Título:

ANÁLISIS FACTORIAL

Presentación:

El Análisis Factorial es el nombre genérico que se da a una clase de métodos estadísticos multivariantes cuyo propósito principal es sacar a la luz la estructura subyacente en una matriz de datos. Analiza la estructura de las interrelaciones entre un gran número de variables no exigiendo ninguna distinción entre variables dependientes e independientes. Utilizando esta información calcula un conjunto de dimensiones latentes, conocidas como FACTORES, que buscan explicar dichas interrelaciones. Es, por lo tanto, una técnica de reducción de datos dado que si se cumplen sus hipótesis, la información contenida en la matriz de datos puede expresarse, sin mucha distorsión, en un número menor de dimensiones representadas por dichos FACTORES. Un Análisis Factorial tiene sentido si se cumplen dos condiciones: PARSIMONIA e INTERPRETABILIDAD.

En esta lección daremos una visión general de dicha técnica y aprenderemos cuáles son los pasos a seguir a la hora de realizar un Análisis Factorial ilustrándolos con ejemplos.

Introducción

¿Qué factores tiene en cuenta una persona cuándo va a comprarse un coche?

¿Qué características distinguen unas marcas de pastas de dientes de otras ?

¿Qué tipos de aptitudes hay que tener en cuenta para evaluar la labor de un vendedor? ¿Cómo se pueden medir?

¿Cuáles son los determinantes de la resistencia de los individuos a innovaciones tecnológicas?

¿Cómo medir el grado de inteligencia de una persona? ¿Existe un único tipo de inteligencia o hay varios? ¿Si existen varios cómo medirlos?

¿Qué factores conforman la personalidad de una persona? ¿Cómo medirlos?

¿Cómo medir el nivel de desarrollo de un país?

¿Qué ratios financieros deben tenerse en cuenta a la hora de evaluar la labor desarrollada por una empresa?

¿QUÉ TIENEN EN COMÚN TODOS ESTOS PROBLEMAS? ¿CÓMO RESOLVERLOS?

En esta lección trataremos de responder a estas cuestiones.

Objetivos

1) Definir qué es el Análisis Factorial y cuáles son sus objetivos.
2) Indicar cuáles son las etapas a seguir en la realización de un Análisis Factorial.
3) Formular el modelo del Análisis Factorial e interpretar el significado de sus parámetros
4) Analizar el grado de deseabilidad de un Análisis Factorial sobre un conjunto de datos a partir del análisis de la matriz de correlación de las variables observadas.
5) Seleccionar el método apropiado para la extracción de los factores.
6) Determinar el número de factores a extraer.
7) Aprender a interpretar el significado de un factor.
8) Conocer distintos métodos de rotación de factores
9) Conocer distintos métodos de cálculo de las puntuaciones factoriales y cómo usarlas para interpretar los resultados obtenidos
10) Validar el modelo resultante de un Análisis Factorial

Apartados

Contenidos

1.- ¿QUÉ ES UN ANÁLISIS FACTORIAL?

