real de factores en la serie de datos. En ésta, como en otras cuestiones, existen opiniones dispares. La práctica investigadora, como es usual, se impondrá en esta polémica.

B) Porcentaje de la varianza total atribuible a cada factor

Un criterio alternativo y equivalente al anterior es el porcentaje de la varianza total que es explicada por cada factor común o componente principal. Se obtiene de la división de cada *autovalor* por la varianza total. Ésta es igual a la suma de los elementos de la diagonal principal de la matriz de correlación. En ACP, la varianza total es igual al número total de variables; en AFC, a la suma de las *comunalidades* iniciales de cada variable. Pero más que el porcentaje de varianza explicada por cada factor específico, interesa el porcentaje *acumulado* de varianza de factores sucesivos. Éste expresa el porcentaje de varianza que lograría explicarse con la incorporación de ese factor o componente y los precedentes.

A diferencia del criterio de *autovalor*, en éste no se ha establecido ningún punto de corte absoluto que determine el número de factores a incluir, o desconsiderar, en el modelo factorial. El procedimiento presenta cierta arbitrariedad en la decisión de qué proporción de la varianza total es suficiente para dejar de añadir factores al modelo. La decisión normalmente depende, como destacan Batista y Martínez (1989: 53) de: el número de variables originales, la magnitud de las correlaciones de R, la significatividad estadística de los factores y, sobre todo, del conocimiento y experiencia que el investigador posea de las variables consideradas.

No obstante, se han pronunciado algunas propuestas de porcentajes acumulados de varianza mínimos para tomar como referentes. Afifi y Clark (1990) proponen que como mínimo se explique, con factores sucesivos, el 80% de la varianza total. Hair *et al.* (1992; 1999) diferencian entre ciencias “duras” y ciencias sociales. Para las “ciencias duras” sugieren que la extracción factorial no concluya hasta que se logre al menos explicar el 95% de la varianza total con los factores retenidos, o hasta que el último factor explique sólo una pequeña proporción (menos del 5%). En las “ciencias sociales” el porcentaje mínimo desciende hasta el 60% de la varianza total y, a veces, incluso menos. Esta apreciable reducción porcentual la justifican por la “menor precisión” de la información que suele analizarse en las ciencias sociales.

EJEMPLO DE APLICACIÓN DE CRITERIOS DE VARIANZA (AUTOVALORES) EN LA DECISIÓN DEL NÚMERO DE FACTORES COMUNES O DE COMPONENTES PRINCIPALES A EXTRAER

En la salida del programa SPSS, los *autovalores* "iniciales" figuran en la tabla conjunta llamada "varianza total explicada", junto a las sumas de las saturaciones al cuadrado previo y posterior a la *rotación* factorial. La tabla siguiente corresponde al procedimiento de *componentes principales*, la variedad analítica factorial más popular y de más fácil realización. Recuérdese que una práctica usual en la investigación social es realizar un ACP para determinar en cuántas dimensiones latentes pueden agruparse las variables de interés. Después, puede realizarse un AFC conociendo ya el número de factores a extraer.

Las tablas de "varianza total explicada" en AFC proporcionan la misma información sobre los *autovalores* iniciales (la variabilidad previa a la extracción factorial) que en ACP. Por esta razón se evita su exposición repetida. En lo que difieren es en nombrar "factores" y no "componentes" y, obviamente, en las sumas de las saturaciones al cuadrado. Pero, como en la decisión sobre el número de factores o componentes a retener el protagonista común son los *autovalores*, la exposición se limita a su intervención en esta fase del proceso de análisis. Lo referido a las sumas de las saturaciones se posterga al apartado 5.6. En dicho apartado sí se ejemplifican distintos métodos de extracción factorial, al constatarse diferencias en sus resultados. En general, los distintos procedimientos factoriales van a describirse sólo cuando sus resultados no coincidan con ACP.

Varianza total explicada

Compon.	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	3,619	25,849	25,849	3,619	25,849	25,849	2,416	17,260	17,260
2	1,665	11,891	37,740	1,665	11,891	37,740	2,046	14,611	31,871
3	1,114	7,958	45,698	1,114	7,958	45,698	1,861	13,290	45,162
4	1,015	7,253	52,951	1,015	7,253	52,951	1,090	7,789	52,951
5	,950	6,786	59,737						
6	,868	6,199	65,936						
7	,752	5,372	71,308						
8	,736	5,259	76,567						
9	,673	4,805	81,373						
10	,600	4,283	85,656						
11	,574	4,103	89,758						
12	,559	3,994	93,752						
13	,470	3,361	97,113						
14	,404	2,887	100,000						

Método de extracción: Análisis de componentes principales.

La tabla recoge tantos *componentes principales* como variables, aunque se quiere reducir sensiblemente su número. En ACP el criterio más aplicado es que los componentes al menos

presenten un *autovalor* igual a 1. Del componente se pide que explique más variabilidad que la correspondiente a una variable. Recuérdese que el *autovalor* indica la cantidad de *varianza* total explicada por cada componente principal. Como las variables están expresadas en su forma estandarizada (para simplificar la información y, sobre todo, para evitar la influencia indebida de las unidades de medición en la ponderación de los componentes), su *varianza* es 1. Al ser 14 las variables analizadas, la *varianza total* es 14. Puede constatarse que la suma de todos los *autovalores* de la tabla es igual al número de variables. Exactamente, $3,619 + 1,665 + 1,114 + \dots + ,404 = 13,999$ (14). Asimismo, el producto de todos los *autovalores* ($3,619 \times 1,665 \times 1,114 \times \dots \times ,404$) es igual al *determinante* de la *matriz de correlación*: ,07667.

Los *autovalores* figuran ordenados por su tamaño. El primer componente se halla integrado por la combinación lineal de variables que más *varianza* explica: 3,619. Respecto a la variabilidad total (13,999), se traduce a un 25,85% de *varianza* explicada ($3,619 / 13,999 \times 100$). En cambio, el último componente apenas extrae *varianza* de las variables originales: ,404; que, en relación con la *varianza* total, tan sólo representa el 2,9%. En general, los *autovalores* de los últimos factores o componentes se aproximan a cero, cuando las *correlaciones* entre las variables son muy elevadas. Esta situación no se da en este ejemplo. Las *correlaciones* son leves, a excepción de cuatro variables, a decir por sus *comunalidades* "iniciales", como se vio en el subapartado 5.4.3.

La tabla corrobora asimismo la afirmación de Tabachnick y Fidell (1989) de que en ACP el número de componentes con *autovalores* mayores de 1 suele ser un número comprendido entre el número total de variables dividido entre 3 y dividido entre 5. En este caso, $14 / 3 = 4,667$ y $14 / 5 = 2,8$. Entre 5 y 3 componentes estaría la decisión. En realidad, son 4 los componentes a extraer porque presentan *autovalores* mayores o iguales a 1. El primer componente explica el 25,85% de la *varianza* total de las variables mientras que el cuarto componente sólo explica el 7,25%. En total, con la combinación de las 14 variables en 4 componentes logra explicarse el 52,95% de la *varianza* (porcentaje acumulado, que indica el porcentaje de *varianza* atribuible a un factor o componente y a aquellos que le preceden en la tabla). Este porcentaje es importante en magnitud, aunque ligeramente inferior al mínimo ideal en las ciencias sociales: el 60% de la *varianza* total. Si bien, distintos autores (Hair *et al.*, 1992; 1999) reconocen que la "menor precisión" de la información en las ciencias sociales influye en que se acepten soluciones factoriales que logren explicar incluso un menor porcentaje de *varianza* que el aquí obtenido. Explicar el 60% de la *varianza* total supone incrementar a 5 el número de componentes. Pero antes de aumentar el número de componentes, hay que valorar otros criterios que después se exponen. Aquí sólo se señala que el componente 5 presenta un *autovalor* próximo a 1 (,950), permitiendo su incorporación al modelo factorial.

C) El gráfico de sedimentación (o "scree test")

En 1966 Catell propone (en "The scree test for the number of factors", *Multivariate Behavioral Research*, 1: 245-276) el empleo de un gráfico, que denomina "scree", en la decisión de cuántos factores incluir en un modelo factorial. En este gráfico se representan los *autovalores* correspondientes a cada factor en el eje vertical, y los factores o componentes, en su orden de extracción, en el eje horizontal. De la conjunción de cada

factor con su correspondiente *autovalor* resulta una curva decreciente que conecta puntos sucesivos. La curva siempre es decreciente, en consonancia con la disposición de los *autovalores*. Al primer factor o componente siempre le corresponde el *autovalor* más elevado, mientras que el último factor se caracteriza por tener el de menor cuantía.

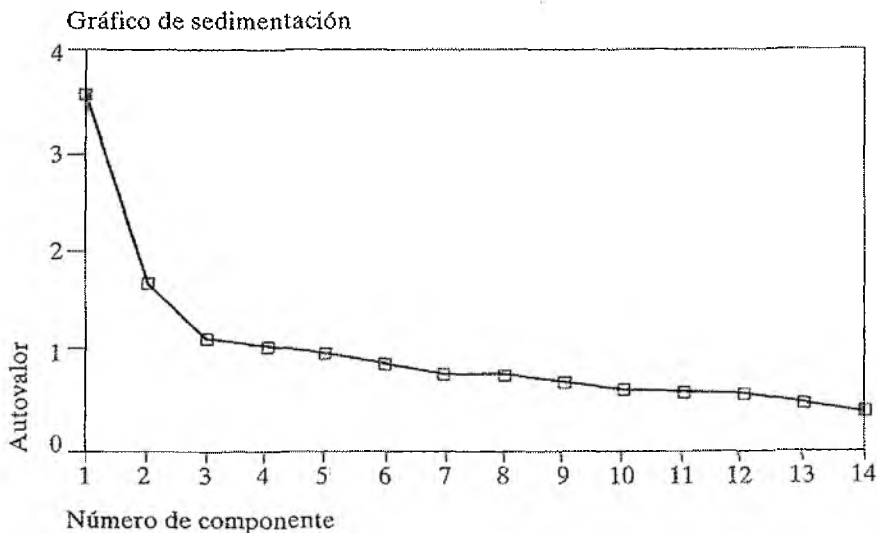
Catell escoge el término, aplicado en geología, “scree” (traducido al castellano como “desmoronamiento”, “escombros”...) por la semejanza del gráfico a los escombros que se forman al pie, o en la parte inferior, de una pendiente de una montaña rocosa. Los *autovalores* grandes forman el “acantilado” y los pequeños los “escombros” o cascotes.

De acuerdo con este criterio, el número de factores o componentes está delimitado por el punto de inflexión de la trayectoria de caída de la pendiente. Es decir, cuando la pendiente descendiente comienza a nivelarse. Catell sugiere que se tomen todos aquellos factores o componentes situados antes del punto de inflexión.

En la práctica, la aplicación de este criterio supone la incorporación al modelo factorial de 1 y, a veces, incluso 2 o 3 factores más de los seleccionados siguiendo exclusivamente el criterio de los *autovalores* (Hair *et al.*, 1992; 1999).

EJEMPLO DE GRÁFICO DE SEDIMENTACIÓN

El recurso al gráfico de *sedimentación* adquiere mayor relevancia cuando varios *autovalores* se aproximan a 1, como sucede en el ejemplo presente, y quiere comprobarse si se pueden extraer menos componentes o factores de los inicialmente elegidos siguiendo el criterio de los *autovalores*.



Este gráfico, también llamado de la “varianza total asociada a cada factor”, interviene igualmente en la decisión del número de componentes o factores a retener. Lo indica el punto de inflexión de la curva descendente, donde ésta comienza a nivelarse. La nivelación empieza a producirse, en este ejemplo, en el punto correspondiente al *autovalor* del tercer componente (1,114), muy próximo al *autovalor* del componente cuatro (1,015). Esto significa que la aplicación de la prueba de “desmoronamiento” (o “scree test”) lleva a la solución de agrupar las 14 variables en 3 componentes principales. En contra de la afirmación de Hair *et al.* (1992; 1999), su aplicación en los datos aquí analizados no incrementa el número de componentes, sino lo reduce.

Este tercer criterio de selección de factores es calificado negativamente de “arbitrario” y “subjetivo”. Kaiser (1970) lo considera “ambiguo” además de “subjetivo”. Estos calificativos se deben al hecho de que puede haber más de una interpretación de la trayectoria de caída, e, incluso la misma consideración de la pendiente como “pronunciada” o “no pronunciada” puede ser “arbitraria”. Dillon y Goldstein (1984) comparten estas críticas y advierten, además, de algunas complicaciones que pueden suscitar la práctica de este criterio, aparentemente sencillo. Puede que no haya ningún punto de inflexión claro u obvio, en cuyo caso esta prueba no sería concluyente. Pero, también, puede darse la situación contraria: la existencia de varios puntos de inflexión (por ejemplo, en la primera mitad de los autovalores), que provoquen indecisión sobre cuál de los puntos de inflexión refleja el número correcto de factores que compondrán el modelo factorial.

En caso de duda, la recomendación usual es efectuar varios análisis factoriales especificando, cada vez, un número de factores diferentes. Si los *residuos* (que figuran en la matriz de correlación residual) son pequeños, el análisis factorial realizado se considera bueno. La presencia de varios *residuos* moderados (de ,05 o ,10) o grandes (> ,10) sugiere la necesidad de incorporar otro factor más al modelo analítico (Tabachnick y Fidell, 1989).

La aplicación del criterio de *desmoronamiento* mejora, no obstante, cuando confluyen varios elementos: tamaño muestral elevado, altas *comunalidades* y *coeficientes factoriales* próximos a 1 en varias variables en cada factor o componente (Gorsuch, 1983). Cuando coinciden estas situaciones, es más probable que su aplicación resulte menos “ambigua” y, en consecuencia, más clara.

D) Significatividad

El criterio de *significatividad* estadística sólo se aplica cuando el procedimiento factorial es de *máxima verosimilitud* o de *mínimos cuadrados*. Como ya se dijo en el subapartado 5.5.1, ambos procedimientos de extracción factorial aplican, para comprobar la significatividad del modelo, la prueba de contraste *chi-cuadrado* con “ $1/2 [(p - k)^2 - p - k]$ ” grados de libertad. Su uso suele relacionarse con la obtención de un núme-

ro de factores superior al indicado por otros criterios. De manera especial, cuando el tamaño de la muestra es elevado y se incluyen muchas variables en el análisis

También se puede hacer uso del *test de Barlett*, en la evaluación de los factores, de manera conjunta y por separado. El procedimiento es similar al descrito en el subapartado 5.4.3. La hipótesis nula (H_0) niega la existencia de factores en la serie de datos analizada, mientras que la *alternativa* (H_1) la afirma. Para dicho contraste de hipótesis se acude, asimismo, a la distribución *chi-cuadrado*, pero con " $1/2 (p^2 - p)$ " grados de libertad.

Como complemento de las pruebas de significación referidas se propone el criterio de "significatividad sustantiva" (Kim y Mueller, 1978b). Éste se aplica con posterioridad a la significatividad *estadística*. Una vez que se ha comprobado que los factores extraídos son estadísticamente significativos, puede procederse a aumentar o a disminuir el número de factores para, de esta forma, asegurarse que tiene el número mínimo de factores que es compatible con sus datos. El incremento del número de factores suele producirse si los datos se desvían significativamente del modelo asumido. La disminución sucede, en cambio, cuando el modelo inicial se acepta como adecuado.

EJEMPLO DE APLICACIÓN DEL CRITERIO DE SIGNIFICATIVIDAD ESTADÍSTICA

La prueba de bondad de ajuste χ^2 se realiza cuando el procedimiento de extracción de factores es "máxima verosimilitud" y "mínimos cuadrados generalizados". Además, su uso exige el cumplimiento del supuesto de *normalidad* multivariable, siendo sensible incluso a leves incumplimientos de dicho supuesto. Se aplica una vez obtenida la *matriz factorial* (apartado 5.6) y previo a la *rotación* de los ejes factoriales. Mediante ella se comprueba la adecuación del modelo factorial en la identificación de las dimensiones *latentes* de las variables: si con el número de factores comunes extraídos logran explicarse las correlaciones de la serie de variables observadas.

En ambos procedimientos factoriales se llega a una agrupación similar de las 14 variables en 4 factores comunes, como se verá en el apartado 5.6. Igualmente, se logra un porcentaje acumulado de varianza explicada similar: 40,376% en *máxima verosimilitud* y 40,651% en *mínimos cuadrados generalizados*. Ambos porcentajes son inferiores a los obtenidos en ACP: 52,951%.

Respecto a la prueba χ^2 , ésta resulta bastante significativa en ambos procedimientos factoriales ($p < ,0005$). La significatividad del estadístico χ^2 lleva al rechazo de H_0 , que afirma que "se ha logrado extraer toda la varianza poblacional mediante los 4 factores extraídos de la muestra analizada". El rechazo de dicha hipótesis lleva, en consecuencia, a considerar la conveniencia de incrementar el número de factores para, de esta forma, poder reproducir adecuadamente las correlaciones entre las variables observadas. A decir por la prueba, las predicciones del modelo con 4 factores comunes "no" son muy buenas estimaciones de dichas correlaciones.

La aplicación de la prueba χ^2 confirma su principal crítica: el llevar a modelos con más factores comunes de los deseables con la finalidad de mejorar el ajuste del modelo a los

Método de extracción	Chi-cuadrado	g.l.	Sig.
Máxima verosimilitud	91,942	41	,000
Mínimos cuadrados generalizados	89,328	41	,000

* Los grados de libertad son iguales a $1/2 ((p - k)^2 - p - k) = 1/2 ((14 - 4)^2 - 14 - 4) = 41$.

datos. Especialmente, en muestras grandes como la presente (2.492 casos), debido a que el valor χ^2 es directamente proporcional al tamaño muestral. También afecta el número de variables que el modelo incluya. En suma, puede llevar a ir incrementando el número de factores a extraer hasta alcanzar un ajuste razonablemente bueno ($p > ,05$), que lleve a la aceptación de H_0 . Pero, como ya se ha dicho antes, en la decisión de incrementar el número de factores también intervienen criterios anteriormente descritos, además de la llamada significatividad "sustantiva". La solución factorial ante todo ha de tener significado lógico que facilite su interpretación. Esto lleva, necesariamente, a la lectura de la *matriz factorial*, preferiblemente la "rotada". A la vista de la composición de cada factor puede decidirse su inclusión o exclusión del modelo final.

