Capítulo 5

El Análisis Factorial y la construcción de indicadores e índices sociales

Las técnicas de análisis factorial se han consolidado en la investigación social. En concreto, su aplicación en la determinación de indicadores e índices cuenta ya con una extensa tradición que se concreta en:

- La aplicación de esta técnica nos permite identificar las dimensiones y/o ***indicadores*** más significativos del concepto sometido a examen. El elemento identificativo de esta técnica es su capacidad en sintetizar información, lo que consigue eliminando del conjunto de variables iniciales aquellas que ofrecen información redundante junto con las que no se adaptan al modelo de regresión múltiple, desde el que se basa esta técnica. Los estudios de ecología factorial explican la diferenciación social a partir de tres factores, dimensiones o indicadores, a saber: estatus o rango social, estatus familiar y estatus étnico[1].

- Su escasa repercusión en la construcción de ***índices*** (quizás por el excesivo protagonismo que ésta técnica adquirió en los estudios mencionados) no ha impedido que la técnica factorial siga considerándose una herramienta de incuestionable valor para tal fin. En este caso, la adición de las puntuaciones factoriales obtenidas nos da el índice del concepto analizado.

Estas aplicaciones, justifican el hecho de que dediquemos una apartado *ex profeso* al análisis factorial como técnica aplicada en la construcción de indicadores e índices, propósito perseguido en esta investigación.

El análisis factorial es una de las técnicas de análisis multivariable[2] más utilizada en la investigación en Ciencias Sociales. El análisis multivariable parte del principio de ***causalidad múltiple*** (Bisquerra, R., 1989: 2), a partir del cual, los hechos y fenómenos sociales son explicados no sólo por una única causa sino por una gran variedad de ellas.

En este contexto, el objetivo general del análisis multivariable no es otro que el de analizar simultáneamente un importante número de variables observables de una

1. Una exposición de esta técnica en los estudios de ecología factorial se encuentra en el Capítulo 6, Indicadores sociales y análisis urbano: el caso español, de esta Tesis.
2. En la bibliografía especializada, el término multivariate analysis ha sido traducido de distintas formas, a saber: análisis multivariable, análisis multivariante y análisis multivariado. No obstante, y pese a esta triple denominación, el significado y contenido es compartido.

muestra de individuos (Bisquerra, R., 1989: 4 y 35)[3]. En particular, el análisis factorial es una técnica multivariable que tiene por cometido reducir un conjunto de *v* variables aleatorias (interrelacionadas), en un grupo de *f* factores latentes (independientes), de tal manera que *f* factores siempre serán, en número, inferior a las *v* variables iniciales. Los factores reflejan la síntesis de la información redundante de las variables (Ferrán, M., 1996: 421). En última instancia, el éxito de esta técnica viene dado en la medida que su resolución cumpla dos requisitos básicos (Bisquerra, R., 1989: 288):

- El ***principio de parsimonia,*** común a toda la teoría científica, establece que "todo modelo debe ser más simple que los datos en los que se basa" (García Ferrando, M., 1989: 434).

- El número de factores elegidos deben ser ***interpretables***.

En este caso, y si cada una de las variables es susceptible de ser expresada a partir de una serie de ***factores latentes*** desconocidos, estos factores, a su vez, pueden ser considerados como indicadores nuevos, ***indicadores sintéticos***, los cuales resumen el conjunto de información proporcionada por los indicadores originales[4].

1.- Origen y finalidad del Análisis Factorial.

El antecedente del análisis factorial lo encontramos en las *técnicas de regresión* planteadas inicialmente por **Galton**. En 1901 **Pearson**, discípulo del primero, expone el *Método de componentes principales*, previo paso al cálculo del análisis factorial (Bisquerra, R., 1989: 289).

3. La mayoría de análisis multivariables se resuelven a través de cálculo matricial, de ahí que el punto de partida del análisis multivariable sea la matriz de datos originales, en la que las filas representan a los individuos y las columnas a las variables. El valor de la observación del individuo i en la variable j se representa por X_{ij} (Bisquerra, R., 1989, Op. cit., pp. 35-39).

4. Es difícil encontrar en la literatura estadística, e incluso en las investigaciones aplicadas consultadas, la utilización explícita del análisis factorial como técnica para la obtención de indicadores sintéticos, y todo ello pese al carácter "sintético" que caracteriza a esta técnica. El INE expresa en los siguientes términos el objetivo perseguido en el monográfico Disparidades Provinciales, incluido en su publicación de 1997: "El objetivo final del presente análisis es resumir toda la información (el informe recoge 395 indicadores relacionados con distintos campos sociales de interés, no obstante, las páginas que alude la cita, hacen referencia exclusivamente a los indicadores del campo "trabajo", pues lo que se pretende es dejar constancia, a modo ilustrativo, de esta técnica) mediante la obtención de un pequeño grupo de indicadores sintéticos que reflejen de manera sencilla y evidente las disparidades territoriales (...)". (INE, 1997, Op. cit., pp. 607) (la información entre el paréntesis y las cursivas son nuestras).