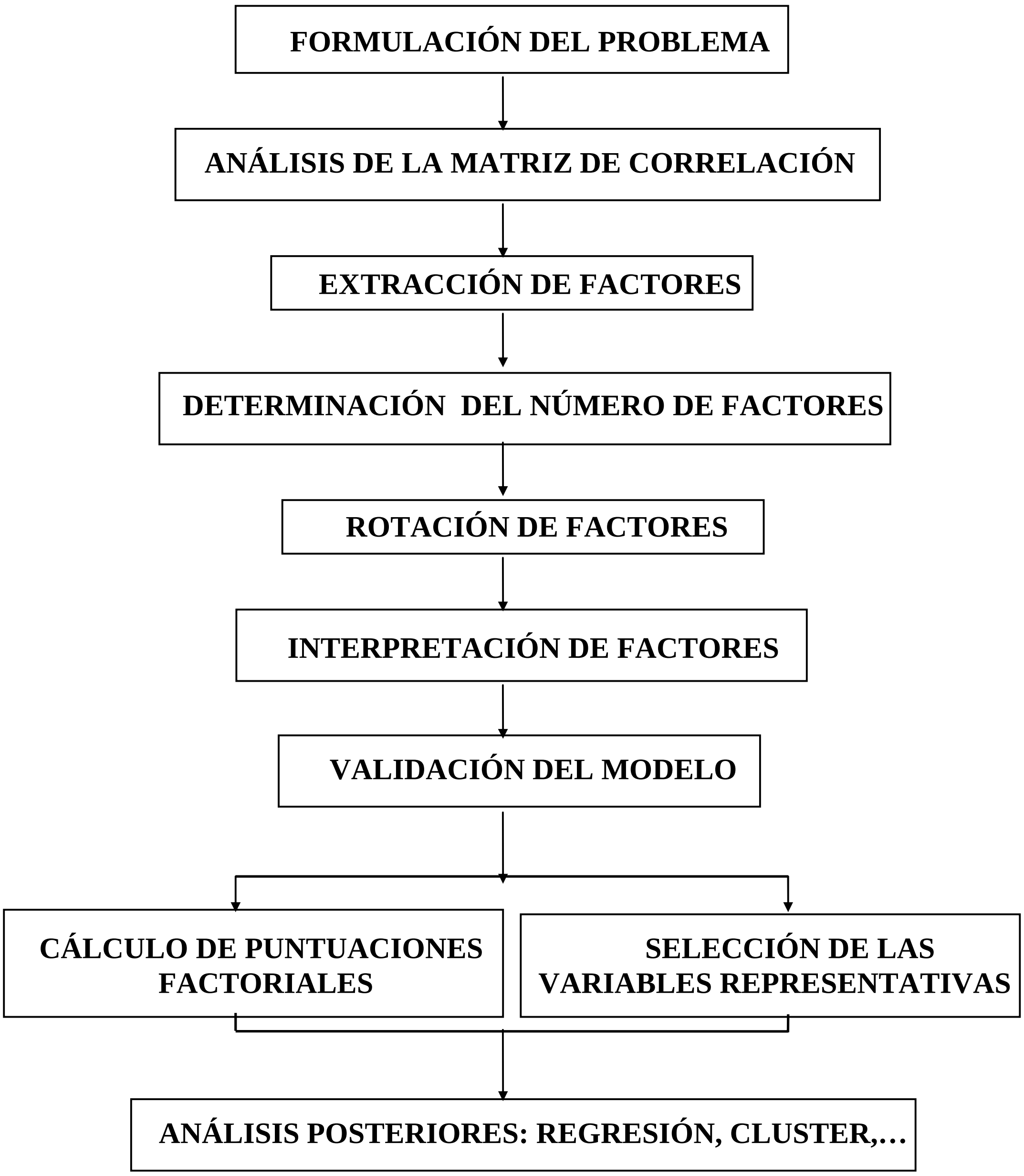
El Análisis Factorial es una técnica estadística multivariante cuyo principal propósito es sintetizar las interrelaciones observadas entre un conjunto de variables en una forma concisa y segura como una ayuda a la construcción de nuevos conceptos y teorías. Para ello utiliza un conjunto de variables aleatorias inobservables, que llamaremos ***factores comunes***, de forma que todas las covarianzas o correlaciones son explicadas por dichos factores y cualquier porción de la varianza inexplicada por los factores comunes se asigna a términos de error residuales que llamaremos ***factores únicos o específicos***.

El Análisis Factorial puede ser ***exploratorio o confirmatorio***. El **análisis exploratorio** se caracteriza porque no se conocen a priori el número de factores y es en la aplicación empírica donde se determina este número. Por el contrario, en el **análisis de tipo confirmatorio** los factores están fijados a priori, utilizándose contrastes de hipótesis para su corroboración.

En esta lección nos centraremos en el Análisis Factorial Exploratorio dado que el Análisis Factorial Confirmatorio se suele estudiar como un caso particular de los Modelos de Ecuaciones Estructurales. Remitimos al lector interesado en éste último al libro de Kline (1998) en el que se hace una buena exposición de dicho tipo de modelos.

2.- ¿CÓMO REALIZAR UN ANÁLISIS FACTORIAL?

En la siguiente figura se ilustran los pasos necesarios para la realización de un Análisis Factorial:



En los puntos siguientes se presenta con detalle en qué consiste cada una de estas etapas.

3.- FORMULACIÓN DEL PROBLEMA.