E) Interpretabilidad

A diferencia del criterio de *significatividad*, el de *interpretabilidad* se aplica en toda la variedad de procedimientos factoriales. Una vez que se ha decidido el número de componentes o factores que compondrán el modelo factorial, y antes de proseguir con los análisis, debe comprobarse si los factores extractados tienen *significado sustantivo*. De los factores se pide, ante todo, que sean interpretables, que tengan significado lógico desde alguna perspectiva teórica. En caso contrario, debería replantearse su incorporación al modelo factorial.

Este último criterio de selección de factores concierne a la capacidad de asignar un significado a los factores. Complementa a los criterios anteriores e, incluso, puede provocar el replanteamiento del modelo factorial en su conjunto. El modelo factorial ha de ser *interpretable* desde alguna perspectiva teórica, además de *parsimonioso*. Los factores carentes de significado sustantivo conforman una estructura latente sin un nexo de unión lógico entre las variables empíricas que lo integran (aunque su correlación o covarianza sea elevada).

La composición de cada factor, junto con la variabilidad explicada, se ofrece en la *matriz factorial*. A partir de sus componentes (preferiblemente de la *matriz rotada*) se procede a la lectura e interpretación del modelo factorial. Si se observa que un factor ha quedado pobremente definido al incluir sólo una variable con un peso o coeficiente factorial bajo, puede considerarse la exclusión del factor del modelo analítico y repetir los análisis sin dicho factor. Lo mismo puede acontecer cuando el factor carezca de interpretación lógica por la entidad de las variables que lo forman.

5.6. La matriz factorial y su interpretación

Elegido el procedimiento de extracción factorial y el número de factores que formarán el modelo analítico, se procede al cálculo de la *matriz factorial* (también llamada "pattern matrix"). Para ello se parte de la matriz de *correlación*, o la de *varianza-covarianza*, dependiendo de cuál haya sido la seleccionada.

La *matriz factorial* básica contiene los factores o componentes y las variables empíricas o indicadores que los integran. Los factores o componentes suelen estar ubicados en las columnas de la matriz. Las variables empíricas se sitúan en las filas. Además se incluyen los pesos o *coeficientes factoriales* λ_{ij} ("factor loadings" en AFC y "component loadings" en ACP). Estos coeficientes tienen una interpretación análoga a los coeficientes de *regresión estandarizados*. Recuérdese que en ACP los componentes principales se explican en función de las variables observadas, mientras en AFC cada variable empírica actúa como la dependiente en la ecuación lineal y los k factores comunes como las independientes o predictoras.

Cuando los factores o componentes no se hallan correlacionados entre sí (quiere esto decir que son *ortogonales*) y las variables empíricas están estandarizadas, como es usual, los pesos o *coeficientes factoriales* equivalen a correlaciones de cada variable X_i con cada factor o componente j . En esta situación de incorrelación de estructuras latentes, la *matriz factorial* coincide con la de *estructura*. Esta última incluye las correlaciones entre variables y factores.

El cuadrado de los *coeficientes factoriales* (λ_{ij}^2) se interpreta de forma análoga al cuadrado de los coeficientes de correlación: proporción de varianza explicada de la variable X_i por cada factor (AFC) o proporción de varianza explicada de los componentes a partir de las variables que lo integran (ACP). En el ejemplo ilustrado a continuación, el coeficiente factorial de la variable X_1 en el factor 1 es igual a 0,725. Elevado al cuadrado ($0,725^2 = ,53$), y multiplicado por 100, significa que el 53% de la varianza de la variable X_1 es explicada por el factor 1. Lo mismo se haría con todos los coeficientes factoriales de cada variable en los factores del modelo, lo que ayuda en la valoración de lo bien que el modelo factorial describe a las variables observadas.

Los *autovalores* ("eigenvalues") para cada factor se obtienen de la suma de los cuadrados de los coeficientes factoriales (λ_{ij}^2) de cada variable empírica en ese factor. Expresan la cantidad de la *varianza total* en los datos que representa cada factor. "Como la *varianza total* explicada por un componente o factor es igual a la suma de cuadrados de los coeficientes para la columna correspondiente, los coeficientes de componentes principales serán más altos, en general, que los coeficientes de factor común" (Dunteman, 1989: 62).

De la división de cada *autovalor* entre la *varianza total* y multiplicada por 100 se calcula el *porcentaje de varianza explicada* por cada factor. Cuando los factores o componentes no se hallan correlacionados, puede conocerse la *varianza total* explicada por el modelo factorial de la suma de los distintos *autovalores*. Tomando los datos del ejemplo, se observa que el modelo factorial compuesto por dos factores comunes logra explicar el 71,4% de la *varianza total* de las cinco variables empíricas. Como es usual, el primer factor explica el mayor porcentaje de *varianza*. Exactamente, el 42,4%.

La suma de los (λ_{ij}^2) para cada variable en todos los factores o componentes proporciona su *comunalidad* (h_i^2). Ésta expresa la proporción de la varianza de cada variable X_i que es explicada por los factores o componentes en el modelo. En la *matriz factorial* del ejemplo se obtiene que el 84,1% de la varianza de la variable X_1 queda explicada por los dos factores juntos. Si a 1 (la varianza total de la variable estandarizada) se le resta la *comunalidad*, se obtiene la proporción de *varianza específica* ($u_i = 1 - h_i^2$). Es decir, la proporción de varianza de cada variable que la solución factorial no logra explicar. El modelo de dos factores comunes del ejemplo logra explicar una proporción considerable de la varianza de las variables, salvo en las variables X_3 y X_4 .

En general, valores de *comunalidad* bajos (próximos a 0,0) expresan que la variable no ha quedado bien definida por el modelo factorial. Esto puede deberse a la exclusión en el modelo de un factor (o componente) de gran relevancia en la explicación del fenómeno que se analiza; o bien, a que la variable tiene una baja proporción de *varianza común* con otras variables. Lo que revierte en una mayor cuantía de *varianza única*.

EJEMPLO DE MATRIZ FACTORIAL "NO ROTADA"

Variables	Factor 1	Factor 2	Comunalidad (h_i^2)	Especificidad (u_i^2)
X_1	0,725(a)*	0,562	0,841 (b)	0,159 (c)
X_2	0,648	-0,739*	0,966	0,034
X_3	0,545*	-0,432	0,484	0,516
X_4	0,487	-0,526*	0,514	0,486
X_5	0,798*	0,354	0,762	0,238
Autovalores (suma de cuadrados)	2,12 (d)	1,45	3,57	1,43
% varianza total	42,4 (e)	29,0	71,4 (f)	28,6
% varianza común	59,4 (g)	40,6		

* Factor en el que más satura las variables.

(a) λ_{11} : *coeficiente factorial* de la variable X_1 en el factor 1. Es la correlación de dicha variable con el factor 1: $\lambda_{11}^2 = 0,725^2 = 0,526$. La proporción de la varianza de la variable X_1 explicada por el factor 1.

(b) h_1^2 : *comunalidad* de la variable $X_1 = \sum_{j=1}^k \lambda_{1j}^2 = 0,725^2 + 0,562^2 = 0,841$. Porción de la varianza de la variable X_1 explicada por los factores comunes.

(c) u_i^2 : especificidad de la variable $X_i = 1 - 0,841 = 0,159$. Proporción de la varianza de la variable X_i no explicada por los factores comunes.

(d) Autovalor para el factor 1 =

$$\sum_{i=1}^p \lambda_{q1}^2 = 0,725^2 + 0,648^2 + 0,545^2 + 0,487^2 + 0,798^2 = 2,12$$

(e) % de varianza entre todas las variables que es explicada por el factor 1 = Autovalor / $p \times 100 = (2,12/5) \times 100 = 42,4$.

(f) % de varianza entre todas las variables que es explicada por los dos factores = $(3,57/5) \times 100 = 71,4$.

(g) Varianza entre todas las variables que es explicada por el factor 1 como un porcentaje de la que es explicada por todos los factores =

$$\left(\frac{\sum_{i=1}^p \lambda_{q1}^2}{\sum_{i=1}^p h_i^2} \right) \times 100 = (2,12/3,57) \times 100 = 59,4$$

La *matriz factorial* del ejemplo incluye los integrantes básicos: las variables empíricas, los factores y los coeficientes factoriales. A éstos se añaden las comunales, las especificidades, los autovalores y los porcentajes de varianza explicada por la solución factorial. En la mayoría de los paquetes estadísticos, como el programa SPSS, estos últimos valores figuran en tablas aparte de la matriz factorial: los *autovalores* y los porcentajes de varianza explicados por la solución factorial en la tabla "varianza total explicada", las *comunales* en una tabla que reúne las *comunales* "iniciales" y las posteriores a la "extracción" factorial. Ello facilita la comparación entre ambas *comunales*. De interés, sobre todo en AFC, donde las *comunales iniciales* expresan *varianza común*. Indican la correlación múltiple cuadrada de cada variable observada X_i (que actúa a modo de variable dependiente) respecto a las " $p - 1$ " variables restantes (las independientes o predictoras); es decir, la proporción de la varianza total en X_i que es explicada por las demás variables.

Las *comunales finales* suelen diferir, en cuantía, de las *iniciales*. Tanto en ACP como AFC, sus valores corresponden a las correlaciones múltiples cuadradas de cada variable X_i (la dependiente), pero ahora respecto a los factores actuando como las variables independientes en una ecuación de regresión múltiple. Representan la proporción de la varianza de cada variable que logra predecirse por la estructura *latente*. Estas *comunales* suelen estimarse mediante procedimientos iterativos, de ajuste continuo entre las *matrices de correlación observada y la reproducida*.

La *matriz de correlación reproducida* contiene, en su diagonal principal, las estimaciones de comunalidad a partir de la suma de los cuadrados de los coeficientes factoriales en los factores retenidos para cada fila de la matriz factorial. A partir de esta matriz se procede a un nuevo análisis factorial, que genera nuevas estimaciones de comunalidad. El procedimiento concluye cuando convergen las *matrices de coeficientes factoriales*.

Por último, hay que advertir que, cuando las *comunales* son pequeñas (próximas a 0,0) y/o varían considerablemente, los *coeficientes factoriales* de componentes

principales serán considerablemente más grandes que los correspondientes a AFC (Dunteman, 1989).

EJEMPLO DE MATRIZ DE CORRELACIÓN REPRODUCIDA

Para ilustrar lo dicho sobre la matriz de correlaciones reproducidas (tabla A), se selecciona la correspondiente al ACP. Además, se añade la tabla de comunalidades (tabla B) para que pueda constatar que la diagonal de la susodicha matriz está formada por las estimaciones de la comunalidad después de haberse producido la "extracción" factorial.

Tabla A. Matriz de correlaciones reproducidas

	X ₁₅	X ₁	X ₂	X ₃	X ₄	X ₅	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄
X ₁₅	,531 ^b													
X ₁	,302	,465 ^b												
X ₂	-,249	-,133	-,766 ^b											
X ₃		-,287	,448 ^b	-,083	-,132	-,279	,352	-,370	-,296	-,588	,130	,052	-,570	-,306
X ₄		,089	-,282	-,132	-,426	,426	-,465	-,248	-,221	,216	,166	-,189	-,419	
X ₅		,214	,058	-,194	,177	-,087	-,055	,186	,114	-,028	,093			
X ₆		-,092	,231	-,092	,102	,136	,181	-,554	-,488	,153	,168			
X ₇		-,385	,417 ^b	-,385	,418	,232	,216	-,320	-,266	,183	,395			
X ₈		-,421 ^b	-,449	-,421 ^b	-,449	-,224	-,281	,176	,136	-,258	-,389			
X ₉			,261		,484 ^d	,261	,301	-,179	-,128	,272	,422			
X ₁₀			-,099			-,099	,274 ^b	-,013	,319	,239				
X ₁₁			-,115					-,004	,708	,249				
X ₁₂			,673 ^b					-,010	-,189					
X ₁₃			,619 ^b					-,089	-,242					
X ₁₄								,010	-,189					
								,691 ^b	,219					
									,385 ^b					
Residuo ^a														
X ₁₅		000	-,054	,032	,019	-,017	-,070	,076	,081	,113	,010	,005	,174	,038
X ₁			,068	,099	,003	,089	-,140	,097	,106	-,025	-,032	-,010	-,038	,129
X ₂				,256	,029	-,074	,012	-,060	-,148	,028	-,064	-,090	,057	-,061
X ₃					,225	-,086	,155	-,131	,108	,030	-,105	-,006	,054	-,038
X ₄						-,090	,003	,015	-,004	-,019	,113	,198	-,028	,033
X ₅							,102	-,143	-,049	,017	,066	,044	,011	-,078
X ₆								,060	,025	,015	-,001	-,018	,009	,121
X ₇									-,003	-,010	,025	-,011	-,006	-,113
X ₈										-,113	-,028	-,056	-,058	-,060
X ₉											,000	-,019	-,136	,021
X ₁₀												-,159	-,006	,038
X ₁₁													-,051	,071
X ₁₂														-,011
X ₁₃														
X ₁₄														

Método de extracción: Análisis de componentes principales.

^a Los residuos se calculan entre las correlaciones observadas y las reproducidas. Hay 46 (50%) de residuos no redundantes con valores absolutos > ,05.

^b Comunalidades reproducidas.

• X₁₅ "simpatía hacia norteafricano (marroquí...); X₁ "leyes inmigración"; X₂ "ideología política"; X₃ "sexo"; X₄ "edad"; X₅ "número de inmigrantes"; X₆ "regularizar a inmigrantes"; X₇ "entrada inmigrantes"; X₈ "partido racista"; X₉ "casar con marroquí"; X₁₀ "estudios"; X₁₁ "ingresos"; X₁₂ "vecino marroquí"; X₁₃ "inmigrante delincuente".

Obsérvese que en la matriz de *correlaciones reproducidas* se añaden los *residuos*. En el análisis factorial éstos se definen como la diferencia entre el coeficiente de correlación *observado* y el *reproducido* o estimado desde el modelo obtenido. Por ejemplo, el *residuo* de la variable "leyes de inmigración" (X_{15}) respecto a la variable "simpatía hacia marroquí" (X_{15}) es igual a ,000; valor nulo porque ambas correlaciones coinciden. En la matriz de *correlaciones* (subapartado 1.3.2) puede comprobarse que la correlación *observada* entre ambas variables es ,302. Este valor coincide con la correlación *reproducida* (,302) –tabla A–. La menor coincidencia entre ambas correlaciones se produce precisamente en las dos variables que, a decir por sus *comunalidades*, menor varianza comparten con el resto de las variables: "sexo" (X_3) e "ideología política" (X_2). La correlación *observada* entre ambas variables es $-.031$, mientras que la *reproducida* es $-.287$. Su *residuo* es igual a ,256 ($-.031 - [-.287]$). Le sigue la correlación entre las variables "sexo" y "edad": correlación *observada* = $-.058$; correlación *reproducida* = $-.282$; *residuo* = ,225.

Tabla B. *Comunalidades*

	<i>Inicial</i>	<i>Extracción</i>
Simpatía marroquí	1,000	,531
Leyes inmigración	1,000	,465
Ideología política	1,000	,766
Sexo	1,000	,448
Edad	1,000	,507
N.º inmigrantes	1,000	,417
Regularizar inmigrante	1,000	,421
Entrada inmigrante	1,000	,484
Partido racista	1,000	,274
Casar con marroquí	1,000	,731
Estudios	1,000	,673
Ingresos	1,000	,619
Vecino marroquí	1,000	,691
Inmigrante delincuente	1,000	,385

Método de extracción: Análisis de componentes principales.

En la tabla A se dice que la mitad de los *residuos* son "no redundantes", al ser su valor absoluto superior a ,05. Como en la mayoría de los análisis estadísticos, en el análisis factorial interesa que los *residuos* sean bajos porque indica que el ajuste del modelo es bueno. En caso contrario, debería reconsiderarse el modelo. Para más información, véase el apartado 5.7.

5.6.1. *La rotación de factores*

La *matriz factorial* muestra la relación habida entre las variables y los factores o componentes, pero su interpretación no suele ser sencilla. Es frecuente que varias variables presenten coeficientes factoriales elevados en más de un factor, cuando lo que interesa es que la mayor parte de su variabilidad quede explicada por un solo factor y no por varios. Esto lleva al desarrollo del principio de "estructura simple", según el cual las

variables han de saturar básicamente en un factor. Quiere esto decir que sus coeficientes factoriales han de ser elevados en un solo factor o componente y bajos en el resto. Un *coeficiente factorial* próximo a 0,0 indica inexistencia de relación entre la variable empírica y el factor. Un valor elevado (positivo o negativo), cercano a 1,0, expresa lo contrario.

El principio de “estructura simple” fue propuesto por Thurstone en 1947 (*Multiple Factor Analysis*, Chicago University Press). En el análisis factorial *booleano* este principio no es tan exigido. Una variable puede tener un coeficiente factorial elevado (próximo a 1,0) en más de un factor, lo cual elimina la necesidad de efectuar una *rotación* factorial. Salvo esta excepción, en la generalidad de los análisis factoriales se demanda el principio de “estructura simple”. Éste incluye los supuestos siguientes:

- a) Cada factor o componente ha de tener unos pocos coeficientes factoriales elevados y los demás próximos a cero.
- b) Cada variable empírica ha de presentar un coeficiente factorial elevado sólo en un factor.
- c) Los factores o componentes no han de tener la misma distribución. Deben mostrar un modelo diferente de coeficientes factoriales elevados y bajos.