No obstante, el origen de esta técnica se atribuye a **Spearman** quien en 1904 daría a conocer su trabajo sobre la inteligencia (Bisquerra, R., 1989: 291). La teoría clásica de la inteligencia establece que sus diferentes manifestaciones responden a un factor general, el factor “G” de Spearman (Cuadras, C., 1991: 864). De este modo, el análisis factorial se vincula originariamente con las aplicaciones en el campo de la Psicología proporcionando un excelente modelo explicativo matemático a las teorías de capacidad y comportamiento.

Si Spearman es considerado el padre de la técnica, el que la popularizaría sería **Thurstone** (García Ferrando, M., 1989: 434) quien, en la década de los cuarenta, aplicaría el análisis factorial para identificar y diferenciar los principales factores que intervienen en la inteligencia humana (capacidad verbal, espacial y cuantitativa). A Thurstone debemos, principalmente, la relación entre las *correlaciones* y las *saturaciones* de las variables en los factores, así como el desarrollo teórico y metodológico de las rotaciones factoriales (Bisquerra, R., 1989: 291). Estas rotaciones, solución que facilita la interpretación de los factores, inicialmente fueron gráficas. Con posterioridad, **Kauser** (1958) lo transcribiría a un modelo matemático, el modelo de rotación Varimax, al que con posterioridad le sucederían otros (Bisquerra, R., 1989: 291).

Vemos, pues, como han sido los distintos autores que han participado de alguna u otra forma, los que han “modelado” y determinado lo que hoy conocemos y aplicamos como técnicas factoriales. En esta trayectoria evolutiva, y en función del objetivo perseguido por los autores, podemos diferenciar dos tendencias: el análisis factorial exploratorio y el análisis factorial confirmatorio. El ***análisis factorial exploratorio*** se inicia con Spearman (1904), continua con Thurstone (1947) y culmina con Harman (1989) y tiene por objeto “explorar la dimensionalidad latente sobre un conjunto de variables expresadas a través de sus factores comunes, cuya estructura debe ser lo mas simple posible”. Por su parte, el ***análisis factorial confirmatorio***, abanderado por Jöreshog (1983) y Sörbom (1985) “(...) sea por razones de tradición científica, sea por comparación con otros análisis previos, sea por sentido común, se suele realizar con un conocimiento previo de la estructura de los factores” (Bisquerra, R., 1989: 334; Cuadras, C., 1991: 887).

El desarrollo de las técnicas multivariables en general, y en particular el análisis factorial, precisan de un importante número de operaciones, y máxime cuando en el análisis interviene un número importante de variables. Con la llegada y difusión de los ordenadores no solo se han superado los inconvenientes ligados a la laboriosidad de éstas técnicas, sino que se han convertido en una de las más utilizadas en la investigación en las Ciencias Sociales (Bisquerra, R., 1989: 291)[5].

En definitiva, las técnicas de análisis factorial se han consolidado, difundido y aplicado en un importante número de investigaciones en virtud a su capacidad reductora, esto es, por ofrecer la posibilidad de acercase, simplificadamente, a la realidad compleja.

En las páginas que siguen, el análisis factorial se presenta, y posteriormente se aplicará, como una técnica que "consiste en resumir la información obtenida de una matriz de datos con *v* variables. Para ello, se identifican un número menor de factores *F*, siendo $F < v$. Los factores representan a las variables originales con una pérdida mínima de información" (Bisquerra, R., 1989: 287).

2.- El modelo teórico del Análisis Factorial

Como ya hemos expuesto, el objetivo del análisis factorial no es otro que el de extraer los factores latentes, no observables directamente, de un conjunto de variables sobre un número de individuos, o lo que es lo mismo, la obtención de un conjunto de factores que expliquen la *covariación* entre dichas variables. En consecuencia, el modelo matemático del análisis factorial se asemeja al ***modelo de regresión múltiple***[6] (Bisquerra, R., 1989: 287). En este caso, y a diferencia de la ecuación del modelo de regresión, los factores no son variables simples, sino dimensiones que engloban a un conjunto determinado de variables, pudiendo ser explicadas las variables (linealmente) en función de los factores seleccionados.