En la formulación del problema debe abordarse la selección de las variables a analizar así como la de los elementos de la población en la que dichas variables van a ser observadas.

Aunque pueden realizarse análisis factoriales con variables discretas y/o ordinales lo habitual será que las variables sean cuantitativas continuas y en lo que sigue nos ceñiremos a este caso. Es importante, en todo caso, que dichas variables recojan los aspectos más esenciales de la temática que se desea investigar y su selección deberá estar marcada por la teoría subyacente al problema. No tiene sentido incluir variables que no vengan fundamentadas por los aspectos teóricos del problema porque se corre el riesgo de que los resultados obtenidos ofrezcan una estructura factorial difícil de entender y con escaso contenido teórico relevante.

Es muy aconsejable en este paso que el analista tenga una idea más o menos clara de cuáles son los factores comunes que quiere medir y que elija las variables de acuerdo con ellos y no al revés porque se corre el riesgo de encontrar factores espúreos o que los factores queden mal estimados por una mala selección de las variables.

Así mismo, la muestra debe ser representativa de la población objeto de estudio y del mayor tamaño posible. Como regla general deberán existir por lo menos cuatro o cinco veces más observaciones (tamaño de la muestra) que variables. Si el tamaño de la muestra es pequeño y esta relación es menor, los resultados deben interpretarse con precaución.

Conviene hacer notar, finalmente, que los resultados del análisis no tienen por qué ser invariantes a cambios de origen y escala por lo que se aconseja, si las unidades de medida de las variables no son comparables, estandarizar los datos antes de realizar el análisis.

3.1.- El modelo del Análisis Factorial

Sean X_1, X_2,…, X_p las p variables objeto de análisis que supondremos en todo lo que sigue, que están tipificadas. Si no lo estuvieran el análisis se realizaría de forma similar pero la matriz utilizada para calcular los factores no sería la matriz de correlación sino la de varianzas y covarianzas.

El investigador mide estas variables sobre n individuos, obteniéndose la siguiente *matriz de datos*:

Sujetos	Variables $\mathbf{X}_1$ $\mathbf{X}_2$ … $\mathbf{X}_p$
1 **2** **…** **n**	x_{11} x_{12} … x_{1p} x_{21} x_{22} … x_{2p} ………. x_{n1} x_{n2} … x_{np}

El modelo del Análisis Factorial viene dado habitualmente por las ecuaciones:

$$\mathbf{X}_1 = a_{11}\mathbf{F}_1 + a_{12}\mathbf{F}_2 + \ldots + a_{1k}\mathbf{F}_k + \mathbf{u}_1$$

$$\mathbf{X}_2 = a_{21}\mathbf{F}_1 + a_{22}\mathbf{F}_2 + \ldots + a_{2k}\mathbf{F}_k + \mathbf{u}_2$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\mathbf{X}_p = a_{p1}\mathbf{F}_1 + a_{p2}\mathbf{F}_2 + \ldots + a_{pk}\mathbf{F}_k + \mathbf{u}_p$$

donde $\mathbf{F}_1,\ldots,\mathbf{F}_k$ ($k<<p$) son los **factores comunes y** $\mathbf{u}_1,\ldots\mathbf{u}_p$ los **factores únicos o específicos** y los coeficientes $\{a_{ij};\ i=1,\ldots,p;\ j=1,...,k\}$ las **cargas factoriales.**

Se supone, además, que los factores comunes están a su vez estandarizados ($E(F_i) = 0$; $Var(F_i) = 1$), los factores específicos tienen media 0 y están incorrelados ($E(u_i) = 0$; $Cov(u_i,u_j) = 0$ si $i\neq j$; $j, i=1,\ldots,p$) y que ambos tipos de factores están incorrelados ($Cov(F_i,u_j) = 0$, $\forall i=1,..,k$; $j=1,\ldots,p$.