Tanto en ACP como en AFC lo habitual es que el primer factor sea un factor general, que agrupe la mayoría de las variables observadas. Casi todas ellas presentan coeficientes factoriales elevados en dicho factor, lo que determina que éste sea el factor que mayor proporción de varianza explique. Los factores o componentes siguientes explican cada vez menor variabilidad. Esto guarda relación con el decreciente número de variables con coeficientes elevados. Además, estos factores también suelen caracterizarse por ser bipolares. Es decir, por incluir coeficientes factoriales positivos y negativos.

Si el investigador busca “simplificar” la estructura factorial, tendrá que proceder a la *rotación* de la *matriz factorial*. La *rotación* consiste en girar los ejes factoriales para que éstos se aproximen a las variables empíricas. La finalidad es facilitar la interpretación de la matriz factorial, forzando a las variables a definirse más en una dimensión latente, con preferencia a otras. De esta manera se obtiene una mayor diferenciación entre los factores que logran un perfil más definido. El número de factores o componentes se mantiene inmutable, al igual que el porcentaje de varianza total explicada por el modelo factorial inicial y las *comunalidades* de las variables. Lo que varía es la composición de los factores, al cambiar los *coeficientes factoriales* de cada variable en cada factor. Esto también altera la proporción de variabilidad explicada por cada factor. No se olvide que con la rotación se redistribuye la varianza entre todos los factores.

EJEMPLO DE MATRIZ FACTORIAL “ROTADA”

Para ilustrar en qué consiste la “rotación” factorial, a continuación se incluye la representación gráfica y la matriz factorial “rotada” que corresponde al primer ejemplo expuesto a principio del apartado 5.6.

- a) El gráfico (figura A) incluye la representación de los coeficientes factoriales en una serie de coordenadas cuyos ejes lo forman los dos factores del modelo. Los *coeficientes factoriales* se representan en los ejes respectivos, tomando las puntuaciones obtenidas en cada factor. De la intersección de ambas puntuaciones se obtiene la posición de las variables empíricas respecto de los ejes factoriales.

Este gráfico difiere de los convencionales en que cada punto representa una variable y los ejes su factor común o componente principal. En los gráficos convencionales, los puntos representan casos individuales y cada eje una variable. Estos gráficos corresponden al análisis de variables. Cuando se está interesado en el análisis de casos, son los casos (y no las variables) los que se representan en los ejes factoriales. Las coordenadas de cada caso vienen dadas por sus correspondientes puntuaciones factoriales.

Con la *rotación* de los ejes factoriales (en el sentido de las agujas del reloj) se persigue la aproximación de los ejes a los indicadores. Estos nuevos ejes se seleccionan para que atraviesen los conglomerados de los puntos que representan a los distintos indicadores. El gráfico a continuación ilustra cómo se procede en la *rotación ortogonal*. Adviértase que el nuevo eje, que representa al factor rotado 1 atraviesa el conglomerado compuesto por las variables empíricas X_2 , X_3 y X_4 . El eje para el factor 2 se aproxima a las variables X_1 y X_5 . Ambos ejes son perpendiculares entre sí, como es característico de la *rotación ortogonal*. Este procedimiento de *rotación*, como después se verá, se asienta en el supuesto de inexistencia de correlación entre los factores. Como se trata de ejes no correlacionados, la distancia entre ellos ha de ser de 90° , tanto antes como después de la rotación. Tras la rotación lo que se obtiene es una mejor definición de los factores.

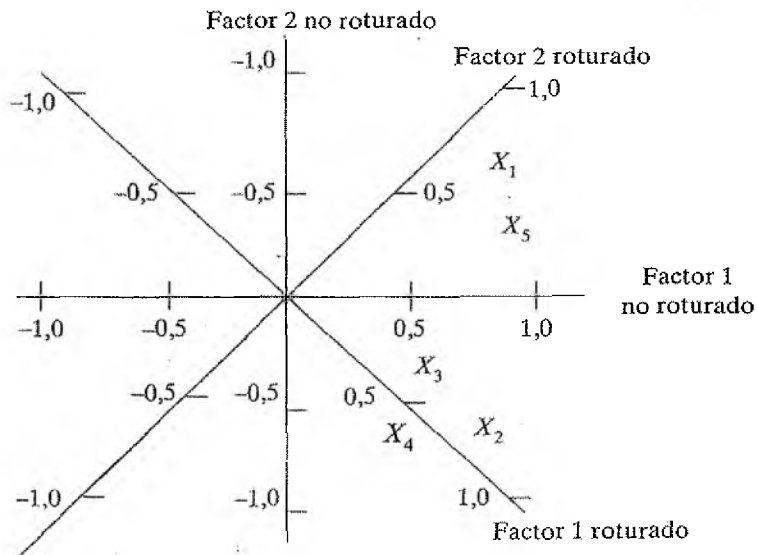


Figura A.

- b) Los paquetes estadísticos ofrecen también otros gráficos para la solución factorial rotada, cuyas coordenadas se corresponden con los coeficientes factoriales para la matriz factorial *rotada*. En dicho *gráfico de coeficientes factoriales* tras la rotación (en este caso *ortogonal varimax*), puede igualmente comprobarse el éxito de la rotación hecha.

En la figura B puede observarse la existencia de dos conglomerados de variables situados en los extremos de los ejes factoriales. El conglomerado que componen las variables X_2 , X_3 y X_4 se sitúa en el extremo positivo del factor 1, mientras que el conglomerado integrado por las variables X_1 y X_5 se ubica en el extremo positivo del factor 2. Todo lo cual expresa que con la rotación ortogonal se ha alcanzado una estructura "simple". Las variables han quedado localizadas al final de cada eje, al presentar coeficientes factoriales elevados en sólo un factor. Ninguna se sitúa en el origen (0,0), lo que indicaría, por el contrario, la existencia de bajos coeficientes factoriales en ambos factores. Cuando esto sucede, significa que las variables no están asociadas a ningún factor.

En general, una hilera de puntos a lo largo de los ejes factoriales indica que el factor no ha quedado claramente definido por las variables que en él saturan. Asimismo, si los conglomerados de variables se hubiesen situado a medio camino entre los dos ejes factoriales, significa que se precisa realizar una *rotación oblicua*, porque los ejes factoriales se encuentran correlacionados entre sí. Esta localización de los conglomerados de variables también puede deberse a la presencia de otro factor que habría que incorporar al modelo factorial. En ambas situaciones el investigador deberá introducir modificaciones y realizar de nuevo los análisis.

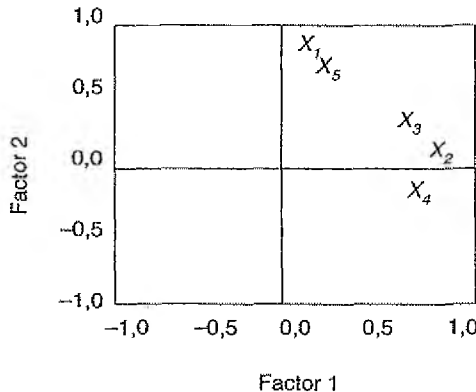


Figura B. Gráfico de coeficientes factoriales tras la rotación ortogonal varimax.

- c) La *matriz factorial rotada*. Si se compara con la obtenida antes de la rotación, pueden advertirse las diferencias en la solución factorial antes y después de la rotación. La estructura factorial incluida en la matriz no rotada está menos definida que la alcanzada tras la rotación. Si antes de la rotación, la mayoría de las variables empíricas presentan coeficientes factoriales elevados (todos superiores a 0,30) en el factor 1,

después de la rotación dos variables (X_1 y X_5) quedan más definidos en el factor 2. Sus coeficientes factoriales en el factor 1 dejan de ser significativos (inferiores a 0,30).

Asimismo, se observa una alteración notoria en la composición de los factores. Las variables que presentan un mayor peso o coeficiente factorial, previo a la rotación, en el factor 1 son X_1 , X_3 y X_5 . Después de la rotación ortogonal, el factor 1 queda claramente configurado por las variables X_2 , X_3 y X_4 . La ponderación de estas variables en el factor 2 es escasa. De manera especial, la variable X_2 , con un coeficiente en el factor 2 de 0,072. El factor 2 queda definido por las variables X_1 y X_5 , solución nada coincidente con la configuración previa a la rotación factorial. Con la rotación se ha logrado una mayor definición y diferenciación entre los factores, al forzarse a los indicadores a definirse más por un factor que por otro.

Tabla A. Matriz factorial rotada por el procedimiento ortogonal varimax

Variables	Factor 1	Factor 2	Comunalidad (h_i^2)	Especificidad (u_i^2)
X_1	0,167	0,902 *	0,841	0,159
X_2	0,980 *	0,072	0,966	0,034
X_3	0,659 *	0,224	0,484	0,516
X_4	0,709 *	-0,108	0,514	0,486
X_5	0,258	0,834 *	0,762	0,238
Autovalores (suma de cuadrados)	1,992	1,576	3,57	1,43
% varianza total	39,8	31,5	71,4	28,6
% varianza común	55,8	44,2		

* Factor en el que más satura las variables.

De la comparación entre ambas matrices factoriales también puede constatarse la variación en *varianza* explicada por cada factor. Si antes de la rotación el factor 1 presentaba un *autovalor* de 2,12, que suponía el 42,4% del total de *varianza explicada* (71,4%), después de la rotación su *autovalor* desciende a 1,992, que traducido a *varianza explicada* se queda en el 39,8%. Este descenso en *varianza* del factor 1 es consecuencia de una mayor definición del factor 2 tras la rotación. Este factor queda claramente configurado por sólo dos variables (X_1 y X_5), aunque con una saturación muy elevada: X_5 con un coeficiente factorial de 0,834 y X_1 de 0,902. Este factor pasa de explicar el 29% de la *varianza total* al 31,5%.

En suma, la *rotación* ha ocasionado una redistribución de la *varianza total* entre los factores, sin alterar la proporción total de *varianza explicada* por el modelo de dos factores. Tampoco se han modificado las *comunalidades*; es decir, la proporción de la *variabilidad* de cada variable empírica que explica el modelo factorial de dos factores comunes.

El ejemplo anterior corresponde a un procedimiento específico de rotación: la rotación *ortogonal* y, dentro de ella, a la variedad más popular llamada “varimax”. Pero, aunque ésta sea la opción más popular, no quiere decir que sea la única posible. En el cuadro 5.1 se resumen los principales procedimientos de rotación. Se diferencia entre rotación “ortogonal” y la “oblicua”.

CUADRO 5.1. *Procedimientos de rotación factorial*

<p>A. ROTACIÓN ORTOGONAL:</p> <ol style="list-style-type: none"> 1. Varimax 2. Quartimax 3. Equamax <p>B. ROTACIÓN OBLICUA:</p> <ol style="list-style-type: none"> 1. Oblimin 2. Quartimin 3. Biquartimin o covarimin 4. Oblimax 5. Promax
--

A) Rotación ortogonal

Parte del supuesto de incorrelación entre las dimensiones latentes. Todo factor común o componente principal es independiente de los demás, lo que determina que los ejes factoriales se mantengan perpendiculares entre sí, formando un ángulo de 90°. Esta variedad genérica incluye los tres procedimientos principales siguientes:

A.1. Varimax

Es el procedimiento de rotación factorial ortogonal más empleado y el que se aplica, por defecto, en la mayoría de los programas estadísticos. Fue propuesto por Kaiser en 1958 (en “The varimax criterion for analytic rotation in factor analysis”, *Psychometrika*, 23: 187-200). Su finalidad es simplificar la estructura factorial maximizando la varianza de los coeficientes factoriales cuadrados (λ_{ij}^2) para cada factor. De ahí le viene el nombre “varimax”, acorde con su objetivo principal: que la VARIANZA se MAXIMICE para ayudar a la interpretación de los factores.

La simplificación máxima se alcanza cuando en cada *columna* de la matriz factorial (donde se sitúan los factores) existen sólo unos y ceros (coeficientes factoriales muy elevados en unas variables y muy bajos en el resto).

En la aplicación de la rotación *varimax* pueden emplearse coeficientes factoriales “brutos” o “normalizados”. Lo habitual es utilizar coeficientes factoriales “normalizados” (o

estandarizados) para equilibrar la influencia de variables con distintas comunalidades en la solución factorial. Para ello se aplica el criterio de *normalización* de Kaiser que consiste en ajustar los coeficientes factoriales dividiendo, cada uno, por la raíz cuadrada de la *comunalidad* de la variable correspondiente. O, igualmente, dividiendo cada coeficiente factorial cuadrado por la comunalidad de la variable respectiva.

La aplicación de los coeficientes factoriales *normalizados* proporciona a cada variable empírica igual peso en la rotación. De esta forma se evita que aquellas variables con *comunalidades* más elevadas influyan más en la solución factorial final. Cuando se aplique este criterio, se está ante una *rotación varimax normalizada*. Ésta se muestra muy adecuada cuando las variables presentan *comunalidades* muy dispares (elevadas en unas variables y muy bajas en otras).

A.2. Quartimax

En vez de simplificar las columnas de la matriz factorial (como sucede en la rotación *varimax*), busca la simplificación de las *filas* (variables) de la matriz. Esto lleva a maximizar la varianza de los coeficientes factoriales cuadrados para cada variable. Cada variable deberá tener una correlación elevada en unos pocos factores y baja en el resto, de manera que sea mínimo el número de factores necesarios para explicar su variabilidad.

Analíticamente *quartimax* es más sencillo que *varimax*. Si bien, *varimax* ofrece una separación más clara de los factores (Kim y Mueller, 1978b; Dillon y Goldstein, 1984; Hair *et al.*, 1992, 1999). En *quartimax* muchas variables pueden presentar coeficientes factoriales elevados en un mismo factor, al buscar la simplificación de las filas. En cada fila debe haber sólo un coeficiente factorial elevado, el resto ha de situarse próximo a cero. Esto puede provocar el efecto no deseado de que muchas variables presenten coeficientes factoriales elevados en un mismo factor. Habría un único factor general integrado por la mayoría de las variables que ponderan mayoritariamente en este factor.

A.3. Equimax

Este tercer procedimiento de rotación ortogonal se presenta como una síntesis de los dos anteriores. Busca tanto la simplificación de las *filas* (variables) como de las *columnas* (factores) de la matriz factorial. Pese a ello no ha tenido el desarrollo esperado, siendo de los tres procedimientos de rotación el menos aplicado en la práctica.

B) Rotación oblicua

Muestra mayor adecuación a situaciones muy habituales de interrelación entre dimensiones latentes. Al permitirse la correlación entre los factores o componentes, los

ejes factoriales no son perpendiculares, pudiendo estar separados por un ángulo inferior a 90° . El coseno del ángulo muestra la correlación entre los dos ejes factoriales que se relacionan. La consideración de dimensiones latentes interrelacionadas comporta una mayor complejidad en la interpretación de la solución factorial, aunque ésta sea teórica y empíricamente más realista.

La correlación mutua entre los factores introduce modificaciones importantes en la matriz factorial. En primer lugar, los coeficientes factoriales no pueden ya interpretarse como coeficientes de correlación simple, como sucede en la rotación *ortogonal*, sino sólo como coeficientes de regresión de cada variable empírica X_i en función de factores mutuamente relacionados. De la *matriz factorial no rotada* derivan dos matrices diferentes: la matriz de los coeficientes factoriales (o la “matriz de configuración”: “pattern matrix”) y la matriz de correlaciones entre los factores y las variables (la “matriz de estructura”: “structure matrix”). En la rotación *ortogonal* ambas matrices coinciden, en la *oblicua* difieren. Además, las correlaciones entre las variables empíricas (indicadores) y las dimensiones latentes pueden estar infladas por el solapamiento entre los factores o componentes. Las variables pueden correlacionar con unos factores directamente y con otros indirectamente, a través de su correlación con otro factor a él correlacionado.

Para conocer el grado de correlación entre los factores o componentes, la generalidad de los programas estadísticos ofrecen una matriz de correlaciones entre los factores (“factor correlation matrix”). Ésta permite comprobar el grado de intercorrelación entre los factores. Información de interés en la decisión de qué modalidad de rotación aplicar con la finalidad de mejorar la solución factorial.

La interrelación entre los factores también altera la interpretación de las *comunalidades* de las variables y de la proporción de *varianza* explicada por cada factor. Las *comunalidades* no coinciden con la suma de los coeficientes factoriales al cuadrado de cada variable en cada factor o componente. Tampoco la proporción de *varianza* explicada por el factor se obtiene a partir de la suma de los coeficientes cuadrados de las variables que saturan en dicha dimensión latente. Los factores están correlacionados, lo que entorpece la interpretación de la suma de los coeficientes factoriales cuadrados rotados oblicuamente en términos de proporción de *varianza*.

La rotación *oblicua* incluye igualmente una amplia variedad de alternativas. Éstas pueden sintetizarse en las siguientes:

B.1. Oblimin

Es uno de los procedimientos de rotación *oblicua* más populares, al estar incluido en la mayoría de los programas estadísticos, como el SPSS. Suele aplicarse con coeficientes factoriales *normalizados*, obtenidos de la división del coeficiente factorial al cuadrado por la raíz cuadrada de la *comunalidad* de la variable respectiva.

Este procedimiento de rotación resulta de la combinación de dos criterios:

- *Quartimin*: que se asemeja al *quartimax* en la rotación ortogonal. Minimiza la suma de los productos de los coeficientes factoriales y produce factores de mayor oblicuidad.
- *Covarimin*: minimiza la covarianza de los coeficientes factoriales cuadrados. Tiende a generar factores menos oblicuos y, como los dos criterios anteriores, también suele utilizar coeficientes factoriales *normalizados*.