En el modelo del análisis factorial, si los factores son inferidos a partir de las variables observadas, cada variable será expresada como una combinación lineal de factores no observables directamente. Se admite, pues, que "un conjunto de variables aleatorias, X_1, X_2,...X_n, se explicarán por un conjunto de factores comunes, F_1...F_m (siendo $f < v$) y n factores únicos, U_1...U_n" de acuerdo con el siguiente modelo factorial lineal (Cuadras, C., 1991: 867-868):

5. Bisquerra Alzina introduce el término estadística informática. Con este neologismo el autor se refiere "al estudio de la estadística aplicada a la investigación empírica mediante procedimientos informáticos. (...) trata de los análisis más adecuados para cada investigación, sin preocuparse demasiado del proceso mecánico de cálculo ni de la demostración de las fórmulas matemáticas. Estos conocimientos son la confluencia de la estadística aplicada y de la informática" (Bisquerra, R., 1989, Op. cit., pp. 25-16).

6. "La regresión múltiple estudia las relaciones entre una variable dependiente y un conjunto de variables independientes" (Bisquerra, R., 1989, Op. cit., pp. 197). Regresión y correlación son dos aspectos del mismo análisis pues la regresión múltiple permite la correlación múltiple (mide la intensidad de la relación entre un conjunto de variables independientes y una dependiente).

$$X_1 = a_{11}F_1 + \ldots + a_{1m}F_m + d_1U1$$
$$X_2 = a_{21}F_1 + \ldots + a_{2m}F_m + d_2U2$$
$$X_n = a_{n1}F_1 + \ldots + a_{nm}F_m + d_nUn$$

en donde:
a_{11} ..... a_{1m} pesos factoriales de los Factores Comunes[7];
F_1 ..... F_m m Factores Comunes, tales que f < v[8];
d_n es el *peso* factorial del Factor Unico;
U_1 es el *factor único*[9].

En el modelo factorial se supone que: (1) los *factores comunes* y los *únicos* están mutuamente incorrelacionados (son factores independientes para los que no se da la interdependencia lineal); y, (2) los *m* factores son en número menor que las *n* variables, puesto que se desea explicar las variables por un número más reducido de variables o factores (Cuadras, C., 1991: 868-869).

Los principios fundamentales del modelo factorial lineal vienen dados por dos teoremas (Cuadras, C., 1991a: 93-96):

1er Teorema: la varianza total de una variable (S^2_i) puede ser explicada por la suma de tres tipos de varianzas independientes entre sí:

- Varianza total explicada por los factores de 1 a k,

$$(S^2_1, S^2_2, S^2_3, \ldots, S^2_k)$$

- Varianza de la unicidad (S^2_u), varianza no explicada por los factores comunes y que responde a la varianza específica propia de la variable y a la varianza de error debida a errores aleatorios.

7. Este coeficiente indica el grado de relación, de peso y/o de saturación de cada factor común sobre la variable.
8. Los factores comunes debemos entenderlos como "la dimensionalidad influyente que relaciona y explica las relaciones y asociaciones existentes entre las variables" (Cuadras, C., 1991a, Op. cit., pp.86). Explican las correlaciones, variabilidad, entre las variables.
9. Los factores únicos explican, por sí mismos, una parte de la variabilidad de cada variable, sin relación con los demás ni con los factores comunes. Explican la varianza de la variable Xi no explicada por el factor común e influyen, exclusivamente, en la variable Xi (i = 1,..... n).

La notación matemática de la varianza total de una variable es:

$$S^2_i = S^2_1 + S^2_2 + S^2_3, \dots, S^2_k + S^2_u$$

y dividido por S^2_i:

$$1 = a^2_{i1} + a^2_{i2} + \dots + a^2_{ik} + S^2_{iu}$$

en donde:

- $a^2_{i1} + a^2_{i2} + \dots + a^2_{ik}$ es la proporción de varianza total de la variable explicada por los factores 1 a k y,
- S^2_{iu} la proporción de la varianza total de la variable explicada por la unicidad.

Lo que se anota como:

$$1 = h^2_i + S^2_{iu}$$

en donde:

- $h^2_i = a^2_{i1} + a^2_{i2} + \dots + a^2_{ik}$ es la comunalidad o parte de la varianza total de la variable que se relaciona con el resto de variables,
- S^2_{iu} es la unicidad o varianza total explicada no por los factores comunes, sino por la especificidad de la variable (E^2_i) y la varianza del error (e^2_i).

En síntesis, el primer teorema demuestra que:

$$1 = h^2_i + E^2_i + e^2_i$$

2º Teorema: la proporción de la varianza total de una variable explicada por cada uno de los factores comunes puede ser expresada como un coeficiente de determinación (r^2). La raíz cuadrada de esta proporción nos da un coeficiente de correlación (saturación) entre el factor y la variable.