Si, además, los factores están incorrelados ($Cov(F_i,F_j) = 0$ si $i \neq j$; j, i=1,…,k) estamos ante un **modelo con factores ortogonales.** En caso contrario el **modelo se dice que es de factores oblícuos.**

Expresado en forma matricial

$$\mathbf{x} = \mathbf{Af} + \mathbf{u} \Leftrightarrow \mathbf{X} = \mathbf{FA'} + \mathbf{U} \qquad (1)$$

donde $\mathbf{x} = \begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_p \end{pmatrix}$, $\mathbf{f} = \begin{pmatrix} F_1 \\ F_2 \\ \dots \\ F_k \end{pmatrix}$, $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \dots \\ u_p \end{pmatrix}$, **X** es la matriz de datos,

$\mathbf{A} = \begin{pmatrix} a_{11}\ a_{12} \dots a_{1k} \\ a_{21}\ a_{22} \dots a_{2k} \\ \dots\dots\dots\dots \\ a_{p1}\ a_{p2} \dots a_{pk} \end{pmatrix}$ es la **matriz de cargas factoriales** y

$\mathbf{F} = \begin{pmatrix} f_{11}\ f_{12} \dots f_{1k} \\ f_{21}\ f_{22} \dots f_{2k} \\ \dots\dots\dots\dots \\ f_{p1}\ f_{p2} \dots f_{pk} \end{pmatrix}$ es la **matriz de puntuaciones factoriales**

Utilizando las hipótesis anteriores se tiene que:

$$Var(X_i) = \sum_{j=1}^{k} a_{ij}^2 + \psi_i = h_i^2 + \psi_i;\ i=1,\dots,p$$

donde $h_i^2 = Var\left(\sum_{j=1}^{k} a_{ij}F_j\right)$ y $\psi_i = Var(u_i)$ reciben los nombres de **comunalidad y especificidad de la variable X_i**, respectivamente.

Por lo tanto, la varianza de cada una de las variables analizadas puede descomponerse en dos partes: una, la ***comunalidad*** h_i^2 que representa la varianza explicada por los factores comunes y otra la ***especificidad*** ψ_i que representa la parte de la varianza específica de cada variable. Además se tiene que

$$\mathrm{Cov}\,(X_i, X_\ell) = \mathrm{Cov}\left(\sum_{j=1}^{k} a_{ij}F_j, \sum_{j=1}^{k} a_{\ell j}F_j\right) = \sum_{j=1}^{k} a_{ij}a_{\ell j} \quad \forall i \neq \ell$$

por lo que son los factores comunes los que explican las relaciones existentes entre las variables del problema. Es por esta razón que los factores que tienen interés y son susceptibles de interpretación experimental son los factores comunes. Los factores únicos se incluyen en el modelo dada la imposibilidad de expresar, en general, p variables en función de un número más reducido k de factores.

Ejemplo 1(Resultados de un test de inteligencia)

A un conjunto de estudiantes se les somete a diversos tests en varias materias con el fin de medir sus aptitudes intelectuales

Como consecuencia de dichas pruebas se obtienen una serie de puntuaciones estandarizadas en Matemáticas (Mat), Física (Fca), Química (Qca), Inglés (Ing), Historia (His) y Francés (Fra).

El modelo factorial finalmente estimado viene dado por las ecuaciones

$Mat = 0.8F_1+0.2F_2+u_{Mat}$ $\qquad$ $Ing = 0.2F_1 + 0.8F_2 + u_{Ing}$

$Fca = 0.7F_1 + 0.3F_2 + u_{Fca}$ $\qquad$ $His = 0.15F_1 + 0.82F_2 +u_{His}$

$Qca = 0.6F_1 + 0.3F_2 + u_{Qca}$ $\qquad$ $Fra = 0.25F_1 + 0.85F_2 + u_{Fra}$

con $E[F_i] = E[u_j] = 0 \ \forall i=1,2;\ j\in\{Mat,Fca,Qca,Ing,His,Fra\}$

$Var[F_i] = 1;\ i=1,2;\ Cov(F_1,F_2) = 0$

$Cov(u_i,F_j) = 0;\ \forall i=1,2;\ j\in\{Mat,Fca,Qca,Ing,His,Fra\}$

$Cov(u_i,u_j) = 0;\ \forall i\neq j\in\{Mat,Fca,Qca,Ing,His,Fra\}$

Se tiene que:

$$Var[Mat]= 1 = Var[0.8F_1+0.2F_2+u_{Mat}] =$$
$$= 0.8^2Var[F_1]+0.2^2Var[F_2]+Var[u_{Mat}]+2x0.8x0.2\ Cov(F_1,F_2)+$$
$$+2x0.8\ Cov(F_1,u_{Mat}) + 2x0.2\ Cov(F_2,u_{Mat}) = 0.68+\psi_{Mat}$$