B.2. Oblimax

Se presenta como un criterio alternativo al *quartimin* (el equivalente oblicuo del *quartimax*). Persigue la simplificación de la estructura factorial, aumentando (*maximiz*ando) el número de coeficientes factoriales elevados y bajos. En cambio, disminuye los de rango medio.

B.3. Promax

Es uno de los procedimientos de rotación *oblicua* más novedosos, incorporándose a las últimas versiones de programas como el SPSS. Se caracteriza por conseguir la solución oblicua aplicando algunas funciones de la solución *ortogonal*, como la matriz "target": aquella que el investigador tiene en mente o cree que existe.

"La razón detrás de la rotación promax es que las soluciones ortogonales suelen aproximarse a la solución oblicua, y reduciendo los coeficientes más pequeños a coeficientes de casi cero, se puede obtener una matriz *target* de estructura simple razonablemente buena. Entonces, encontrando los factores oblicuos que mejor ajusten (aplicando el criterio de mínimos cuadrados) a esta matriz *target*, se obtiene la solución oblicua deseada" (Kim y Mueller, 1978b: 40-41).

Este tercer criterio de rotación *oblicua* suele utilizarse cuando se quiere comprobar la congruencia de una estructura factorial con otra que se conoce.

EJEMPLO DE DISTINTOS PROCEDIMIENTOS DE ROTACIÓN ORTOGONAL Y OBLICUA EN ACP

Para que pueda mejor comprenderse la lógica de la *rotación*, se decide aplicar las distintas opciones de *rotación* a un mismo método de extracción factorial: el análisis de *componentes principales*. Los métodos de *factor común* se ejemplifican en el apartado 5.8.

Primero, se expone la matriz de *componentes "no rotada"* (tabla A). Punto de referencia para comprobar la mejora lograda con la *rotación*. En la salida original del programa SPSS (versión 10.0) dicha matriz incluye únicamente las variables, los componentes extractados y las saturaciones o coeficientes factoriales ("component loadings"). La información sobre *co-*

munalidad y los *autovalores* se da en tablas aparte. Los *autovalores* en la tabla común llamada "varianza total explicada" (subapartado 5.5.2). Las *comunalidades* en una tabla específica ya expuesta en el segundo ejemplo del apartado 5.6.

En la matriz *no rotada* (tabla A) se observa una estructura poco definida de los componentes, al haber 8 variables (de las 14 analizadas) con saturaciones "significativas" en más de un componente. Se toma el referente más habitual de $\lambda_{ij} \geq ,30$. Supone que al menos un 9% ($,30^2 = ,09$) de la varianza de la variable ha quedado explicada por el componente. Dicho referente puede incluso reducirse ligeramente a las cantidades expuestas en el subapartado 5.6.2, dado el elevado tamaño muestral.

Por ejemplo, la variable "casar con marroquí" presenta una correlación significativa con tres componentes: 1 ($,639$), 3 ($,442$) y 2 ($,330$). Su correlación con el componente 4 es, en cambio, negativa y más leve ($-,133$). En este último componente la variable que más pondera es "ideología política" ($,815$), seguida a distancia por la variable "sexo" ($-,394$). La presencia de las demás variables en este último componente no es significativa. Por el contrario, el componente 1 se presenta, como es usual, como un componente "general". En él saturan significativamente ($\lambda \geq ,30$) la generalidad de las variables, con la única excepción de la variable "sexo". Lo que concuerda con la información de los *autovalores* dada en el subapartado 5.5.2.

Tabla A. Matriz de componentes^a

	Componente			
	1	2	3	4
casar con marroquí	,639	,330	,442	-,133
simpatía marroquí	-,626	-,273	-,163	,193
entrada inmigrante	,618	,113	-,298	2,78E-03
vecino marroquí	,589	,341	,439	-,189
leyes inmigración	-,586	-2,02E-02	,336	8,51E-02
n.º inmigrantes	,582	-,123	-,243	6,367E-02
inmigrante delincuente	,571	1,20E-02	-,238	4,811 E-02
regularizar inmigrante	-,570	9,74E-02	,285	7,394E-02
partido racista	,449	,138	,106	,206
ingresos	-,329	,690	7,246E-03	,184
estudios	-,452	,674	8,15E-02	9,128E-02
edad	,388	-,536	,261	3,929E-02
sexo	9,040E-03	,279	-,464	-,394
ideología política	,301	,110	-1,85E-03	,815

Método de extracción: Análisis de componentes principales.

^a 4 componentes extraídos.

Previo a la *rotación* factorial, el primer componente presenta un *autovalor* (suma de las saturaciones al cuadrado tras la extracción factorial: $,639^2 + (-,626)^2 + ,618^2 + \dots + ,301^2$) igual a 3,619, que supone un 25,85% de varianza explicada ($3,619/14 \times 100$). El segundo componente explica el 11,89% de la varianza, al ser la suma de sus saturaciones igual a 1,665.

El tercer componente presenta un *autovalor* igual a 1,114, que se traduce a un 7,96% de la varianza total explicada. Por último, el componente 4 es el que menor cantidad de varianza explica: 1,015, es decir, el 7,25%. Recuérdese que la agrupación de las 14 variables en 4 componentes principales logra explicar el 52,95% de la varianza total (véase la tabla "varianza total explicada" en el subapartado 5.5.2).

Adviértase que en ACP, a diferencia de AFC, las medidas de varianza "iniciales" y "posteriores" a la extracción factorial no varían. Pero sí tras producirse la "rotación", como después se verá.

Respecto a las *comunalidades*, ya se vio, en el apartado 5.6, que al ser un ACP, las *comunalidades* "iniciales" son siempre igual a 1,0. Las de interés son las posteriores a la "extracción" factorial. Éstas se obtienen de la suma de las saturaciones al cuadrado de la variable en el conjunto de los componentes extractados. Su valor expresa la proporción de varianza de las variables respectivas que logra explicarse mediante la conjunción de los componentes. Las tres variables cuya variabilidad logra ser mejor explicada son "ideología política" ($h_i^2 = ,766 = ,301^2 + ,110^2 + [-,001847]^2 + ,815^2$), "casar con marroquí" ($h_i^2 = ,731 = ,639^2 + ,330^2 + ,442^2 + [-,133]^2$) y "vecino marroquí" ($h_i^2 = ,691 = ,589^2 + ,341^2 + ,439^2 + [-,189]^2$). En cambio, sólo el 27,4% de la varianza de la variable "partido racista" ha quedado explicada por la solución de cuatro componentes principales y el 38,5% de la variable "inmigrante delincuente". En estas dos últimas variables la mayor parte de su varianza queda sin explicar (*especificidad*): el 72,6% y el 61,5%, respectivamente.

En busca de una mejor definición de los componentes se prueban distintos procedimientos de *rotación*. Primero, se aplica la variedad más usual, *varimax* (tabla B), que persigue maximizar las saturaciones de una variable en un componente y minimizarla en el resto. Para evitar la desigual influencia de variables con diferentes *comunalidades*, se aplica además el criterio de *normalización de Kaiser*. Tras seis iteraciones (aunque se fijó el número máximo de iteraciones utilizadas por defecto en el programa SPSS: 25. Éstos son los pasos que puede seguir el algoritmo para realizar la *rotación*) se obtiene la matriz incluida en la tabla B.

Si previo a la *rotación* ocho variables presentan coeficientes "significativos" en más de un componente, tras la *rotación ortogonal varimax* el número de variables que ponderan en más de un componente se reduce a dos: las variables "sexo" (con $\lambda_{ij} \geq \pm ,30$ en tres componentes) y "simpatía marroquí" (en dos componentes). Las doce variables restantes quedan claramente definidas en un componente. La definición perfecta se produce cuando la variable presenta un coeficiente muy elevado (próximo a 1,00) en un componente y muy bajo (próximo a ,00) en el resto.

El primer componente principal queda más definido por la combinación lineal de cinco variables relacionadas con la "política inmigratoria": "leyes inmigración", "entrada inmigrantes", "regularizar inmigrantes", "número de inmigrantes" e "inmigrante delincuente".

El segundo componente por cuatro indicadores usuales de "simpatía" hacia los inmigrantes" (en concreto, hacia los "marroquíes"): "casar con marroquí", "vecino marroquí", "simpatía marroquí" y "partido racista".

El tercer componente por variables sociodemográficas relacionadas con la "posición social": "estudios", "ingresos" y "edad". En especial, las dos primeras.

El cuarto componente por dos variables que muestran estar relacionadas: "ideología política" y "sexo".

Si se compara la presente agrupación de variables con la obtenida mediante el análisis de conglomerados (capítulo 3), puede constatar que ambas coinciden, aunque la variable

Tabla B. Matriz de componentes rotados^a

	Componente			
	1	2	3	4
leyes inmigración	-,660	-,110	9,350E-02	-,915E-02
entrada inmigrante	,659	,217	-,425E-02	2,200E-02
regularizar inmigrante	-,610	-,210	5,204E-02	5,012E-02
n.º inmigrantes	,579	,111	-,246	9,453E-02
inmigrante delincuente	,576	,160	-,146	7,861E-02
casar con marroquí	,177	,830	-,554E-02	8,724E-02
vecino marroquí	,141	,818	-,341E-02	2,472E-02
simpatía marroquí	-,351	-,636	4,038E-02	4,251E-02
partido racista	,275	,332	-,332E-02	,295
estudios	-,207	-,400E-02	,793	-,931E-03
ingresos	-,173	6,978E-02	,754	,122
edad	5,780E-02	,143	-,675	,168
ideología política	,276	4,828E-03	9,794E-02	,825
sexo	,325	-,625E-02	,310	-,493

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización varimax con Kaiser.

^a La rotación ha convergido en 6 iteraciones.

“ideología política” finalmente no se incluyese en dicho análisis por las razones expuestas en dicho capítulo.

Tras la *rotación*, ni las *comunalidades* ni el porcentaje global de la varianza total explicada (52,95%) varían. Lo que sí cambia es la varianza explicada por cada componente, al variar las saturaciones de las variables en cada componente. La *rotación* produce una redistribución de la varianza que ocasiona una pérdida de la variabilidad explicada por el primer componente, con el aumento consiguiente en los demás componentes. El componente 1 pasa de un *autovalor* de 3,619 a 2,416 (un 17,26% de la varianza total); el componente 2, de 1,665 a 2,046 (14,61% de la varianza total); el componente 3, de 1,114 a 1,861 (el 13,29%); y el componente 4, de 1,015 a 1,090 (7,79%). Véase la tabla de “varianza total explicada” en subapartado 5.5.2.

La tabla C incluye la matriz de *rotación* para transformar los coeficientes de la matriz de componentes “no rotados” a “rotados”: la matriz de transformación de los componentes.

Tabla C. Matriz de transformación de las componentes

Componente	1	2	3	4
1	,720	,576	-,351	,163
2	,084	,440	,894	-,003
3	-,686	,636	-,247	,253
4	,059	-,265	,128	,954

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización varimax con Kaiser.

Después se prueban otros procedimientos de *rotación ortogonal*: *quartimax* (tabla D) y *equamax* (tabla E). *Quartimax* busca la simplificación de las filas de la matriz, en vez de las columnas. Si se compara la tabla D con la tabla B, puede observarse que salvo la variable "partido racista" (que queda más definida en el componente 1 que en el 2), el resto de las variables saturan en los mismos componentes en ambos procedimientos de rotación ortogonal. El número de variables con coeficientes significativos en más de un componente aumenta a tres: "sexo", "simpatía marroquí" y "partido racista".

Tabla D. Matriz de componentes rotados^a

	Componente			
	1	2	3	4
entrada inmigrante	,681	,141	-9,97E-03	-1,27E-02
leyes inmigración	-,675	-3,59E-02	6,164E-02	-5,73E-02
regularizar inmigrante	-,628	-,140	2,152E-02	8,236E-02
n.º inmigrantes	,603	4,496E-02	-,218	6,317E-02
inmigrante delincuente	,800	9,423E-02	-,118	4,768E-02
partido racista	,326	,300	-2,02E-02	,278
casar con marroquí	,275	,805	-4,41E-02	6,956E-02
vecino marroquí	,234	,798	-2,43E-02	9,159E-03
simpatía marroquí	-,419	-,592	2,078E-02	6,641E-02
estudios	-,248	-1,52E-02	,782	7,208E-03
ingresos	-,194	9,054E-02	,745	,135
edad	,114	,136	-,671	,159
ideología política	,310	-2,21E-02	,170	,811
sexo	,276	-,100	,328	-,505

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización *quartimax* con Kaiser.

^a La rotación ha convergido en 6 iteraciones.

Las *comunalidades* no varían respecto a la solución de rotación *varimax*, pero sí la proporción de varianza explicada por cada componente, debido a la pequeña alteración producida en su composición. Aunque continúa explicándose el 52,95% de la varianza total, aumenta la aportación del componente 1 en detrimento de los otros componentes, que ven disminuida su aportación a la varianza total explicada.

La aplicación del procedimiento de rotación *equamax* (tabla E) provoca, en cambio, una redistribución de la varianza total explicada por los componentes más equitativa. El primer componente explica el 15,75% de la varianza (*autovalor* igual a 2,205); el componente 2, el 15,31% (2,143); el componente 3, el 13,59% (1,903); y el componente 4, el 8,304% (1,163). Si se comparan la tabla E con la D y la B, puede observarse que la combinación lineal de variables que forman cada componente en la solución *equamax* se asemeja, en general, más a la obtenida mediante la rotación *varimax*. Aunque aumenta a cuatro las variables con coeficientes significativos en más de un componente: "sexo", "ideología política", "simpatía marroquí" y "partido racista".

Tabla E. Matriz de componentes rotados^a

	Componente			
	1	2	3	4
entrada inmigrante	,640	,251	-6,68E-02	7,918E-02
leyes inmigración	-,640	-,143	,117	-,146
regularizar inmigrante	-,597	-,244	7,497E-02	-3,31E-03
n.º inmigrantes	,554	,138	-,267	,143
inmigrante delincuente	,553	,188	-,167	,128
casar con marroquí	,120	,835	-6,16E-02	,122
vecino marroquí	9,077E-02	,824	-3,94E-02	5,601E-02
simpatía marroquí	-,315	-,655	5,382E-02	-1,03E-03
partido racista	,230	,338	-4,15E-02	,324
estudios	-,174	-4,89E-02	,800	-2,93E-02
ingresos	-,157	5,869E-02	,761	,107
edad	1,134E-02	,140	-,675	,178
ideología política	,214	-3,95E-03	9,300E-02	,844
sexo	,376	-2,98E-02	,295	-,468

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización equamax con Kaiser.

^a La rotación ha convergido en 6 iteraciones.

Por último, se comprueba si distintos procedimientos de *rotación oblicua* logran una mayor definición de las variables en los componentes. Se parte de la hipótesis de que los componentes pueden estar intercorrelacionados. Primero, se prueba el procedimiento de rotación oblicua más popular, *oblimin*, y después el procedimiento *promax*. Éstas son las dos opciones de rotación oblicua dadas en el programa SPSS. Las tablas F ("matriz de configuración") y G ("matriz de estructura") corresponden a la rotación oblicua *oblimin*, con el criterio de *normalización* de Kaiser. Recuérdese que en la *rotación oblicua* ambas matrices varían, a diferencia de la *rotación ortogonal*. Como el procedimiento *oblimin* permite controlar la extensión de la "oblicuidad" mediante *delta* (" δ "), éste se fija en 0, siguiendo la recomendación de Harman (1976) de que *delta* sea 0. La oblicuidad es mayor cuando *delta* es 0 y menor cuando es negativo.

La *matriz de estructura* (tabla G) muestra las correlaciones entre las variables y los componentes. Se obtiene multiplicando la matriz de configuración por la matriz de correlaciones de los componentes. Recuérdese que una característica que distingue a la *rotación oblicua* es que, al permitir que los componentes estén intercorrelacionados, los coeficientes dejan de ser simples coeficientes de correlación de la variable con el componente, sino coeficientes de regresión de cada variable empírica respecto a los componentes que están mutuamente relacionados.

De la comparación de las matrices con las obtenidas con la rotación ortogonal puede concluirse que la composición de los componentes principales no difiere, en general, de los modelos precedentes; si bien, ha logrado una definición más clara de los componentes, al haber menos variables que saturan de forma significativa en más de un componente. La diferencia principal es el intercambio en la composición de los componentes 2 (ahora caracterizado por la combinación de las variables "estudios", "ingresos" y "edad") y 3 ("casar con marroquí", "vecino marroquí", "simpatía marroquí" y "partido racista").

Tabla F. Matriz de configuración^a

	Componente			
	1	2	3	4
leyes inmigración	-,661	4,666E-02	8,862E-03	-9,55E-02
entrada inmigrante	,644	5,982E-03	,113	1,834E-02
regularizar inmigrante	-,596	1,140E-02	-,119	5,669E-02
n.º inmigrantes	,563	-,206	2,171E-03	8,480E-02
inmigrante delincuente	,557	-,103	5,938E-02	7,125E-02
estudios	-,142	,794	1,333E-02	4,728E-02
ingresos	-,140	,770	,115	,168
edad	-3,74E-02	-,667	,116	,109
casar con marroquí	-1,36E-02	-1,69E-03	,858	1,085E-02
vecino marroquí	-4,43E-02	1,352E-02	,857	-6,12E-02
simpatía marroquí	-,223	-6,26E-03	-,626	9,649E-02
partido racista	,198	1,610E-02	,287	,275
ideología política	,261	,168	-,104	,856
sexo	,408	,301	-8,24E-02	-,463

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización oblimin con Kaiser.

^a La rotación ha convergido en 8 iteraciones.