La ecuación anterior quedaría:

$$1 = a^2_{i1} + a^2_{i2} + \dots + a^2_{ik} + E^2_i + e^2_i$$

en donde:

- $a^2_{i1} + a^2_{i2} + \dots + a^2_{ik}$ son los coeficientes de correlación de la variable i con los factores 1, 2, 3,... k;

De lo anterior, se deduce que la correlación entre dos variables es igual a la suma de los productos de sus números pesos o saturaciones para cada uno de los factores, que anotado matemáticamente queda:

$$r_{xy} = r_{x1}r_{y1} + r_{x2}r_{y2} + \ldots + r_{xk}r_{yk}$$

y que de forma generalizada se anota:

$$r_{xy} = \Sigma a_{xy}a_{yk}$$

así las correlaciones estimadas entre los factores y las variables pueden ser utilizadas como estimación de las correlaciones entre las variables.

3.- Análisis Factorial de Componentes Pricipales.

En la actualidad, el desarrollo matemático e informático alcanzado por los modelos de análisis multivariable se encuentra muy desarrollado y consolidado, ofreciendo una gran riqueza analítica. No obstante, la validez de todo el proceso metodológico asociado al análisis factorial (que a continuación desarrollamos) solo se garantiza de seguir dos pasos previos (Bisquerra, R., 1989: 4):

- **Análisis exploratorio de los datos**. Desarrollado inicialmente por Tukey, se conoce hoy por las siglas ***EDA*** (*Exploratory Data Analysis*). Incluye las medidas de la estadística descriptiva univariable (centralidad, variabilidad, asimetría, comprobación de los supuestos paramétricos,...) y tiene por finalidad comprender, de forma individual, cada una de las variables. Las representaciones gráficas "tallos y hojas" y la "caja y bigotes" son muy aplicadas en esta primera fase[10].

- **Análisis bivariable.** Una vez analizada la idoneidad de las variables y como paso previo a la mayoría de los análisis multivariables, resulta de gran utilidad identificar y conocer como se relacionan pares de variables. La ***matriz de correlaciones*** (en la que se muestra la interdependencia lineal entre las variables) a través del ***coeficiente de correlación de Pearson***, será la que determine el grado de relación lineal entre dos variables aleatorias, a la vez que ofrece información sobre la variabilidad de una variable que queda expli-

10. Al análisis de la matriz de correlaciones le precede, metodológicamente, el análisis exploratorio de las variables que formarán parte de esa matriz (con demasiada frecuencia este análisis es omitido). Este análisis exploratorio tiene como finalidad examinar las características principales de las variables para asegurarnos si cumplen, o no, el requerimiento teórico (la distribución de las variables debe aproximarse a la normalidad) para que los coeficientes de correlación lineal utilizados midan adecuadamente la relación existente entre las variables. Para una exposición informática sobre la depuración de la matriz de datos ver Bisquerra, R., 1987, Op. cit., pp.37-77.

cada por la otra variable (Cuadras, C., 1991: 862).

La Tabla 5.1 recoge la secuencia metodológica a llevar a cabo cuando el análisis factorial sea la técnica seleccionada. Junto a los pasos, el cuadro incluye información adicional que es explicada en las páginas que le siguen (Bisquerra, R., 1989: 295-299, 300 y 316).

Tabla 5.1:
Secuencia Metodológica en el Análisis Factorial

	Paso	Objetivo	Indicadores/métodos
Indispensables	1.- Matriz de correlación	Examinar la correlación y la asociación lineal entre las variables	I.- El determinante; II.- Test de esfericidad de Barlett; III.- Indice de KMO de Kaiser-Meyer-Olkin; IV.- Correlación anti-imagen; V.- Medida de adecuación de la muestra (MSA); VI.- Coeficiente de Correlación Múltiple
Indispensables	2.- Extracción de factores	Resumir el conjunto de variables en un subconjunto de factores, de tal manera que aún siendo en número menor, ofrezcan la misma información	I.- Método centroide; II.- Componentes Principales; III.- Factor Principal; IV.- Factorización de ejes principales; V.- Factorización alpha; VI.- Factorización de imagen; VII.- Máxima verosimilitud; VIII.- Mínimos cuadrados no ponderados; IX.- Mínimos cuadrados generalizados; X.- Mínres; XI.- Segunda generación de Kaiser
Simplificación de la interpretación factorial	3.- Rotación de Factores	Pretende seleccionar la solución factorial más sencilla e interpretable	I.- Ortogonales: I.I.- Quartimax; I.II.- Ecuamax; I.III.- Ortomax; I.IV.- Varimax; II.- Oblicuas: II.I.- Cuartimín; II.II.- Oblimax; II.III.- Oblimin; II.IV.- Ortoblícua
Aplicaciones posteriores	4.- Puntuaciones factoriales	Permiten determinar en qué medida los factores seleccionados se dan en los individuos o en otras unidades de análisis	

Fuente: Elaboración propia.

a. ***Matriz de correlaciones***.