Se sigue, por lo tanto, que la comunalidad del resultado en matemáticas es $h^2_{Mat} = 0.68$ y su especificidad es $\psi_{Mat} = 0.36$

Siguiendo el mismo razonamiento se tiene que:

$$h^2_{Fca} = 0.58,\ \psi_{Fca} = 0.42$$

$$h^2_{Qca} = 0.45,\ \psi_{Qca} = 0.55$$

$$h^2_{Fra} = 0.785,\ \psi_{Fra} = 0.215$$

$$h^2_{Ing} = 0.64,\ \psi_{Ing} = 0.36$$

$$h^2_{His} = 0.6949,\ \psi_{His} = 0.3051$$

La matriz de cargas factoriales es $\mathbf{A} = \begin{pmatrix} 0.8 & 0.2 \\ 0.7 & 0.3 \\ 0.6 & 0.3 \\ 0.2 & 0.80 \\ 0.15 & 0.82 \\ 0.25 & 0.85 \end{pmatrix}$

Así mismo, se tiene que, por ejemplo

$$\mathrm{Cov}(Mat,Fca) = \mathrm{Cov}(0.8F_1+ 0.2F_2+ u_{Mat}, 0.7F_1 + 0.3F_2 + u_{Fca}) =$$

$$= 0.8x0.7\mathrm{Var}(F_1) + 0.2x0.3\mathrm{Var}(F_2) +$$

$$+ (0.8x0.3+0.2x0.7)\ \mathrm{Cov}(F_1,F_2) + 0.7\ \mathrm{Cov}(u_{Mat},F_1) +$$

$$+ 0.3\ \mathrm{Cov}(u_{Mat},F_2) + 0.8\ \mathrm{Cov}(F_1,u_{Fca}) + 0.2\ \mathrm{Cov}(F_2,u_{Fca}) +$$

$$+ \mathrm{Cov}(u_{Mat},u_{Fca}) = 0.56 + 0.06 + 0 = 0.62$$

Razonando de esta manera la matriz de varianzas y covarianzas que, al estar las puntuaciones estandarizadas coincide con la matriz de correlaciones, vendrá dada por:

$$\begin{pmatrix} 1 & \mathbf{0.62} & \mathbf{0.54} & 0.32 & 0.284 & 0.37 \\ \mathbf{0.62} & 1 & \mathbf{0.51} & 0.38 & 0.351 & 0.43 \\ \mathbf{0.54} & \mathbf{0.51} & 1 & 0.36 & 0.336 & 0.405 \\ 0.32 & 0.38 & 0.36 & 1 & \mathbf{0.686} & \mathbf{0.73} \\ 0.284 & 0.351 & 0.336 & \mathbf{0.686} & 1 & \mathbf{0.7345} \\ 0.37 & 0.43 & 0.405 & \mathbf{0.73} & \mathbf{0.7345} & 1 \end{pmatrix}$$

Se observa, en particular, que las calificaciones en las asignaturas del bloque de Ciencias (Mat, Fca y Qca) y las del bloque de Letras (Ing, His y Fra) están más correlacionadas entre sí que con las asignaturas del otro bloque.

Ejemplo 2 (Análisis de un conjunto de ratios financieros)

Como ejemplo ilustrativo de cómo llevar a cabo un Análisis Factorial estudiaremos el comportamiento de un conjunto de ratios financieros listados en la Tabla 1 y medidos sobre una muestra de empresas del sector del transporte de algunos países europeos participantes en el proyecto BACH listados en la Tabla 2.