Tabla G. Matriz de estructura

	Componente			
	1	2	3	4
entrada inmigrante	,687	-,114	,360	8,633E-02
leyes inmigración	-,673	,165	-,266	-,150
regularizar inmigrante	-,639	,113	-,337	-1,16E-02
n.º inmigrantes	,603	-,309	,259	,158
inmigrante delincuente	,602	-,210	,298	,139
estudios	-,260	,808	-,135	-8,07E-02
ingresos	-,207	,752	-7,00E-03	6,298E-02
edad	,121	-,693	,209	,229
casar con marroquí	,314	-,112	,855	,169
vecino marroquí	,276	-8,29E-02	,829	,102
simpatía marroquí	-,453	9,597E-02	-,693	-3,55E-02
partido racista	,324	-9,40E-02	,411	,340
ideología política	,259	1,055E-02	,133	,831
sexo	,295	,317	-5,22E-02	-,493

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización oblimin con Kaiser.

Al permitirse que los componentes estén correlacionados, las sumas de los cuadrados de las saturaciones no se pueden añadir para obtener la varianza total, a diferencia de los procedimientos de rotación *ortogonales*. Sólo se informa de los *autovalores* totales: componente 1 = 2,822, componente 2 = 2,030, componente 3 = 2,618 y componente 4 = 1,227.

La interrelación de los componentes también altera la interpretación de las *comunalidades*, cuyo valor no coincide con la suma de las saturaciones al cuadrado de la variable en cada componente. La divergencia es mayor, cuanto más correlacionados estén los componentes. Para constatarlo, calcúlese las *comunalidades* y compárense con las expuestas en el apartado 5.6, que son las proporcionadas por el programa para todo ACP, indistintamente del método de *rotación* aplicado.

Las correlaciones de los componentes aparecen en la tabla H. Dicha matriz no coincide con la matriz *identidad* (caracterizada por tener la diagonal principal integrada por unos y el resto por ceros), aunque las correlaciones no son elevadas. Ello explica la poca variación en la composición de los componentes respecto a la *rotación ortogonal*. Los componentes más correlacionados son el 1 y el 3 (.380) y los menos el 1 con el 4 (.07455). Estos dos últimos componentes son casi ortogonales al estar escasamente correlacionados. El solapamiento en la varianza de ambos es casi nula ($.07455^2 = .0056 = 5,6\%$) mientras que el solapamiento en la varianza de los componentes 1 y 3 alcanza el 14,44% (.380²). Adviértase que en la mayoría de las situaciones la inexistencia de correlación entre los componentes es una propiedad deseable porque supone que los componentes extraídos realmente miden dimensiones diferentes en los datos.

Tabla H. Matriz de correlaciones de los componentes

Componente	1	2	3	4
1	1,000	-,159	,380	7,455E-02
2	-,159	1,000	-,130	-,151
3	,380	-,130	1,000	,186
4	7,455E-02	-,151	,186	1,000

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización oblimin con Kaiser.

La escasa correlación de los componentes puede llevar a reconsiderar la conveniencia de proceder a la *rotación oblicua*. Se prueba la *rotación promax*, obteniéndose correlaciones ligeramente superiores entre los componentes. La más elevada (tabla I) se da entre los componentes 1 y 2 (.418) y la más baja entre los componentes 1 y 4 (.113). Ello se debe a que mediante la *rotación promax* se obtiene una composición similar de los componentes a la conseguida mediante los procedimientos *ortogonales*. El componente 2 está mayoritariamente integrado por la combinación lineal de las cuatro variables que miden "simpatía" hacia los inmigrantes y el componente 3 por los indicadores de "posición social". Si se comparan las tablas J y K con las correspondientes a la *rotación oblimin* (tablas F y G) se observará que ambos componentes están intercambiados en ambas soluciones factoriales.

Tabla I. Matriz de correlaciones de los componentes

Componente	1	2	3	4
1	1,000	,418	-,187	,113
2	,418	1,000	-,145	,198
3	-,187	-,145	1,000	-,142
4	,113	,198	-,142	1,000

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización promax con Kaiser.

Tabla J. Matriz de configuración^a

	Componente			
	1	2	3	4
leyes inmigración	-,675	4,205E-02	3,795E-02	-8,83E-02
entrada inmigrante	,659	8,098E-02	1,581E-02	1,075E-02
regularizar inmigrante	-,611	-9,05E-02	1,893E-03	6,372E-02
n.º inmigrantes	,573	-2,74E-02	-,199	8,036E-02
inmigrante delincuente	,569	3,086E-02	-9,55E-02	6,586E-02
vecino marroquí	-3,81E-02	,858	2,006E-02	-5,05E-02
casar con marroquí	-7,68E-03	,856	5,008E-03	1,154E-02
simpatía marroquí	-,235	-,615	-1,49E-02	9,907E-02
partido racista	,200	,275	2,002E-02	,273
estudios	-,140	2,578E-02	,793	4,123E-02
ingresos	-,140	,126	,769	,163
edad	-4,41E-02	,112	-,668	,117
ideología política	,254	-,121	,167	,853
sexo	,426	-9,66E-02	,308	-,472

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización promax con Kaiser.

^a La rotación ha convergido en 6 iteraciones.

Tabla K. Matriz de estructura

	Componente			
	1	2	3	4
entrada inmigrante	,691	,357	-,120	9,884E-02
leyes inmigración	-,674	-,263	,170	-,161
regularizar inmigrante	-,642	-,334	,120	-2,34E-02
n.º inmigrantes	,608	,257	-,314	,168
inmigrante delincuente	,607	,296	-,216	,150
casar con marroquí	,351	,855	-,120	,180
vecino marroquí	,311	,829	-9,02E-02	,113
simpatía marroquí	-,478	-,691	,104	-4,72E-02
partido racista	,342	,410	-9,61E-02	,347
estudios	-,273	-,140	,810	-8,23E-02
ingresos	-,212	-1,16E-02	,754	6,265E-02
edad	,141	,214	-,693	,229
ideología política	,269	,130	1,580E-02	,834
sexo	,275	-5,66E-02	,310	-,487

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización promax con Kaiser.

• Con cualquiera de los procedimientos de *rotación* puede cubrirse el objetivo de lograr un modelo factorial ajustado al principio de estructura “simple”, integrado por dimensiones latentes claramente definidas mediante los indicadores utilizados. Si bien, la *rotación ortogonal*, en general, y el procedimiento *varimax*, en particular, son los más aplicados en la investigación empírica. Y ello pese a que la *rotación oblicua* se ajusta más a situaciones habituales de dimensiones latentes “interrelacionadas”.

El éxito de los procedimientos de *rotación ortogonal* en gran parte se debe a su amplia difusión en la generalidad de los programas estadísticos. También ha contribuido su mayor sencillez de realización y de interpretación. Mientras en la *rotación ortogonal* el modelo factorial se obtiene de la interpretación de la matriz factorial, en la *rotación oblicua* se precisa conocer las matrices factoriales, de *estructura* y la de *correlación* entre factores. Además, la sola interpretación de la *matriz factorial* es más compleja, al estar sus datos referidos a factores o componentes interrelacionados.

A las ventajas destacadas de la *rotación ortogonal* hay que añadir su mayor adecuación, cuando el análisis factorial se aplica como paso previo a otros análisis multivariados. Sirva de ejemplo su aplicación para eliminar la *multicolinealidad* en análisis como el de regresión. Para este propósito se precisa obtener dimensiones latentes *ortogonales* (no correlacionadas), que agrupen a indicadores muy correlacionados entre sí. Lo mismo sucede cuando se quiere formar grupos (de sujetos o variables) muy cohesionados entre sí y diferentes de los otros grupos. En estas situaciones la *rotación ortogonal* se convierte en la mejor opción, al adecuarse a los propósitos del análisis. En cambio, la *rotación oblicua* se adecua más cuando “el analista simplemente está interesado en obtener constructos o dimensiones teóricamente significativos” porque, como afirman Hair *et al.* (1992: 234-235), “es teórica y empíricamente más realista”.

En situaciones de indecisión en la elección del procedimiento de *rotación*, lo más conveniente es seguir la recomendación, destacada por Tabachnick y Fidell (1989: 637), de “solicitar la *rotación oblicua* con el número deseado de factores y observar las correlaciones entre los factores. Si éstas exceden de ,30, existe un 10% (o más) de solapamiento en la varianza entre los factores. Esta varianza se considera suficiente para garantizar la *rotación oblicua*, a menos que existan razones apremiantes para la *rotación ortogonal*”.

Este proceder convierte a la observación de la *matriz de correlación factorial* en clave en la decisión de qué modalidad de *rotación* realizar. Si esta matriz coincide con la *matriz identidad*, significa que los factores no se hallan correlacionados entre sí, lo que favorece el uso de la *rotación ortogonal*. Por el contrario, cuando las correlaciones entre los factores son $\geq ,30$, la *rotación oblicua* se convierte en la opción deseable en la consecución de una estructura factorial más ajustada a la realidad observada.

5.6.2. Lectura e interpretación de la matriz factorial

De la *matriz factorial* (preferiblemente la *rotada*) se obtiene la configuración del modelo factorial: las variables empíricas que conforman cada dimensión latente. Es-

ta información la dan los pesos o coeficientes factoriales (los "factor loadings" o "component loadings") de cada variable empírica en cada factor o componente. La magnitud de dicho coeficiente es lo que define la composición de cada dimensión, que agrupa a las variables que ponderan en ella. Normalmente las variables empíricas se integran en aquella dimensión latente en la que mayor coeficiente factorial presenten. Pero la asignación de variables empíricas (o indicadores) a factores o componentes ni es siempre sencilla, ni se halla libre de polémica.

Primero, ha de decidirse a partir de qué valor el *coeficiente factorial* se considera "significativo". O, dicho en otros términos, qué cantidad mínima de la varianza de una variable ha de ser explicada por un factor o componente. La propuesta más comparada es tomar como valor mínimo $\pm ,30$. Todo coeficiente factorial $\geq \pm ,45$ se estima "significativo" y a partir de $\pm ,50$, como "muy significativo". Cuanto más se aproxime el coeficiente factorial a 1,00, mayor es la relación de la variable con el factor. Comrey (1973) califica a los coeficientes $\geq \pm ,70$ (50% de varianza que se superpone) de "excelentes"; los $\geq \pm ,63$ (40% de varianza) de "muy buenos"; $\geq \pm ,55$ (30% de varianza) "bueno"; $\geq \pm ,45$ (20% de varianza) "justo"; y $\geq \pm ,32$ (10% de varianza) "pobre", aunque suficiente. Tabachnick y Fidell (1989) sólo señalan "como regla" interpretar las variables con coeficientes factoriales $\geq \pm ,30$.

En tamaños muestrales superiores a 100 casos, estos referentes pueden reducirse. En muestras inferiores deben, en cambio, aumentarse. Pero, en la decisión de qué punto de corte adoptar en la valoración de un coeficiente factorial como "significativo" no sólo interviene el tamaño muestral. También ha de considerarse el número de variables empíricas. El aumento del número de variables, al igual que el incremento del tamaño muestral, permite rebajar la cuantía mínima del coeficiente para considerarlo "significativo". Hair *et al.* (1992: 239-240) incluyen ambos aspectos en su propuesta:

- Para tamaños muestrales ≥ 50 casos, los coeficientes factoriales $\geq \pm ,30$ son "significativos"; $\geq \pm ,40$, "importantes"; y $\geq \pm ,50$, "muy significativos".
- Para tamaños muestrales de 100 casos, se recomiendan coeficientes factoriales $\geq \pm ,19$ y $\geq \pm ,26$ para niveles de significación de ,05 y ,01, respectivamente. Si la muestra es de 200 casos, los coeficientes recomendados son ligeramente inferiores: $\geq \pm ,14$ y $\geq \pm ,18$; Cuando la muestra es superior a 300 casos, los coeficientes factoriales de referencia descienden aún más: $\geq \pm ,11$ y $\geq \pm ,15$ para los niveles de significación de ,05 y ,01.
- Si se considera además el número de variables y de factores o componentes, los coeficientes factoriales de referencia también varían. El pasar del primer factor a los posteriores ha de suponer, igualmente, un incremento del nivel aceptable para que el coeficiente se considere "significativo". Así, por ejemplo, para un tamaño muestral de 100 unidades y un nivel de significación de ,05, un coeficiente factorial significativo en el factor número 5, con 20 variables empíricas, es $\geq \pm ,216$; pero, con 50 variables empíricas, desciende ligeramente a $\geq \pm ,202$. Para el factor décimo, con 20 variables, el coeficiente mínimo necesario aumenta a $\geq \pm ,261$; y con 50 variables desciende a $\geq \pm ,214$.

En resumen, los incrementos en el tamaño de la muestra y en el número de variables empíricas repercuten, directamente, en un ligero descenso en la cuantía del coeficiente factorial para que se juzgue “significativo”. Por el contrario, el aumento del número de factores supone un incremento en la magnitud de los coeficientes en los últimos factores para que puedan considerarse “significativos”. Sólo se valora la cuantía del coeficiente para conocer el grado de relación de la variable empírica con la dimensión latente. Los signos se interpretan como en cualquier coeficiente de correlación. Indican la dirección de la relación entre el indicador y la dimensión latente. El signo *positivo* expresa que la relación es positiva: ambas variables, la observada y la latente, avanzan en la misma dirección; el aumento o disminución del valor de una de ellas provoca el aumento o disminución de la otra. El signo *negativo* expresa, en cambio, que la relación es negativa: al aumento de una variable le sigue una disminución en el valor de la otra, en la cuantía expresa en el coeficiente, y, a la inversa, la disminución de una de ellas supone el aumento de la otra.

Es habitual que una misma dimensión latente incluya coeficientes factoriales positivos en unas variables y negativos en otras. Cuando esto sucede, se dice que el factor o componente es “bipolar”.

Una vez decidida la cuantía mínima para que el coeficiente factorial sea “significativo”, se procede a la lectura comparativa de los coeficientes factoriales de cada variable en cada factor. Para cada variable se destaca (subrayándolo, rodeándolo con un círculo, poniendo un asterisco, o como se quiera) el coeficiente de mayor magnitud en valor absoluto. Si alguna de las variables empíricas no presenta ningún coeficiente factorial que sea “significativo”, puede considerarse su eliminación del modelo factorial. Esta decisión suele adoptarse cuando la variable presenta, además, una baja *comunalidad* o se considera irrelevante para los objetivos de investigación (Tabachnick y Fidell, 1989). El adoptar esta decisión conlleva la repetición de los análisis, excluyendo las variables eliminadas. Salvo que el investigador decida, deliberadamente, interpretar la matriz factorial original (que incluye variables con coeficientes factoriales no significativos), sin introducir ninguna modificación; es decir, ignorando aquellas variables con coeficientes factoriales no significativos.

Para facilitar la interpretación de los factores se puede proceder, en tercer lugar, a agrupar las variables empíricas con coeficientes factoriales elevados para cada factor. El significado sustantivo de un factor lo dan las variables con mayor coeficiente factorial. Se debe buscar qué significado “común” comparten estas variables para, posteriormente, asignar un nombre o *etiqueta* al factor. La *etiqueta* ha de reflejar el significado común de las variables que ponderan en el factor. Obviamente, las variables con mayor coeficiente factorial serán las que más influyan en el nombre del factor.

Ésta es la fase del análisis factorial de mayor “subjetivismo”. El investigador es libre de asignar el nombre que quiera al componente o factor. La única condición que se impone es que la *etiqueta* elegida refleje y sintetice el contenido “común” de las variables que ponderen en dicha dimensión latente.

EJEMPLO DE INTERPRETACIÓN DE DISTINTOS MODELOS AFC

Expuestos los modelos ACP, queda describir las soluciones de *factor común* (AFC). Para abreviar la exposición, sólo se interpreta el modelo final obtenido tras aplicar el procedimiento de rotación de los ejes factoriales más popular: *varimax*. La tabla A corresponde a la matriz de factores rotados para el método de *eje o factor principal*. Éste es el método de extracción factorial de mayor aplicación en AFC, junto al de *máxima verosimilitud* (Kim y Mueller, 1978b; Dillon y Goldstein, 1984; Afifi y Clark, 1990). Como su realización ha sido posterior a ACP, se ha seguido el mismo criterio de extraer 4 dimensiones latentes (ahora llamadas "factores comunes") que expliquen las correlaciones de una serie de variables observadas. Como se dijo en el subapartado 5.2.2, en AFC se diferencia entre *varianza común* y *específica*, de modo que cada variable observada queda definida por la conjunción de factores "comunes" y "únicos". También se busca un modelo sencillo e interpretable. Lo que lleva a la *rotación* para que cada variable se relacione con el menor número de factores posible.

Tabla A. Matriz de factores rotados^a

	Factor			
	1	2	3	4
entrada inmigrante	,593	,196	-1,61E-02	5,744E-02
leyes inmigración	-,541	-,143	7,588E-02	-,116
regularizar inmigrante	-,523	-,192	2,934E-02	-1,31E-02
n.º inmigrantes	,497	,136	-,182	7,120E-02
inmigrante delincuente	,451	,180	-,127	,136
partido racista	,266	,238	-5,80E-02	,156
casar con marroquí	,214	,794	-5,12E-02	8,784E-02
vecino marroquí	,223	,656	-2,51E-02	4,879E-02
simpatía marroquí	-,385	-,479	2,121E-02	-3,54E-02
estudios	-,244	-4,14E-02	,741	1,698E-02
ingresos	-,232	4,845E-02	,565	,155
edad	,107	,123	-,536	,143
sexo	8,615E-02	1,068E-03	,161	-3,01E-02
ideología política	,162	8,049E-02	-2,83E-02	,549

Método de extracción: Factorización del eje principal.

Método de rotación: Normalización *varimax* con Kaiser.

^a La rotación ha convergido en 6 iteraciones.