El primer paso del análisis factorial es el análisis de la ***matriz de correlación*** (*correlation matrix*) (Bisquerra, R., 1989: 283). La matriz de correlación (representada por *R*) se elabora a partir de la *matriz de datos originales* y en una primera aproximación, la podemos identificar con una matriz de "similitudes" o "proximidades", ya que cuanto más elevados sean los coeficientes de correlación mayor será la relación entre las variables (Bisquerra, R., 1989: 37; García Ferrando, M., 1989: 432).

Según el modelo expuesto, el análisis factorial es una técnica que analiza la correlación lineal entre las variables. Si las variables no estuvieran asociadas linealmente las correlaciones entre ellas serían nulas, y por ello, no existiría asociación lineal entre las variables por lo que sería incorrecto, y vacío de contenido, someter las variables a un análisis factorial (Ferrán, M., 1996: 423).

Para medir el grado de asociación entre las variables existe un número importante de coeficientes estadísticos (recogidos en Tabla 5.1). De todos ellos, es el ***coeficiente de correlación múltiple*** el más conocido. Este indicador, mide el grado de intercorrelación o de asociación lineal entre las variables, de tal manera que cuando éstos son bajos, las variables podrán ser eliminadas; y cuando sean altos, la matriz puede ser considerada adecuada para un análisis del tipo factorial (Bisquerra, R., 1989: 230). No obstante, sucede que en no todas las ocasiones una correlación baja, o viceversa, es sinónimo de inexistencia de factores compartidos. Por ello, para garantizar que los datos se ajustan, o no, a un modelo de análisis factorial es interesante someterlos a otros tests como el ***determinante*** de la matriz de correlaciones, el de esfericidad de ***Barlett*** o, el de ***Kaiser-Meyer-Olkin***. Si los test resultaran significativos, el análisis factorial se evidencia como una técnica idónea para interpretar la información contenida en esta matriz (Ferrán, M., 1996: 424).

Además de los test de lineabilidad expuestos, otras condiciones a considerar y que nos garantizan la idoneidad del análisis factorial, son: el análisis no podrá contener variables de tipo cualitativo, pues el modelo parte de la supuesta combinación lineal de los factores como expresión de las variables; y, las variables deben tener unidad experimental, pues de lo contrario extraer factores carece de sentido (Cuadras, C., 1991a: 102-103).

b. ***Extracción de Factores***.

El análisis factorial no es un concepto unitario en tanto que engloba una gran variedad de técnicas que siguiendo distintos procedimientos (aunque muy próximos) y con sus ventajas e inconvenientes, tienen la finalidad de extraer los factores subyacentes a un conjunto inicial de variables.

Las distintas técnicas también se exponen en Tabla 5.1 que precede a estas líneas, siendo las técnicas de Componentes Principales, Máxima Verosimilitud y Mínimos Cuadrados, las más utilizadas. Aquí exponemos el análisis de Componentes Principales pues los supuestos sobre los que se basa[11] aseguran la posibilidad de utilizar (como es nuestro propósito) con posterioridad a este análisis, otro tipo de técnicas.

El análisis de ***Componentes Principales*** (introducido por Hotelling en 1932) se define como "una técnica estadística que permite transformar un conjunto de variables, intercorrelacionada, en otros conjuntos de variables no correlacionadas denominados factores" (Bisquerra, R., 1989: 301). El objetivo de la técnica es explicar la mayor cantidad de la varianza de las variables originales a través del menor número de factores o componentes.

La base del análisis de Componentes Principales es la matriz de correlación a partir de la cual se obtendrán las ecuaciones lineales que representan la transformación lineal de las variables originales en relación con los componentes resultantes. Puesto que el *criterio de extracción* de los factores es el de Componentes Principales, el *primer factor principal (F1)*, será aquella combinación que "explica la mayor parte de la variabilidad, varianza, de las variables" (Ferrán, M., 1996: 425). Obtenido éste, se le resta a las variables, y sobre la variabilidad restante se elige el *segundo factor principal (F2)* aquel que, incorrelacionado con el *F1*, explica el máximo de variabilidad, y así sucesivamente (Cuadras, C., 1991a: 117). En conclusión, "los *m* factores principales son tales que la variabilidad explicada por cada uno de ellos es máxima" (Cuadras, C., 1991: 878).

El paquete estadístico utilizado para tal fin, nos dará tantos componentes principales como variables intervienen, con lo cual no obtendremos una reducción de la información, objetivo último de esta técnica. Al conservar todos los componentes principales o factores para explicar cada una de las variables, la proporción de varianza de la variable explicada por los factores (***comunalidad***), será igual a 1 para todas las variables.