Tabla 1: Ratios Analizados

Ratio	Fórmula
Ratio1	$\frac{\text{Activo Total}}{\text{Recursos Propios}} = \frac{AT}{RP}$
Ratio2	$\frac{\text{Beneficios antes de Impuestos}}{\text{Recursos Propios}} = \frac{BAIT}{RP}$
Ratio3	$\frac{\text{Beneficios Netos}}{\text{Recursos Propios}} = \frac{BN}{RP}$
Ratio4	$\frac{\text{Cash Flow}}{\text{Deuda a Largo Plazo}} = \frac{CF}{DLP}$
Ratio5	$\frac{\text{Cash Flow}}{\text{Recursos Propios}} = \frac{CF}{RP}$
Ratio6	$\frac{\text{Deuda a Largo Plazo}}{\text{Activo Total}} = \frac{DLP}{AT}$
Ratio7	$\frac{\text{Deuda a Largo Plazo}}{\text{Capital Propio}} = \frac{DLP}{CP}$
Ratio8	$\frac{\text{Deuda Total}}{\text{Recursos Propios}} = \frac{DT}{RP}$
Ratio9	$\frac{\text{Pasivo Circulante}}{\text{Deuda Total}} = \frac{PC}{RP}$

Tabla 2: Países participantes en el proyecto BACH

Pais	**Abreviatura**
Austria	AUT
Bélgica	BEL
Dinamarca	DNK
Finlandia	FIN
Francia	FRA
Alemania	GER
Italia	ITA
Holanda	NLD
Portugal	POR
España	SPA
Suecia	SWE

Dicho proyecto busca armonizar las cuentas financieras de las empresas de los países participantes con el fin de comparar los resultados de la actividad empresarial en cada país y en cada sector.

Los datos corresponden a valores medios de dichos ratios medidos a lo largo del periodo 1980-2001. Nuestros objetivos son analizar cuáles son los factores subyacentes a la evolución conjunta de dichos ratios y realizar un estudio comparativo por países.

4.- ANÁLISIS DE LA MATRIZ DE CORRELACIÓN

Una vez formulado el problema y obtenida la matriz de datos **X** el siguiente paso a realizar es el examen de la matriz de correlaciones maestrales R = (r_{ij}) donde r_{ij} es la correlación muestral observada entre las variables X_i y X_j. La finalidad de este análisis es comprobar si sus características son las más adecuadas para realizar un Análisis Factorial.

Uno de los requisitos que debe cumplirse para que el Análisis Factorial tenga sentido es que las variables estén altamente intercorrelacionadas. Por tanto, si las correlaciones entre todas las variables son bajas, el Análisis Factorial tal vez no sea apropiado. Además, también se espera que las variables que tienen correlación muy alta entre sí la tengan con el mismo factor o factores.

A continuación presentamos diferentes indicadores del grado de asociación entre las variables.

4.1.- Test de esfericidad de Barlett

Una posible forma de examinar la matriz de correlaciones es mediante el *test de esfericidad de Bartlett* que contrasta, bajo la hipótesis de normalidad multivariante, si la matriz de correlación de las variables observadas, R_ρ, es la identidad. Si una matriz de correlación es la identidad significa que las intercorrelaciones entre las variables son cero. Si se confirma la hipótesis nula (H_0: $|R_\rho| = 1$ o $R_\rho = I$) significa que las variables no están intercorrelacionadas.

El test de esfericidad de Bartlett se obtiene a partir de una transformación del determinante de la matriz de correlación. El estadístico de dicho test viene dado por:

$$d_R = -\left[n-1-\frac{1}{6}(2p+5)\right] \log|R| = -\left[n-\frac{2p+11}{6}\right] \sum_{j=1}^{p} \log\left(\lambda_j\right)$$

donde n es el número de individuos de la muestra y λ_j (j = 1,…,p) son los valores propios de R.

Bajo la hipótesis nula este estadístico se distribuye asíntóticamente según una distribución χ^2 con p(p-1)/2 grados de libertad.

Si H_0 es cierta los valores propios valdrían uno, o equivalentemente, su logaritmo sería nulo y, por tanto, el estadístico del test valdría cero. Por el contrario, si con el test de Bartlett se obtienen valores altos de χ^2, o equivalentemente, un determinante bajo, esto significa que hay variables con correlaciones altas (un determinante próximo a cero indica que una o más variables podrían ser expresadas como una combinación lineal de otras variables).

Así pues, si el estadístico del test toma valores grandes se rechaza la hipótesis nula con un cierto grado de significación. En caso de no rechazarse la hipótesis nula significaría que las variables no están intercorrelacionadas y en este supuesto debería reconsiderarse la aplicación de un Análisis Factorial.