En la *rotación* se ha aplicado, igualmente, el criterio de *normalización* de Kaiser y han convergido 6 iteraciones. Si se compara esta solución con la obtenida mediante ACP con la rotación *varimax* (subapartado 5.6.1), puede observarse que la composición de los factores es similar a la de los componentes con dos excepciones evidentes: una, la variable "sexo", que no presenta ningún coeficiente factorial significativo en alguno de los cuatro fac-

tores (el coeficiente más elevado es ,161 y corresponde al factor 3); y dos, la variable "partido racista", con coeficientes algo más elevados, pero también inferiores al referente habitual ($\pm ,30$).

En busca de una explicación, se analizan las *comunalidades* (expuestas en el subapartado 5.4.3). En AFC, las *comunalidades* "iniciales" difieren de 1,0, al analizarse sólo la *varianza común* de cada variable y no la total como en ACP. Sus estimaciones de las *comunalidades iniciales* son, en cada variable, R^2 múltiple que resulta de regresionar el resto de las variables con la considerada. En la susodicha tabla puede comprobarse que la variable "casar con marroquí" es, de las 14 consideradas, la que mayor proporción de su *varianza* logra ser explicada por la combinación de las 13 variables restantes ($,429 = 42,9\%$ de su *varianza*). Le sigue la variable "vecino marroquí" ($,368$) y "estudios" ($,351$). Por el contrario, las variables "ideología política" ($,08422$) y "sexo" ($,033$) apenas tienen *varianza* compartida con las demás variables. El conocimiento del valor de las otras variables no ayuda a su predicción. Recuérdese que en el análisis de conglomerados (capítulo 3) la variable "ideología política" fue la última en unirse a la agrupación de las variables en conglomerados; la variable "sexo", la penúltima.

Como las *comunalidades* "iniciales" se distancian de 1,00, se prevé que la solución de factorización de *ejes principales* difiera de la obtenida con ACP.

Tras la "extracción" factorial se vuelven a estimar las *comunalidades* considerando ahora los coeficientes factoriales. Las nuevas estimaciones de las *comunalidades* reemplazan a las antiguas y prosigue el proceso iterativo de estimación de los factores comunes. Éste concluye cuando apenas cambian las estimaciones de *comunalidad*.

Como se preveía, tras la extracción factorial, las *comunalidades* finales también difieren de las obtenidas en ACP. Los resultados confirman la tendencia habitual a conseguir *comunalidades* en general inferiores en la factorización de *ejes principales* respecto a ACP. La razón está en que sólo se factoriza la *varianza* "común".

Tras la extracción factorial las *comunalidades* indican la proporción de la *varianza* de las variables que puede ser explicada por los factores comunes. El valor 1,00 expresa que toda la *varianza* de la variable logra ser explicada por los factores comunes; el valor ,00, que su *variabilidad* no se consigue explicar. La *varianza* "no" explicada por los factores comunes se atribuye al componente "único" o *varianza* "específica" de la variable.

La nula *comunalidad* de la variable "sexo" (véase la tabla B del ejemplo del subapartado 5.4.3) lleva a considerar la necesidad de eliminarla de los análisis y volver a estimar el modelo factorial, en busca de su mejora. Máxime cuando dicha variable no presenta ningún coeficiente factorial "significativo" en alguno de los factores comunes. Además, su relevancia para los objetivos del estudio es escasa.

También puede considerarse la eliminación de la variable "partido racista", con coeficientes algo superiores, pero igualmente no significativos. Asimismo, su *comunalidad* tras la "extracción" es baja: sólo el 15,5% de su *varianza* queda explicada por los cuatro factores comunes. Pese a ello, su eliminación no es tan evidente, a diferencia de la variable "sexo", al haber quedado muy definida en un factor: el factor 4, con un coeficiente factorial de ,549 (tabla A), que supone el 30,1% ($,549^2$) de su *variabilidad*. En general, se recomienda valorar la eliminación de las variables cuya *comunalidad* sea inferior a ,50 porque su *variabilidad* no queda suficientemente explicada por el modelo factorial. Si su valor es inferior a ,30, la eliminación de la variable se convierte en imperiosa.

La tabla B ofrece los datos de *varianza* total explicada por la factorización de *ejes principales*. El porcentaje de *varianza* total explicada por la conjunción de los cuatro factores co-

munes (35,937%) es bastante inferior al obtenido en ACP (52,951%). Compárense los porcentajes de varianza de las dos métodos factoriales.

Tabla B. Varianza total explicada

Factor	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones el cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	3,619	25,849	25,849	3,022	21,588	21,588	1,842	13,154	13,154
2	1,665	11,891	37,740	1,137	8,122	29,709	1,520	10,854	24,008
3	1,114	7,958	45,698	,541	3,864	33,574	1,247	8,904	32,912
4	1,015	7,253	52,951	,331	2,363	35,937	,423	3,024	35,937
5	,950	6,786	59,737						
6	,868	6,199	65,936						
7	,752	5,372	71,308						
8	,736	5,259	76,567						
9	,673	4,805	81,373						
10	,600	4,283	85,656						
11	,574	4,103	89,758						
12	,559	3,994	93,752						
13	,470	3,361	97,113						
14	,404	2,887	100,000						

Método de extracción: Factorización de ejes principales.

La aplicación del método de *máxima verosimilitud* logra elevar hasta 40,376 el porcentaje de varianza total explicada (tabla C) con una composición similar en sus factores (tabla D). Como en ACP, la proporción de varianza explicada en cada factor se obtiene de la suma de las saturaciones al cuadrado de todas las variables. De nuevo, las correlaciones son ponderadas por el inverso de la varianza *única* de las variables mediante un proceso iterativo.

Tabla C. Varianza total explicada

Factor	Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones el cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	1,203	8,592	8,592	1,876	13,400	13,400
2	2,735	19,532	28,124	1,502	10,731	24,131
3	1,131	8,081	36,205	1,269	9,065	33,196
4	,584	4,171	40,376	1,005	7,181	40,376

Método de extracción: Máxima verosimilitud.

Tabla D. Matriz de factores rotados^a

	Factor			
	1	2	3	4
entrada inmigrante	,599	,186	-2,24E-02	2,245E-02
leyes inmigración	-,547	-,141	8,037E-02	-7,79E-02
regularizar inmigrante	-,526	-,179	4,674E-02	-8,40E-03
n.º inmigrantes	,495	,132	-,186	4,206E-02
inmigrante delincuente	,456	,186	-,131	5,530E-02
partido racista	,283	,227	-6,55E-02	,105
casar con marroquí	,225	,809	-5,68E-02	4,767E-02
vecino marroquí	,239	,638	-3,23E-02	2,895E-02
simpatía marroquí	-,390	-,474	3,749E-02	-1,39E-02
estudios	-,231	-2,68E-02	,769	1,792E-02
ingresos	-,214	6,269E-02	,550	9,000E-02
edad	,114	,125	-,529	7,109E-02
sexo	9,529E-02	-1,80E-02	,153	-4,00E-02
ideología política	,166	7,487E-02	-3,76E-02	,982

Método de extracción: Máxima verosimilitud.

Método de rotación: Normalización varimax con Kaiser.

^a La rotación ha convergido en 6 iteraciones.

Para abreviar la exposición de los resultados sólo se informa de las matrices factoriales rotadas de los cuatro métodos de factorización restantes: *mínimos cuadrados generalizados* (tabla E), *mínimos cuadrados no ponderados* (tabla F), factorización *alfa* (tabla G) e *imagen* (tabla H). Compárense los cuatro modelos factoriales y calcúense las *comunalidades* y el porcentaje de la varianza total explicada por cada factor. Excepto en la factorización *alfa*, que se distingue porque ni las *comunalidades* ni los *autovalores* son exactamente la suma de los coeficientes factoriales al cuadrado. En cambio, la factorización *imagen* se diferencia porque las *comunalidades* no varían, al "no" ser iterativamente reestimadas.

El considerar que se analiza a la población de individuos y/o variables también afecta a que en la factorización *alfa* no se realicen pruebas de significación, a diferencia de *máxima verosimilitud* y *mínimos cuadrados generalizados*. Asimismo, a diferencia del método de *máxima verosimilitud*, en la factorización *alfa* es frecuente que las variables con menor *comunalidad* obtengan coeficientes factoriales más elevados que mediante otros métodos factoriales como *máxima verosimilitud*. Para comprobar dichas afirmaciones se proporcionan las *comunalidades* (tabla I) y la varianza total explicada (tabla J) mediante la factorización *alfa*.

Como muestra de la similitud de la agrupación de variables, compárense los gráficos de saturaciones en el espacio rotado de ACP (figura A) y AFC de *ejes principales* (figura B). En ambos gráficos las coordenadas se corresponden con los pesos o coeficientes factoriales de la solución obtenida tras la rotación *varimax*. Estos gráficos 3-D son la ilustración usual cuando son tres o más los componentes o factores extraídos. En nuestro caso son cuatro. El gráfico incluye los coeficientes de los tres primeros componentes (ACP) o factores comunes (AFC). La disposición de las variables en el espacio rotado es muy similar y son varias las va-

Tabla E. Matriz de factores rotados^a

	Factor			
	1	2	3	4
entrada inmigrante	,606	,189	-4,81E-02	2,677E-02
leyes inmigración	-,543	-,146	,106	-8,43E-02
regularizar inmigrante	-,521	-,185	7,158E-02	-1,42E-02
n.º inmigrantes	,489	,138	-,211	5,017E-02
inmigrante delincuente	,450	,191	-,151	6,206E-02
partido racista	,278	,231	-7,36E-02	,109
casar con marroquí	,214	,813	-5,46E-02	5,119E-02
vecino marroquí	,230	,644	-3,40E-02	3,194E-02
simpatía marroquí	-,382	-,479	5,037E-02	-1,83E-02
estudios	-,188	-4,16E-02	,776	-3,73E-03
ingresos	-,184	5,301E-02	,570	7,419E-02
edad	8,118E-02	,136	-,536	8,569E-02
sexo	,105	-1,96E-02	,149	-4,31E-02
ideología política	,156	7,628E-02	-1,99E-02	,984

Método de extracción: Mínimos cuadrados generalizados.

Método de rotación: Normalización varimax con Kaiser.

^a La rotación ha convergido en 6 iteraciones.

Tabla F. Matriz de factores rotados^a

	Factor			
	1	2	3	4
entrada inmigrante	,599	,189	-1,98E-02	2,553E-02
leyes inmigración	-,545	-,142	8,011E-02	-7,41E-02
regularizar inmigrante	-,523	-,183	3,555E-02	-6,86E-03
n.º inmigrantes	,499	,134	-,186	4,159E-02
inmigrante delincuente	,461	,183	-,128	6,135E-02
partido racista	,275	,242	-5,88E-02	,100
casar con marroquí	,224	,800	-5,16E-02	4,556E-02
vecino marroquí	,233	,649	-2,58E-02	2,636E-02
simpatía marroquí	-,392	-,473	2,453E-02	-1,32E-02
estudios	-,235	-3,77E-02	,762	1,940E-02
ingresos	-,214	5,666E-02	,556	8,935E-02
edad	,119	,128	-,521	7,101E-02
sexo	8,913E-02	-2,81E-03	,161	-3,39E-02
ideología política	,164	8,005E-02	-3,80E-02	,982

Método de extracción: Mínimos cuadrados no ponderados.

Método de rotación: Normalización varimax con Kaiser.

^a La rotación ha convergido en 6 iteraciones.

Tabla G. Matriz de factores rotados^a

	Factor			
	1	2	3	4
entrada inmigrante	,594	,191	-4,53E-02	5,082E-02
leyes inmigración	-,529	-,123	,104	-7,98E-02
regularizar inmigrante	-,509	-,185	5,29E-02	-1,75E-02
inmigrante delincuente	,507	,143	-,111	5,830E-02
n.º inmigrantes	,455	,144	-,243	9,117E-02
partido racista	,279	,239	-4,95E-02	,116
casar con marroquí	,205	,804	-9,45E-02	8,424E-02
vecino marroquí	,248	,665	-1,64E-02	3,016E-02
simpatía marroquí	-,420	-,449	1,503E-02	-3,25E-03
estudios	-,237	-4,30E-02	,718	2,256E-02
ingresos	-,158	3,336E-02	,646	8,312E-02
edad	,140	,105	-,457	7,188E-02
sexo	,109	-4,05E-03	,174	-5,42E-02
ideología política	,165	8,517E-02	-3,47E-02	,740

Método de extracción: Factorización alfa.

Método de rotación: Normalización varimax con Kaiser.

^a La rotación ha convergido en 5 iteraciones.

Tabla H. Matriz de factores rotados^a

	Factor			
	1	2	3	4
casar con marroquí	,574	,219	-4,17E-02	3,064E-02
vecino marroquí	,539	,189	-2,81E-02	2,008E-02
simpatía marroquí	-,457	-,292	4,519E-02	-6,41E-03
partido racista	,261	,217	-6,15E-02	6,135E-02
entrada inmigrante	,281	,419	-7,18E-02	1,793E-02
leyes inmigración	-,236	-,403	,107	-2,25E-02
n.º inmigrantes	,216	,380	-,184	1,144E-02
regularizar inmigrante	-,260	-,378	7,323E-02	1,530E-03
inmigrante delincuente	,244	,354	-,135	5,058E-02
ideología política	,145	,167	-2,00E-02	,120
estudios	-8,65E-02	-,218	,495	-2,00E-03
ingresos	-2,29E-03	-,180	,436	4,238E-02
edad	,128	,132	-,400	5,239E-02
sexo	1,336E-02	5,488E-02	,123	-6,06E-03

Método de extracción: Factorización imagen.

Método de rotación: Normalización varimax con Kaiser.

^a La rotación ha convergido en 8 iteraciones.

Tabla I. Comunalidades

	Extracción
simpatía marroquí	,378
leyes inmigración	,313
ideología política	,583
sexo	4,497E-02
edad	,244
n° inmigrantes	,295
regularizar inmigrante	,296
entrada inmigrante	,394
partido racista	,151
casar con marroquí	,704
estudios	,575
ingresos	,451
vecino marroquí	,505
inmigrante delincuente	,294

Método de extracción: Factorización alfa.

Tabla J. Varianza total explicada

Factor	Suma de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	3,006	21,473	21,473	1,854	13,241	13,241
2	1,030	7,357	28,830	1,495	10,679	23,920
3	,631	4,510	33,340	1,273	9,090	33,010
4	,559	3,991	37,331	,605	4,321	37,331

Método de extracción: Factorización alfa.

riables que convergen en un mismo punto. Su disposición por encima o por debajo del punto 0,0 depende de la correlación de la variable con el componente o factor: positiva (por encima) o negativa (por debajo). El ideal es que las variables se agrupen al final de los ejes o en su intersección. Ello significa que se ha alcanzado la finalidad perseguida con la *rotación*: la obtención de una "estructura simple". Pero, una interpretación gráfica más detallada exige gráficos de cada dos componentes o factores por separado, como el incluido en el ejemplo del subapartado 5.6.1. Aunque con dichos gráficos se pierde la visión de conjunto, la representación a modo de síntesis del modelo. Por esta razón, se ha optado por los gráficos 3-D.

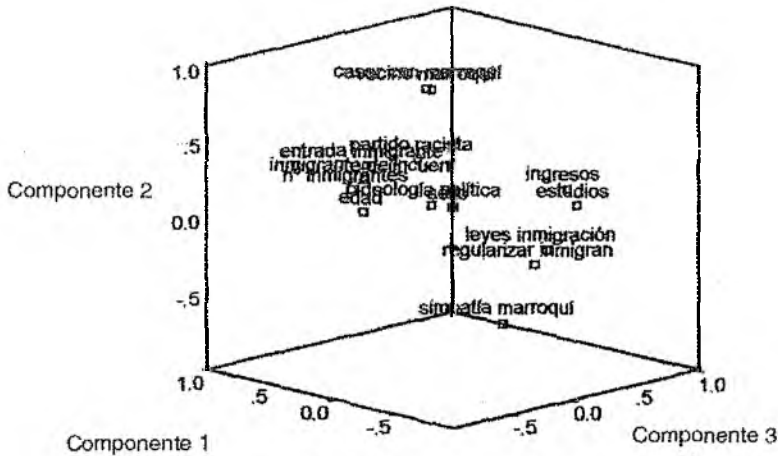


Figura A. Gráfico de componentes en espacio rotado.

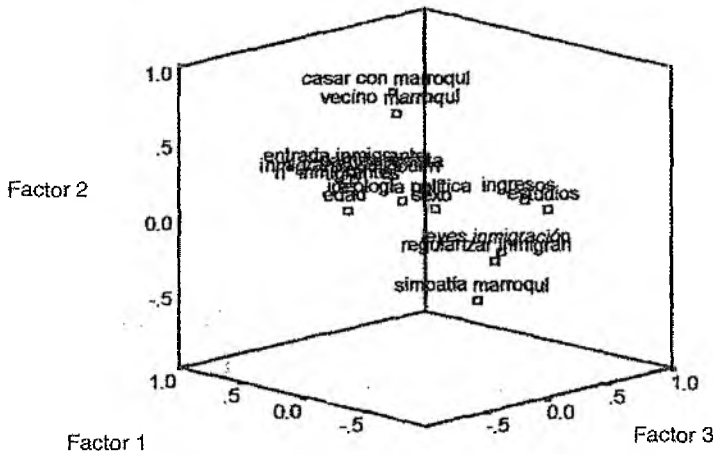


Figura B. Gráfico de saturaciones factorial rotado.

5.7. La evaluación del modelo factorial

En la evaluación del modelo factorial no sólo interviene la “significatividad estadística”, a semejanza de las demás variedades analíticas. También se valora la “significatividad sustantiva”. Esta última es genérica a todo modelo factorial. Hace referencia a la obtención de un modelo estadístico que tenga significado “sustantivo”. Esto

quiere decir que el modelo sea interpretable desde alguna perspectiva teórica; que tenga sentido lógico. En cambio, la "significatividad estadística" del modelo se valora en unos modelos factoriales concretos: de *máxima verosimilitud* y de *mínimos cuadrados generalizados*.