La extracción de estos factores se apoya en el indicador que recoge los valores propios o ***eigenvalues*** de cada variable y puede ser interpretado como la variabili-

11. El análisis de Componentes Principales prescinde de las unicidades para expresar las n variables. Utiliza, por lo tanto, únicamente los factores comunes quedando el modelo factorial:

$$X_i = a_{i1}F_1 + \ldots + a_{in}F_n \quad (i = 1, \ldots n)$$

No exige estimar las comunalidades pues todas valen 1.

dad total explicada por el factor (Bisquerra, R., 1989: 304)[12]. La varianza total es la suma de la varianza de cada una de las variables y que al trabajar con variables estandarizadas, siendo la varianza de éstas 1, el total de la varianza coincidirá con el número de variables que participen en el análisis. La situación ideal se produce cuando los autovalores correspondientes a los primeros factores (esperando que éstos sean mínimos) fueran elevados, pues significaría que entre las variables hay fuertes correlaciones (Ferrán, M., 1996: 426).

Uno de los problemas que el análisis de Componentes Principales deberá resolver es el de la elección del número de factores. Esta es una decisión que adoptará el propio investigador y para la cual cuenta con algunas reglas o criterios:

- Los factores se disponen de mayor a menor ya en las estadísticas iniciales, de tal manera que aquellos que explican la mayor cantidad de varianza total, ocuparán los primeros lugares en las *estadísticas iniciales* y que, en general, suele coincidir con los tres primeros siendo éstos los que explican el mayor porcentaje (PC of VAR) de varianza total.

- Un segundo criterio, y uno de los más utilizados, es la ***regla de Kaiser***. La regla de Kaiser suele ser el criterio que por defecto cuentan los paquetes informáticos y consiste en seleccionar aquellos factores para los que sus *eigenvalues* superan la unidad.

- Un tercer criterio, es el análisis del gráfico ***scree plot*** (o perfil de la falda de una montaña) en donde los factores se sitúan en el eje abcisa y los valores propios en el eje de ordenadas. En este gráfico aparecen diferenciados los factores con valores bajos de los que tienen asociados valores altos, de tal manera que los factores situados por debajo de este punto de inflexión de la gráfica, serán descartados (coincidiendo con valores *eingenvalues* inferiores a la unidad) (Bisquerra, R., 1989: 307-308).

Una vez obtenidos los factores, cada una de las variables podrá ser expresada como combinación lineal de los mismos, lo que queda reflejado en la ***matriz factorial*** (*factor matrix*). La matriz factorial no es más que una reproducción sencilla de la matriz de correlaciones inicial en donde cada columna es un factor (factores se-

12. Los autovalores o (eigenvalues) pueden transformarse en un porcentaje de varianza total de la muestra (PCT of VAR) explicada por los factores, lo que se conseguiría a partir de una simple regla de tres. Por su parte, el cálculo de los porcentajes acumulados (CUM PCT), implica la suma "acumulada" de esos porcentajes. Los autovalores, el PCT of VAR y el CUM PCT aparen junto a cada valor, en las estadísticas iniciales (initial statistics) que siguen a la matriz de correlación (correlation matrix) en toda aplicación de esta técnica. Las estadísticas iniciales proporcionan toda la información relativa al conjunto de variables y factores extraídos y que inicialmente coincidirán en número.

leccionados), las filas son las variables y los *Fij* son considerados como índices de correlación (cuando los factores son ortogonales, esto es, cuando los factores no están correlacionados) entre el factor *i* y la variable *j*. Estos coeficientes, en el análisis factorial, reciben el nombre de pesos, cargas, ponderaciones o ***saturaciones factoriales*** (*factor loading*), e indican el peso que cada variable asigna a cada factor. Cuando las saturaciones (en valores absolutos) de las variables son altas, esta variable se asocia con el factor. El análisis factorial, en consecuencia, adquiere protagonismo cuando todas las variables están saturadas en alguno de los factores (Bisquerra, R., 1989: 302-304).

El índice *Fij* de la matriz factorial, desagrega los valores propios que aparecían en las estadísticas iniciales. Así, mientras que en las estadísticas iniciales (parte derecha de la misma) aparecía el total de la varianza explicada por cada factor (cuya suma es el número de variables que intervienen), en la matriz factorial se recoge, para cada variable, la varianza atribuida a cada factor. La suma de estas varianzas, en cada factor o columna, debe coincidir con el *eigenvalue* de las estadísticas iniciales.