4.2.- Medidas de adecuación de la muestra

El coeficiente de correlación parcial es un indicador de la fuerza de las relaciones entre dos variables eliminando la influencia del resto. Si las variables comparten factores comunes, el coeficiente de correlación parcial entre pares de variables deberá ser bajo, puesto que se eliminan los efectos lineales de las otras variables. Las correlaciones parciales son estimaciones de las correlaciones entre los factores únicos y deberían ser próximos a cero cuando el Análisis Factorial es adecuado, ya que, estos factores se supone que están incorrelados entre sí. Por lo tanto si existe un número elevado de coeficientes de este tipo distintos de cero es señal de que las hipótesis del modelo factorial no son compatibles con los datos.

Una forma de evaluar este hecho es mediante la ***Medida de Adecuación de la Muestra KMO*** propuesta por Kaiser, Meyer y Olkin. Dicha medida viene dada por

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

donde $r_{ij(p)}$ es el coeficiente de correlación parcial entre las variables X_i y X_j eliminando la influencia del resto de las variables.

KMO es un índice que toma valores entre 0 y 1 y que se utiliza para comparar las magnitudes de los coeficientes de correlación observados con las magnitudes de los coeficientes de correlación parcial de forma que, cuanto más pequeño sea su valor, mayor es el valor de los coeficientes de correlación parciales $r_{ij(p)}$ y, por lo tanto, menos deseable es realizar un Análisis Factorial.

Kaise, Meyer y Olkin aconsejan que si KMO ≥ 0,75 la idea de realizar un análisis factorial es buena, si 0,75 > KMO ≥ 0,5 la idea es aceptable y si KMO < 0,5 es inaceptable.

También se puede calcular una ***Medida de Adecuación Muestral para cada variable*** de forma similar al índice KMO. En esta prueba sólo se incluyen los coeficientes de la variable que se desea comprobar. La fórmula es:

$$\mathrm{MSA}_i = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} r_{ij(p)}^2} \; ; \; i=1,\ldots,p$$

Un valor bajo de MSA_i indica que las hipótesis hechas por el modelo del Análisis Factorial son poco compatibles para el caso de la variable X_i. De esta forma si el KMO es bajo es posible localizar las variables responsables de dichos valores y, si el Análisis Factorial resultara poco exitoso, dichas variables podrían ser eliminadas del análisis siempre y cuando su importancia teórica no lo desaconsejara.

Nuestra experiencia práctica con estos índices nos indican que es peligroso tomarlos como únicas medidas de adecuación de la muestra a las hipótesis del modelo del Análisis Factorial, sobre todo si el número de variables consideradas es pequeño. Conviene complementar dicha información con otras fuentes como pueden ser las comunalidades de cada variable, los residuos del modelo y la interpretabilidad de los factores obtenidos a la hora de tomar la decisión de eliminar una variable del estudio.

Ejemplo 2 (continuación)

En la Tabla 3 se muestran los resultados obtenidos al analizar la matriz de correlación mediante el programa SPSS 10.0

Matriz de correlaciones

	AT/RP	BAIT/RP	BN/RP	CF/DLP	CF/RP	DLP/AT	DLP/KP	DT/RP	PC/DT
AT/RP	1.000	.096	-.358	-.285	-.093	.356	.521	.921	-.180
BAIT/RP	.096	1.000	.711	.162	.553	-.047	-.006	.087	.085
BN/RP	-.358	.711	1.000	.323	.762	-.202	-.260	-.343	.141
CF/DLP	-.285	.162	.323	1.000	.440	-.669	-.717	-.511	.642
CF/RP	-.093	.553	.762	.440	1.000	-.217	-.203	-.108	.227
DLP/AT	.356	-.047	-.202	-.669	-.217	1.000	.940	.550	-.956
DLP/KP	.521	-.006	-.260	-.717	-.203	.940	1.000	.748	-.821
DT/RP	.921	.087	-.343	-.511	-.108	.550	.748	1.000	-.343
PC/DT	-.180	.085	.141	.642	.227	-.956	-.821	-.343	1.000

Significación (Unilateral)