Como se dijo en el subapartado 5.5.2, la aplicación del contraste *chi-cuadrado* exige el cumplimiento del supuesto de *normalidad* multivariable. Además, su valor se ve muy afectado por el tamaño muestral. A medida que aumenta el tamaño de la muestra, se incrementa la probabilidad de que el contraste de hipótesis mediante χ^2 resulte "significativo". La no "significatividad estadística", a un nivel de significación suficientemente bajo ($\leq 0,05$, al menos), lleva a replantear todo el modelo analítico, a su modificación y posterior repetición de los análisis. Puede, por ejemplo, que la no "significatividad" se deba a la no consideración de un factor relevante, lo que puede provocar que se decida añadir un nuevo factor al modelo factorial. Esto revierte negativamente en una pérdida de *parsimonia*.

El efecto negativo del contraste χ^2 en tamaños muestrales elevados incide en su desconsideración como criterio principal en la evaluación del modelo factorial. En la mayoría de las situaciones la comprobación de la adecuación del modelo factorial, a los datos empíricos, se restringe al contraste de las diferencias entre las *correlaciones observadas* en la muestra con las *estimadas* a partir del modelo factorial obtenido. Estas diferencias de correlaciones (*observadas* y *estimadas*) son los *residuos*. Su magnitud se utiliza para afirmar o negar la adecuación del modelo. El modelo factorial se considera válido cuando los *residuos* son pequeños, porque significa que el ajuste del modelo a los datos es bueno. Por el contrario, *residuos* $\geq 0,05$ indican que el modelo factorial derivado no refleja la variabilidad de los datos observados, ante lo cual habría que proceder a su reconsideración.

La mayoría de los programas estadísticos proporciona información de los *residuos* en la *matriz de correlación residual*. Ésta se obtiene de la diferencia entre las matrices de correlación *observada* y la *reproducida* desde los factores extractados, como se ilustró en el segundo ejemplo del apartado 5.6. Lo que el investigador debe decidir es a partir de qué valor el *residuo* es "grande" o "pequeño". El valor adoptado con mayor frecuencia como referente es 0,05: El *residuo* se considera "pequeño", cuando su cuantía es inferior a dicha cantidad. Asimismo, se deja a juicio del investigador la concreción del porcentaje máximo de *residuos* "elevados" que llevan a la aceptación o, en su caso, rechazo del modelo factorial.

Por último, añadir que en la variedad factorial llamada "análisis factorial booleano" el ajuste del modelo se mide, igualmente, comparando las respuestas binarias *observadas* con las *estimadas*. El número total de discrepancias se obtiene de la suma de las diferencias de cada puntuación original X_{ij} respecto de sus correspondientes valores estimados \hat{X}_{ij} (Bisquerra, 1989: 344):

$$d = \sum_{i=1}^N \sum_{j=1}^p |\hat{X}_{ij} - X_{ij}|$$

5.8. Las puntuaciones factoriales

El análisis factorial no siempre concluye con la interpretación del modelo analítico y la asignación posterior de “etiquetas” a las dimensiones latentes, puede proseguir con el cálculo de las *puntuaciones factoriales*. Esto acontece, por ejemplo, cuando se quiere, a partir de los resultados factoriales, aplicar otras técnicas analíticas multivariantes (como regresión múltiple, el análisis discriminante o el de conglomerados). En ellas las *puntuaciones factoriales* actuarán como variables, en representación de los valores de los factores o componentes. Pero, ¿qué se entiende por puntuaciones factoriales?

Las *puntuaciones factoriales* pueden definirse como medidas compuestas de cada factor común o componente principal. Informan de la posición de cada caso concreto (variables, individuos, países, municipios, universidades...) en cada factor o componente. Hay tantas *puntuaciones* como factores para cada caso.

En la *matriz factorial* (preferiblemente la *rotada*) puede extraerse, de cada factor, una variable que lo represente en análisis estadísticos posteriores. Esta *variable* “típica” suele coincidir con la que presenta el mayor coeficiente factorial en dicho factor. Jolliffe (1986) propone que sea aquella de mayor coeficiente factorial entre las que presentan un valor “ $\lambda \geq ,70$ ”. Cuando existen dos o más variables que comparten un mismo coeficiente elevado, la elección de la variable “típica” depende, además, del significado sustantivo de las variables. La variable elegida suele coincidir con aquella que mejor representa, desde la vertiente lógico-sustantiva, la dimensión latente identificada.

Al investigador también se le ofrece otra posibilidad: el cálculo de *variables “sucedáneas”*. Especialmente, cuando es difícil la elección de una única variable en representación del factor. En la obtención de variables “sucedáneas” participan todas las variables con coeficientes factoriales elevados en el mismo factor. De ellas se calcula su total o promedio como representación del factor. Aunque, como advierten Hair *et al.* (1992; 1999), la obtención de variables *sucedáneas* (“surrogate variate”) únicamente es posible cuando se realiza una rotación *ortogonal*. Ésta garantiza que los factores sean “ortogonales”; es decir, incorrelacionados entre sí. Lo que concuerda con la finalidad principal perseguida con la realización de un análisis factorial previo a la ejecución de otras técnicas analíticas multivariantes (de dependencia o de interdependencia). Por ejemplo, el buen desarrollo de un análisis de regresión múltiple exige, como se vio en el capítulo 1, ausencia de *multicolinealidad* entre las variables predictoras. Si, en vez de emplear la serie amplia de variables originales, el investigador decide reemplazarlas por un número inferior de variables *compuestas*, obtenidas del modelo factorial, estas nuevas variables no deberán, asimismo, estar correlacionadas entre ellas. Por esta razón se demanda que los ejes factoriales se roten “ortogonalmente”.

Si, por el contrario, se desea que, en representación de cada dimensión latente, intervengan todas las variables que ponderan en ella con un coeficiente factorial significativo, se tendrá que proceder al cálculo de las *puntuaciones factoriales*.

Una manera sencilla de obtener *puntuaciones factoriales* consiste en sumar los productos de cada coeficiente factorial que satura en el factor por el valor de la variable en cada caso. Al estar las variables *estandarizadas* (Z_i), se evita que aquellas de mayor

heterogeneidad contribuyan más en el cálculo de las *puntuaciones factoriales*. Así, en el ejemplo ilustrado en el apartado 5.6 de un modelo factorial de dos factores y cinco variables empíricas, las *puntuaciones factoriales*, a partir de la matriz factorial rotada son, para cada caso, la suma de los productos del coeficiente por el valor estandarizado de la variable que satura significativamente en el factor: factor 1 = $0,980 Z_2 + 0,659 Z_3 + 0,709 Z_4$; factor 2 = $0,902 Z_1 + 0,834 Z_5$. Este cálculo de las *puntuaciones factoriales* se considera "selectivo" porque únicamente incluye las variables que definen al factor, al saturar significativamente en él. Las *puntuaciones* indican la ubicación espacial de cada unidad de análisis (caso o variable) en cada factor. Permiten conocer su disposición, si coincide o se distancia de la generalidad de las unidades analizadas.

En el *análisis factorial booleano* cada caso tiene la puntuación factorial de 1, si presenta una saturación distinta de cero en cualquiera de las variables en el factor. En caso contrario, su puntuación será 0.

La mayoría de los programas estadísticos convencionales ofrece otras opciones para calcular *puntuaciones factoriales*. La opción más habitual es el llamado "procedimiento de regresión". Supone el cálculo de una serie de ecuaciones de regresión múltiple; una por cada dimensión latente. En cada ecuación, los factores o componentes actúan como la variable dependiente y los indicadores o variables empíricas como las independientes. De la intercorrelación de cada variable con su correspondiente coeficiente factorial (preferiblemente "rotado") resulta una serie de coeficientes de regresión parciales. En el análisis factorial, estos coeficientes se denominan *coeficientes de puntuaciones factoriales*. Su interpretación es análoga a los coeficientes de regresión parcial estandarizados. Recuérdese que las variables se prefieren estandarizadas.

$$\hat{F}_{kj} = \sum_{i=1}^p W_{ki} Z_{ij}$$

Donde: " Z_{ij} " es el valor estandarizado de la variable i en el caso j .

" W_{ki} " es el coef. de la puntuación factorial para el factor K y la variable i .

Estos *coeficientes de puntuaciones factoriales* para cada variable en cada factor se extraen de la *matriz de coeficientes de puntuaciones factoriales*. Cada uno de ellos se multiplica por el valor estandarizado de la variable en el factor. Se suman los productos y se obtiene la puntuación factorial para cada caso.

Como las variables están estandarizadas, la puntuación media para cada dimensión latente es igual a 0. Por esta razón, las puntuaciones negativas (las inferiores a 0) se interpretan como puntuaciones bajas, situadas por debajo de la media.

Además del *procedimiento de regresión*, suelen ofertarse otros dos métodos alternativos para calcular *puntuaciones factoriales*, aunque de uso más restringido. Se trata del:

- a) *Método de Anderson-Rubin*. Se caracteriza por proporcionar puntuaciones factoriales no correlacionadas, aunque los factores originales estén correlacionados, lo cual no siempre sucede con el procedimiento de *regresión*. De es-

te último puede resultar la obtención de puntuaciones factoriales correlacionadas, aunque los factores sean ortogonales. Ésta es una de las razones que hacen que esta opción sea deseable, cuando se busca la obtención de puntuaciones factoriales “no” correlacionadas.

- b) *Método Barlett*. Aplica el procedimiento de *máxima verosimilitud*, que exige que los factores se ajusten a una distribución normal.

En ACP, con cualquiera de los tres procedimientos mencionados, se obtienen las mismas *puntuaciones factoriales*. *Puntuaciones* que no se consideran “estimadas”, sino “exactas”. En AFC, por el contrario, “no es posible ninguna solución exacta para los factores”. En todo caso se está ante “estimaciones” (Dillon y Goldstein, 1984; Nourisis, 1994).

EJEMPLO DE PUNTUACIONES FACTORIALES

Para la obtención de las puntuaciones factoriales se ha utilizado el procedimiento más usual y aplicado por defecto en el programa SPSS: el procedimiento de *regresión*. En él, los factores o componentes actúan como variable dependiente (en cada ecuación lineal) y las variables observadas como las independientes. Las tablas A y B corresponden a las matrices de los coeficientes por los que se multiplican, en cada caso, los valores de las variables (estandarizadas) para el cálculo de las puntuaciones en ACP (tabla A) y AFC (tabla B). La opción elegida de AFC es *ejes principales* por ser el método más popular.

Tabla A. Matriz de coeficientes para el cálculo de las puntuaciones en las componentes

	Componente			
	1	2	3	4
simpatía marroquí	-,027	-,315	-,025	,117
leyes inmigración	-,330	,115	-,039	-,030
ideología política	,114	-,137	,133	,778
sexo	,279	-,087	,202	-,476
edad	-,108	,059	-,378	,115
n.º inmigrantes	,263	-,095	-,060	,031
regularizar inmigrante	-,290	,027	-,051	,109
entrada inmigrante	,312	-,041	,067	-,043
partido racista	,043	,115	,034	,237
casar con marroquí	-,136	,476	,000	,003
estudios	,000	,036	,435	,046
ingresos	-,024	,086	,424	,158
vecino marroquí	-,147	,484	,005	-,053
inmigrante delinciente	,262	-,061	-,003	,017

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización varimax con Kaiser.

Puntuaciones de componentes.

Tabla A. Matriz de coeficientes para el cálculo de las puntuaciones factoriales

	Factor			
	1	2	3	4
simpatía marroquí	-,116	-,128	-,048	,046
leyes inmigración	-,236	,057	-,038	-,047
ideología política	,011	-,040	-,008	,510
sexo	,048	-,001	,064	-,029
edad	-,062	,026	-,240	,135
n.º inmigrantes	,197	-,042	-,017	,014
regularizar inmigrante	-,228	,022	-,065	,052
entrada inmigrante	,300	-,044	,082	-,020
partido racista	,053	,035	,007	,085
casar con marroquí	-,130	,611	-,020	,005
estudios	-,023	,028	,541	,040
ingresos	-,070	,050	,259	,171
vecino marroquí	-,024	,296	,014	-,032
inmigrante delincuente	,161	-,025	,000	,067

Método de extracción: Factorización del eje principal.

Método de rotación: Normalización várimax con Kaiser.

5.8.1. Su contribución en la detección de atípicos

Las *puntuaciones factoriales* cumplen otras funciones en el análisis factorial. Destaca su utilidad en la detección de *atípicos*, en la búsqueda de conglomerados y, en general, para comprender la estructura de los datos.

Los *atípicos* se identifican con variables de escasa correlación múltiple cuadrada con las demás variables y que presentan, asimismo, coeficientes factoriales próximos a cero en todos los factores. Los gráficos de *puntuaciones factoriales* para cada par de factores ayudan a la detección de estos *atípicos*. En dichos gráficos, las puntuaciones factoriales se sitúan en los ejes y los puntos representan los casos concretos. Los *atípicos* coinciden con los puntos que se ubican en los extremos del gráfico, como ilustra el ejemplo a continuación.

Como las *puntuaciones factoriales* se hallan estandarizadas, se considera *atípico* todo aquel que supere el valor 2,0, indistintamente de su signo (positivo o negativo). El valor de referencia puede incluso elevarse a 2,5 o 3,0, depende del tamaño de la muestra. Cuando el signo es positivo, significa que el caso se sitúa a más de dos unidades de desviación típica por encima de la media. Si es negativo, que está a más de dos unidades de desviación típica por debajo de la media.

Cuando la rotación de los ejes factoriales ha sido ortogonal, la nube de puntos suele adoptar una forma circular. Dicha forma se convierte en elipse, si la rotación es oblicua.

La detección concreta de los *atípicos* puede además realizarse mediante el listado de los casos extremos, con puntuaciones factoriales superiores a $\pm 2,0$, e incluso mayor, a semejanza de otros procedimientos analíticos.

EJEMPLO DE DETECCIÓN DE ATÍPICOS MEDIANTE LOS GRÁFICOS DE PUNTUACIONES FACTORIALES

Para ilustrar el uso de los gráficos de puntuaciones factoriales en la identificación de *atípicos*, se han seleccionado los gráficos para los factores obtenidos mediante ACP. Se incluyen dos gráficos que ilustran dos formas distintas. El gráfico para el par integrado por los componentes 3 y 4 (figura B) se ajusta a la forma habitual de una nube de puntos circular. Por el contrario, el gráfico de las puntuaciones para los componentes 1 y 2 (figura A) presenta una nube de puntos más elíptica. En ambos gráficos, todo punto por encima del referente usual de $\pm 2,0$ se considera un posible "atípico". Por encima de $\pm 3,0$, el *atípico* es más evidente.

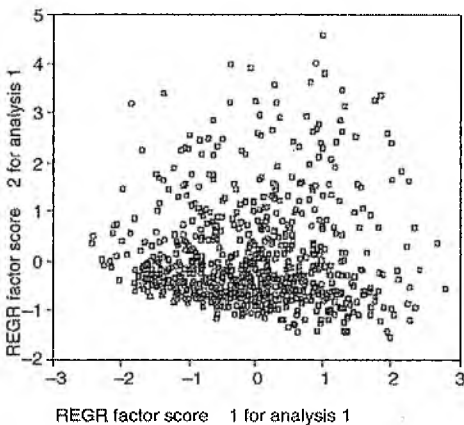


Figura A. Gráfico de puntuaciones factoriales.

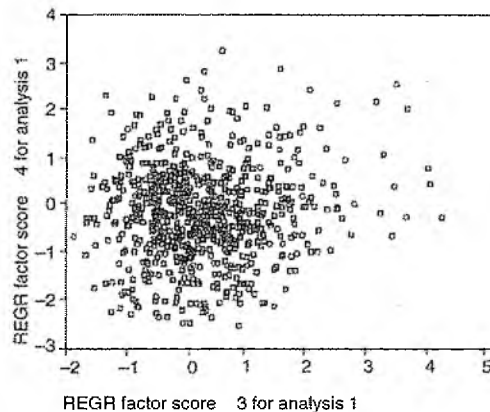


Figura B. Gráfico de puntuaciones factoriales.

LECTURAS COMPLEMENTARIAS

- Batista Foguet, J. M. (1984). "Componentes principales y análisis factorial (exploratorio y confirmatorio)", en Sánchez Carrión, J. J. (ed.): *Introducción a las técnicas de análisis multivariable aplicadas a las ciencias sociales*, Madrid, CIS (Centro de Investigaciones Sociológicas): 23-74.
- Calvo Gómez, F. (1992). "Análisis factorial y las puntuaciones factoriales calculadas por el método selectivo", *Estudios de Deusto*, 40 (1): 71-95.

- Comrey, A. L. (1985). *Manual de análisis factorial*, Madrid, Cátedra.
- Dunteman, G. H. (1989). *Principal components analysis*, Newbury Park, California, Sage.
- Fernández Santana, J. O. (1988). "Comprensión y manejo del análisis factorial", *Revista Internacional de Sociología*, 46 (1): 7-35.
- Gorsuch, R. L. (1983). *Factor analysis*, Hillsdale, N. J., Erlbaum.
- Gorsuch, R. L. (1990). "Common factor analysis versus component analysis: some well and little known facts", *Multivariate Behavioral Research*, 25: 33-39.
- Jolliffe, I. T. (1986). *Principal component analysis*, Nueva York, Springer-Verlag.
- Kim, J. y Mueller, Ch.W. (1978). *Factor analysis: statistical methods and practical issues*, Beverly Hills, Sage
- Velicer, W. F. y Jackson, D. N. (1990). "Component analysis versus common factor analysis: some issues on selecting an appropriate procedure", *Multivariate Behavioral Research*, 25: 1-28.
- Yela, M. (1997). *La técnica del análisis factorial. Un método de investigación en psicología y pedagogía*, Madrid, Biblioteca Nueva.