Por último, las *estadísticas finales* (*final statistics*), y de forma sintetizada, recogen, de un lado, los tres (normalmente) factores principales pues son ellos los que explican el mayor porcentaje de variabilidad total asociando, a cada uno de ellos, su autovalor (porcentaje de la varianza total explicada por cada factor) así como el porcentaje de varianza explicada por cada uno de ellos y el porcentaje acumulado. En segundo lugar, esta estadística, señala la ***comunalidad*** (*communality*) de cada variable y/o proporción de varianza explicada por el conjunto de factores comunes resultantes (Bisquerra, R., 1989: 304). La varianza total de cada una de las variables no es del todo explicada pues, recordemos, a partir de la matriz factorial, solo se recogen unos factores y no todos. Las comunalidades (h^2) se calculan a partir de la matriz factorial y resulta del sumatorio al cuadrado de las ponderaciones factoriales de cada variable, lo que matemáticamente se anota como (Cuadras, C., 1991a: 92):

$$h^2_i = a^2_{il} + \ldots + a^2_{im} \quad i = 1, \ldots, m$$

La comunalidad es un valor que oscila entre 0 y 1. Cuando se aproxima a 1 indica que la variable queda totalmente explicada por los factores, mientras que si se aproxima a 0, los factores no explicarán nada la variabilidad de las variables. La varianza no explicada por los factores comunes se atribuye al ***factor único*** (*U*). La varianza total explicada (1) resulta de sumar las comunalidades más el factor único, expresándolo matemáticamente (Bisquerra, R., 1989: 306):

$$1 = h^2_i + U^2_i$$

Nótese que en la matriz factorial solo aparecen los factores comunes, por lo que el factor único deberemos deducirlo.

Por último, y como indicador de que el modelo al que hemos llegado se ajusta a los datos, el análisis factorial de Componentes Principales concluye calculando la ***matriz de residuales***. Esta matriz aparece dividida en dos partes: en el triángulo superior, se recogen las interrelaciones estimadas entre las variables estimadas a partir del modelo (*reproduced correlation matrix*); y en el triángulo inferior, los ***residuales*** (diferencia entre la correlación observada –*correlation matrix* - y la estimada a partir del modelo –*factor matrix*-). La magnitud de los residuales nos indican lo bien que se ajusta el modelo a los datos, de tal manera que si los residuales son bajos (inferiores a 0,05) podremos considerar que el modelo de Análisis Factorial se adecua a los datos, mientras que si éstos son elevados, las comunalidades serán bajas y deberemos plantearnos la aplicación de este modelo (Bisquerra, R., 1989: 310).

c. ***Rotaciones factoriales.***

La matriz factorial relaciona factores con variables. A partir de ella, deberíamos poder asociar factores con las variables que sintetiza. No obstante, la interpretación de los factores en base a ella es complejo pues en no pocas ocasiones los factores están correlacionados con casi todas las variables. Para solventar estas dificultades interpretativas, y puesto que el fin último perseguido por el Análisis Factorial es el de resumir la información de partida en factores fácilmente interpretables, la ***rotación factorial*** se presenta como la solución que nos permite transformar la matriz inicial en otra de más fácil interpretación, y se basa en la posibilidad de transformar la estructura factorial sin alterar sus propiedades matemáticas (Bisquerra, R., 1989: 312; García Ferrando, M., 1989: 439).

Para algunos autores, la obtención de la matriz factorial es solo el primer paso de la factorización, siguiéndole de forma inmediata, la rotación de la citada matriz pues solo así se consigue una buena interpretación de los factores (Cuadras, C., 1991a: 161 y 163).

La rotación factorial consiste en hacer girar los ejes de coordenadas (que representan los factores de la matriz factorial), hasta que éstos se aproximen a la nube de puntos de la variables representadas, en este caso, a partir de sus pesos. La rotación factorial transforma la matriz factorial inicial en otra matriz denominada ***matriz factorial rotada*** (*rotated factor matrix*), combinación lineal de la primera y por lo que el porcentaje de varianza explicada es igual.

El objetivo de la matriz factorial rotada es simplificar la interpretación factorial,

principio de estructura simple, lo que queda garantizado si ésta cumple una serie de requisitos (Bisquerra, R., 1989: 213):

- La matriz factorial debe tener pocos pesos altos y el resto deberá estar próximos a cero;

- Una variable no debe estar saturada más que en un solo factor;

- No deben existir factores con la misma distribución de pesos altos y bajos.

Del mismo modo que existe una gran variedad de criterios para extraer los factores, también hay distintos procedimientos para realizar las rotaciones factoriales. Todos buscan cumplir el principio de estructura simple y, para todos, las comunalidades y porcentaje de varianza total explicada no cambia, aunque sí el porcentaje de varianza atribuido a cada uno de los factores (Bisquerra, R., 1989: 315).

La variedad de rotaciones también queda recogida en Tabla 5.1. La rotación se efectuaba optando por métodos gráficos o métodos analíticos. Actualmente, los primeros, dada su escasa objetividad, ya no se aplican, debiendo recurrir a los segundos (Cuadras, C. M., 1991a: 164). Estos se dividen siguiendo dos criterios: el criterio de ***rotación ortogonal*** (los factores comunes no están correlacionados y técnicamente son más fáciles de aplicar), y el criterio de ***rotación oblicuo*** (los factores comunes están correlacionados y en consecuencia, son más realistas) (García Ferrando, M., 1989: 443). De los primeros, es la rotación ***varimax*** la más utilizada, mientras que de los segundos es el método de ***oblimín*** el más recomendado (Bisquerra, R., 1989: 316).