EJERCICIOS PROPUESTOS

1. En la misma muestra se vuelve a realizar un análisis de componentes principales excluyendo la variable "ideología política". Interpretense los resultados siguientes y compárense con los expuestos a lo largo del capítulo. Además, calcúlense las comunalidades.

Varianza total explicada

Compon.	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones el cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	3,552	27,322	27,322	3,552	27,322	27,322	2,304	17,725	17,725
2	1,659	12,765	40,086	1,659	12,765	40,086	2,150	16,540	34,265
3	1,114	8,570	48,656	1,114	8,570	48,656	1,871	14,391	48,656
4	,957	7,363	56,019						
5	,874	6,724	62,743						
6	,770	5,921	68,664						
7	,749	5,762	74,426						
8	,690	5,304	79,730						
9	,619	4,765	84,494						
10	,581	4,472	88,966						
11	,560	4,304	93,270						
12	,470	3,619	96,890						
13	,404	3,110	100,000						

Método de extracción: Análisis de componentes principales.

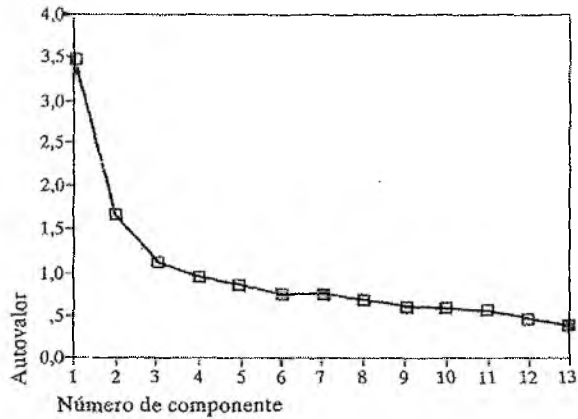


Gráfico de sedimentación

KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin		,821
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	3.159,521
	gl	78
	sig.	,000

Matriz de componentes^a

	Componente		
	1	2	3
casar con marroquí	,640	,341	,442
simpatía marroquí	-,630	-,287	-,162
entrada inmigrante	,620	,121	-,299
vecino marroquí	,592	,354	,438
n.º inmigrantes	,584	-,117	-,243
leyes inmigración	-,583	-2,24E-02	,336
regularizar inmigrante	-,574	-,108	,285
inmigrante delincuente	,570	-7,94E-03	-,238
partido racista	,442	,136	,106
ingresos	-,343	,678	7,543E-03
estudios	-,460	,667	-8,14E-02
edad	,388	-,538	,261
sexo	1,207E-02	,286	-,464

Método de extracción: Análisis de componentes principales.

^a 3 componentes extraídos

Matriz de componentes rotados^a

	<i>Componente</i>		
	1	2	3
entrada inmigrante	,653	,246	-3,53E-02
leyes inmigración	-,648	-,153	9,852E-02
regularizar inmigrante	-,610	-,220	3,429E-02
n.º inmigrantes	,573	,149	-,249
inmigrante delincuente	,567	,194	-,149
casar con marroquí	,152	,833	-6,00E-02
vecino marroquí	,122	,807	-2,99E-02
simpatía marroquí	-,341	-,624	2,459E-02
partido racista	,241	,402	-7,22E-02
estudios	-,223	-2,67E-02	,783
ingresos	-,205	,107	,724
edad	5,228E-02	,157	-,693
sexo	,357	-,157	,381

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización varimax con Kaiser.

^a La rotación ha convergido en 6 iteraciones.

Comunalidades

	<i>Inicial</i>	<i>Extracción</i>
simpatía marroquí	,316	,377
leyes inmigración	,243	,328
sexo	3,078E-02	3,180E-02
edad	,227	,303
n.º inmigrantes	,244	,298
regularizar inmigrante	,232	,309
entrada inmigrante	,286	,395
partido racista	,129	,139
casar con marroquí	,428	,712
estudios	,351	,655
ingresos	,256	,347
vecino marroquí	,368	,465
inmigrante delincuente	,219	,263

Método de extracción: Máxima verosimilitud.

Matriz de configuración^a

	Factor		
	1	2	3
casar con marroquí	,841	8,538E-03	-8,70E-03
vecino marroquí	,649	2,254E-02	-7,52E-02
simpatía marroquí	-,435	-1,37E-02	,283
estudios	5,927E-02	,774	,177
ingresos	,144	,548	,200
edad	8,903E-02	-,515	-5,01E-02
sexo	-3,53E-02	,157	-,120
entrada inmigrante	7,481E-02	2,372E-02	-,596
leyes inmigración	-3,74E-02	3,935E-02	,547
regularizar inmigrante	-8,11E-02	7,662E-03	,514
n.º inmigrantes	3,162E-02	-,151	-,484
inmigrante delincuente	,101	-9,21E-02	-,434
partido racista	,189	-3,27E-02	-,239

Método de extracción: Máxima verosimilitud.
 Metodo de rotación: Normalización oblimin con Kaiser.
^aLa rotación ha convergido en 18 iteraciones.

Matriz de estructura

	Factor		
	1	2	3
casar con marroquí	,844	-,128	-,383
vecino marroquí	,679	-9,42E-02	-,361
simpatía marroquí	-,560	,104	,475
estudios	-,143	,794	,281
ingresos	-3,25E-02	,559	,228
edad	,194	-,538	-,177
sexo	-7,00E-03	,142	-7,74E-02
entrada inmigrante	,337	-8,92E-02	-,625
leyes inmigración	-,288	,138	,570
regularizar inmigrante	-,312	,108	,551
n.º inmigrantes	,272	-,238	-,524
inmigrante delincuente	,310	-,182	-,495
partido racista	,301	-,104	-,329

Método de extracción: Máxima verosimilitud.
 Metodo de rotación: Normalización oblimin con Kaiser.

Prueba de la bondad de ajuste

Chi-cuadrado	gl	Sig.
90,679	42	,000

Varianza total explicada

Factor	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total
1	3,552	27,322	27,322	2,905	22,344	22,344	2,099
2	1,659	12,765	40,086	1,134	8,722	31,066	1,428
3	1,114	8,570	48,656	,587	4,517	35,583	2,319
4	,957	7,363	56,019				
5	,874	6,724	62,743				
6	,770	5,921	68,664				
7	,749	5,762	74,426				
8	,690	5,304	79,730				
9	,619	4,765	84,494				
10	,581	4,472	88,966				
11	,560	4,304	93,270				
12	,470	3,619	96,890				
13	,404	3,110	100,000				

Método de extracción: Máxima verosimilitud.

^a Cuando los factores están correlacionados, no se pueden sumar las sumas de los cuadrados de las saturaciones para obtener una varianza total.

Matriz de correlaciones entre los factores

Factor	1	2	3
1	1,000	-,160	-,446
2	-,160	1,000	,170
3	-,446	,170	1,000

Método de extracción: Máxima verosimilitud.

Método de rotación: Normalización oblimin con Kaiser.

- En la investigación de Vázquez Alonso, Ángel (1991) "Análisis predictivo del rendimiento académico en bachillerato y COU" (*Revista de Educación*, 295: 429-462) se realizó un análisis factorial de componentes principales y de máxima verosimilitud (rotación varimax) con variables que miden el rendimiento de los alumnos: las calificaciones en las asignaturas troncales de cada curso, la inteligencia, las aptitudes académicas (razonamiento abstracto, verbal y numérico). Se quiere descubrir la estructura interna del conjunto de variables que indican la capacidad de los alumnos. La muestra está integrada por 985 alumnos de primer curso, 735 de segundo, 412 de tercero y 409 de COU, pertenecientes a nue-

ve centros públicos de Mallorca (curso 1986/87). Interpretéense los siguientes resultados que corresponden a los alumnos de COU y realícese el gráfico de sedimentación.

Varianza total explicada				Matriz factorial rotada (3 iteraciones)			
Factor	Autovalor	% varianza	% acum.	Variables	Factor 1	Factor 2	Comunalidad
1	7,10180	50,7	50,7	Química	,89740	,20198	,80904
2	2,28210	16,3	67,0	Lengua/literatura	,89357	,10278	,72542
3	,86514	6,2	73,2	Biología	,86052	,22547	,69844
4	,66108	4,7	77,9	Física	,85478	,23176	,75734
5	,57240	4,1	82,0	Geología	,84700	-,21056	,78436
6	,56075	4,0	86,0	Filosofía	,84589	,09945	,84613
7	,50085	3,6	89,6	Matemáticas	,84326	,21507	,79133
8	,34923	2,5	92,1	Dibujo	,80905	,06362	,76174
9	,28636	2,0	94,1	Idioma extranjero	,80846	,21172	,65861
10	,26384	1,9	96,0	Inteligencia general	,06911	,76020	,58268
11	,21674	1,5	97,6	Razonamiento numérico	,15529	,75482	,40260
12	,15524	1,1	98,7	Razonamiento verbal	,17212	,71926	,54696
13	,12277	,9	99,6	Razonamiento abstracto	-,08510	,62877	,59386
14	,06171	,4	100,0	PCOR*	,27574	,59108	,42540

* Puntuación corregida de las pruebas objetivas de Física y Química.

- En el Informe Técnico de la Consejería de Salud y Bienestar Social de la Comunidad de Madrid de 1986, titulado *Zonificación socio-sanitaria, bases para una regionalización de servicios (Mapa de salud y servicios sociales. Comunidad de Madrid)*, bajo la dirección de Alfonso Calvé y Armando Peruga y la coordinación de Benjamín González, se utilizaron las puntuaciones factoriales obtenidas tras la realización de un análisis factorial de componentes principales para realizar tipologías de municipios, zonas básicas y distritos. Complétese la matriz factorial rotada (cálculo de las comunalidades y los autovalores) e intérpretese. Los datos figuran en la primera tabla a continuación (tabla A).
- En el estudio de J. Ignacio Cano, José M. Ruiz Olano y Miguel S. Valles (1988). *El desarrollo social de los pequeños municipios en la Comunidad de Madrid*, Asamblea de Madrid, colección *Estudios Parlamentarios* se comprueba la adecuación de indicadores clásicos en la medición de distintas dimensiones (demográfica, de accesibilidad y socioeconómica) del concepto "desarrollo social" acudiendo a distintas técnicas multivariantes: análisis factorial, de conglomerados y discriminante. Analícese y complétese la matriz factorial (ACP; rotación *varimax*) para el conjunto de los indicadores elegidos (tabla B, página siguiente).

Tabla A (3.º ejercicio)

<i>Variables</i>	<i>Factor 1</i>	<i>Factor 2</i>	<i>Factor 3</i>	<i>Factor 4</i>	<i>Factor 5</i>
Personas por familia	,882	,038	-,273	-,006	,003
Tasa de actividad	,059	,085	,870	,212	-,037
Desocupación	-,072	-,137	-,093	-,004	,885
Tasa participación política	-,334	-,318	-,199	-,346	-,653
Crecimiento 75/81	,274	,834	-,011	,096	-,122
Número de bancos	-,744	-,062	-,445	,030	-,300
Comercios alimentación	-,073	,016	-,564	-,295	-,489
Población 0-14 años	,791	,563	,077	-,023	,022
Población 25-64 años	-,147	-,162	,643	,383	-,161
Fecundidad	,394	,797	,000	-,130	,066
Óptimo fecundo	-,247	,794	,173	-,170	,072
Índice envejecimiento	-,886	-,257	-,174	,043	-,012
Tasa de dependencia	,164	,655	-,316	-,123	-,065
Pob. con bachiller. Sup.	-,146	-,273	,295	,818	,081
Pobl. analfabeta	-,015	-,173	-,050	-,927	,033
Log Pob*	,140	,005	,547	,538	,496
TRF. Secundario	-,256	-,120	-,692	-,035	-,563
TRF. Agrícola	-,268	-,063	-,525	-,615	-,389
TRF. Servicios	-,122	-,346	,346	,761	,273
Log densidad*	,345	,575	,338	,263	,183

* Transformadas tomando el logaritmo de los valores de cada variable.

Tabla B (4.º ejercicio)

<i>Variable</i>	<i>Factor 1</i>	<i>Factor 2</i>	<i>Factor 3</i>	<i>Factor 4</i>	<i>Factor 5</i>	<i>Factor 6</i>
Distancia a Madrid	-,87081	,29295	-,03455	,01409	-,04419	-,04707
Dist. carretera nacional	-,36347	-,02524	-,29794	,08209	-,55653	,33518
Dist. núcleo > 20.000hab.	-,80487	,38022	-,01885	-,01730	-,01614	-,16850
Tasa de jóvenes	,86281	-,10459	-,04098	-,18048	-,09926	-,03697
Tasa de viejos	-,85242	,14609	,08853	-,26919	-,04620	,01540
Tasa de paro	,01964	-,45701	,10234	,02682	,29928	,28748
Dist. centros educativos	-,80565	,26735	,04882	-,02800	,31650	-,22629
Servicios	,57049	,59034	-,08715	,08344	,05014	-,13339
Agricultura	-,77789	,02683	,12958	-,14953	-,03862	,25476
Estudios	,57727	,54517	-,04569	,03318	,10464	,19163
Viviendas secundarias	-,35000	,58294	-,23530	,25980	-,25762	-,17301
Razón de masculinidad	-,31261	,04129	-,07289	,48860	,28554	,26789
Dist. a la farmacia	-,62409	,29992	,12392	-,03574	,35634	,17426
Transporte público	,55759	,23299	,67680	-,05057	,03849	,01741
Razón actividad-dependencia	,13319	-,08035	-,08598	,82106	,27727	,03795
Incremento pobl. 1970-1975	,76002	,20943	-,17123	-,08430	-,10515	,15984
Incremento pobl. 1975-1981	,65282	,37877	-,24803	-,18131	,25365	,23088
Incremento pobl. 1970-1981	,76985	,37868	-,23504	-,18463	,14445	,25509
Industria-construcción	,38214	-,68767	-,07260	,10269	-,00508	-,18467
Teléfonos por habitante	,45465	,25233	,08474	,38809	-,39308	-,21708
Déficit educativo	,29174	-,01323	-,35065	-,17044	,34083	-,60121
Transporte público por pobl.	,43396	,22883	,78844	,10887	-,06900	-,09462

5. Otro ejemplo de la utilización del análisis factorial en la tipologización de la estructura espacial es la clasificación de distintos barrios de Vitoria-Gasteiz realizada por Félix Calvo Gómez y Cristina Lavia Martínez, publicada en 1993, en el artículo "El método selectivo factorial" en el análisis de tipologías urbanas" (*Estudios de Deusto*, 41 (1): 99-121). En él se aplica una alternativa (el "método selectivo") al cálculo tradicional de las puntuaciones factoriales, que consiste en calcularlas sólo en las variables que saturan significativamente en cada factor. En los análisis preliminares se realizó un ACP con rotación *varimax* para las 22 variables que mostraron ser más relevantes en la clasificación de los barrios. Interprete la matriz factorial y las puntuaciones factoriales promedio.

Matriz factorial rotada

Variables	Factor 1	Factor 2	Factor 3	Factor 4
Tasa juventud	,227	-,252	,878	-,017
Tasa vejez	-,204	,018	-,902	-,001
Tasa analfabetismo	,317	,141	-,206	,750
% pobl. estudios primarios	,621	,378	,309	,350
% pobl. estudios superiores	-,782	-,463	-,242	-,121
% emigrantes	,568	,190	,532	,437
% emigr. llegados últimos 5 años	-,344	-,774	-,133	-,016
% emigr. llegados entre 5-10 años	-,108	-,751	,366	-,172
% emigr. llegados hace más de 10 años	,253	,901	-,170	,121
Tamaño medio de la vivienda	-,826	-,411	,008	-,196
% ocupados industria	,637	,472	,523	,054
% ocupados construcción	,421	,283	,181	,627
% ocupados servicios	-,652	-,472	-,511	-,190
Tasa de vascoparlantes	-,677	,090	-,140	-,498
Tasa de paro	,183	,481	-,341	,686
% núcleo familiares con hijos	,052	,136	,954	-,054
% núcleos familias numerosas	-,590	,405	,316	,507
% mujeres ocupadas	-,269	-,739	-,201	-,367
% mujeres amas de casa	,326	,723	,210	,359
% ocupados clase alta	-,832	-,421	-,201	-,202
% ocupados clase media	-,058	-,067	-,039	-,586
% ocupados clase baja	,757	,405	,189	,379
% varianza explicada	25,85	22,06	18,73	14,43

Puntuaciones factoriales promedio para algunos barrios (Vitoria-Gasteiz, 1989)

<i>Barrios</i>	<i>Factor 1</i>	<i>Factor 2</i>	<i>Factor 3</i>	<i>Factor 4</i>
Casco viejo	,782	-,452	-1,691	,910
Ensanche	-1,630	-,794	-1,147	-,349
Lovaina	-,990	-,468	-,963	-,155
Coronación	-,244	,776	-,384	-,341
El Pilar	,574	,465	,821	-,368
Gazaibide	-1,609	,540	1,318	-,502
Txagorritxu	-,308	,112	,148	-,712
San Martín	-,612	-2,133	,406	-,536
Zaramaga	,121	1,058	,481	,513
El Anglo	-,158	,791	-,388	-1,478
Sierras	,709	-,456	,810	-,829
Santiago	,330	-,560	,267	-1,162
Arambizkarra	,824	-,718	1,112	-,392
Arana	,213	,495	,441	,335
Desamparadas	-,685	-,327	-,684	-,595
Judizmendi	,232	,427	-,626	-,220
Santa Lucía	1,016	-1,390	,531	-,327
Errekaleor	,645	-,248	-,715	3,425
Adurza	,774	,241	,336	,272
San Cristóbal	,408	,266	,136	-,214