De todos ellos, es la rotación varimax la que se aplicará en esta investigación reduciendo a ella la exposición de su contenido. La rotación varimax para facilitar la interpretación factorial, se limita a minimizar el número de variables que tienen saturaciones altas en un factor: las variables correlacionadas entre sí presentan saturaciones altas sobre un mismo factor y bajas sobre el resto (Bisquerra, R., 1989: 316; Ferrán, M., 1996: 430). De este modo, si dos variables están fuertemente correlacionadas entre sí, sus saturaciones serán altas en un mismo factor (próximas a 1 en valores absolutos) y, en consecuencia, estarán correlacionadas entre sí (positivamente si las saturaciones comparten signo, y negativamente si es distinto); si las saturaciones altas se presentan en dos factores distintos estarán incorrelacionas.

d. ***Interpretación de los Factores***.

Una vez que ya hemos simplificado la matriz factorial a partir de la rotación de la misma y con ello concluido el proceso de análisis factorial, resta interpretar los factores obtenidos en función de las variables con las que se encuentran asociados (*matriz factorial rotada*) superando, con ello, una de las principales críticas de esta técnica. Su "debilidad" radica en el hecho de que las soluciones apuntadas por el análisis no son únicos dependiendo de: las decisiones adoptadas sobre el modo de factorizar, el tipo de rotación escogido, número de factores seleccionados, etc. Para tal objetivo, los pasos sugeridos por Biquerra Alzina son:

d.1. Estudiar la composición de las saturaciones factoriales significativas de cada factor. Para estudiar estas saturaciones factoriales, y a efectos prácticos, recomienda:

- La representación gráfica de los ejes factoriales. Las representaciones se hacen tomando los factores (representados en el eje de coordenadas –ejes factoriales-) de dos en dos (excepcionalmente de tres en tres factores). Sobre los ejes se situarán los coeficientes de correlación que asocian a las variables con los factores (Bisquerra, R., 1989: 322). A partir de las gráficas resulta sencillo desvelar la estructura latente del factor, pues las variables saturadas de un factor aparecerán agrupadas.

- Ordenar las variables en función del peso de los factores sobre éstas, de tal manera que en la matriz factorial rotada (*SORT*) aparezcan las variables con ponderaciones altas para el mismo factor agrupadas.

- Eliminar las saturaciones bajas ocupando sus espacios blancos (*BLANK*).

d.2. Intentar dar un nombre a los factores. El nombre debe adecuarse a la estructura de sus saturaciones, esto es, conociendo su contenido. En esta última fase juega un importante papel el marco teórico en el que se apoye el objeto de estudio, sin olvidar la experiencia del investigador (Bisquerra, R., 1989: 327). En última instancia, el nombre asignado tendrá carácter aproximativo (García Ferrando, M., 1989: 436).

e. ***Puntuaciones Factoriales***.

Una vez que hemos identificado y nombrado los factores o componentes latentes de un conjunto de variables, puede ser de utilidad conocer qué puntuaciones obtienen los sujetos o unidades de análisis, esto es: las variables son sustituidas por las unidades de análisis, lo que nos permitirá analizar las similitudes que se den entre los individuos respecto a sus puntuaciones y en el conjunto de variables observadas (Ferrán, M., 1996: 433).

El cálculo de las ***puntuaciones factoriales*** (*factor scores*) de cada individuo nos indicará en qué medida los factores se dan en las unidades de análisis. Por otro lado, las puntuaciones factoriales solo serán exactas si el método de extracción de factores ha sido el de componentes principales, de no ser así, las puntuaciones factoriales solo reflejarán estimaciones de las mismas (Bisquerra, R., 1989: 330-332).

El cálculo de la matriz de puntuaciones factoriales[13] adquieren verdadero protagonismo cuando se somete, con posterioridad, en otro tipo de análisis como el análisis de *cluster* o conglomerados, siendo éste su uso potencial.

13. El cálculo de la matriz de puntuaciones factoriales se realiza a partir de la matriz original de datos y la matriz de coeficientes de puntuaciones factoriales rotada y se basa en el modelo de regresión múltiple, siguiendo la fórmula (Bisquerra, R., 1989, Op. cit., pp. 330):

$$F_{ij} = F_{i1}Z_1 + F_{i2}Z_2 + \ldots + Fi_vZ_v$$

en donde:

F_{ij} es la puntuación factorial del individuo j en el factor i;

Cada F_{il} es la ponderación factorial de la variable "l" en el factor i;

Zl son las puntuaciones individuales en cada variable en puntuaciones estandarizadas.