	AT/RP	BAIT/RP	BN/RP	CF/DLP	CF/RP	DLP/AT	DLP/KP	DT/RP	PC/DT
AT/RP		.106	.000	.000	.113	.000	.000	.000	.009
BAIT/RP	.106		.000	.017	.000	.270	.469	.130	.134
BN/RP	.000	.000		.000	.000	.004	.000	.000	.033
CF/DLP	.000	.017	.000		.000	.000	.000	.000	.000
CF/RP	.113	.000	.000	.000		.002	.004	.081	.001
DLP/AT	.000	.270	.004	.000	.002		.000	.000	.000
DLP/KP	.000	.469	.000	.000	.004	.000		.000	.000
DT/RP	.000	.130	.000	.000	.081	.000	.000		.000
PC/DT	.009	.134	.033	.000	.001	.000	.000	.000	

KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		.585
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	2346.465
	gl	36
	Sig.	.000

Matrices anti-imagen

		AT/RP	BAIT/RP	BN/RP	CF/DLP	CF/RP	DLP/AT	DLP/KP	DT/RP	PC/DT
Covarianza anti-imagen	AT/RP	2.896E-02	-1.59E-02	6.740E-03	-4.22E-02	2.208E-02	1.547E-03	1.276E-02	-1.79E-02	1.082E-02
	BAIT/RP	-1.59E-02	.329	-.161	-9.22E-03	5.360E-02	6.412E-04	-7.52E-03	3.453E-03	-4.94E-03
	BN/RP	6.740E-03	-.161	.158	2.620E-02	-.132	-1.49E-03	3.604E-03	1.415E-03	7.549E-04
	CF/DLP	-4.22E-02	-9.22E-03	2.620E-02	.188	-.110	-1.83E-02	-1.46E-03	2.800E-02	-3.53E-02
	CF/RP	2.208E-02	5.360E-02	-.132	-.110	.255	7.256E-03	4.092E-03	-1.84E-02	1.462E-02
	DLP/AT	1.547E-03	6.412E-04	-1.49E-03	-1.83E-02	7.256E-03	8.967E-03	-7.04E-03	-1.15E-03	1.058E-02
	DLP/KP	1.276E-02	-7.52E-03	3.604E-03	-1.46E-03	4.092E-03	-7.04E-03	1.544E-02	-8.93E-03	-1.87E-03
	DT/RP	-1.79E-02	3.453E-03	1.415E-03	2.800E-02	-1.84E-02	-1.15E-03	-8.93E-03	1.189E-02	-7.67E-03
	PC/DT	1.082E-02	-4.94E-03	7.549E-04	-3.53E-02	1.462E-02	1.058E-02	-1.87E-03	-7.67E-03	1.718E-02
Correlación anti-imagen	AT/RP	.434[a]	-.163	9.971E-02	-.571	.257	9.599E-02	.603	-.963	.485
	BAIT/RP	-.163	.597[a]	-.708	-3.71E-02	.185	1.181E-02	-.105	5.524E-02	-6.57E-02
	BN/RP	9.971E-02	-.708	.616[a]	.152	-.657	-3.96E-02	7.301E-02	3.269E-02	1.450E-02
	CF/DLP	-.571	-3.71E-02	.152	.570[a]	-.502	-.446	-2.71E-02	.592	-.620
	CF/RP	.257	.185	-.657	-.502	.561[a]	.152	6.518E-02	-.333	.221
	DLP/AT	9.599E-02	1.181E-02	-3.96E-02	-.446	.152	.675[a]	-.599	-.111	.852
	DLP/KP	.603	-.105	7.301E-02	-2.71E-02	6.518E-02	-.599	.716[a]	-.659	-.115
	DT/RP	-.963	5.524E-02	3.269E-02	.592	-.333	-.111	-.659	.511[a]	-.537
	PC/DT	.485	-6.57E-02	1.450E-02	-.620	.221	.852	-.115	-.537	.567[a]

a. Medida de adecuación muestral

Se observa que el valor del KMO = 0.585 está en el límite de los valores recomendados por Kaiser, Meyer y Olkin poniendo de manifiesto que, en este caso muy probablemente, el proceso de reducción de datos no sea muy espectacular. El test de esfericidad de Bartlett, sin embargo, rechaza la hipótesis de diagonalidad de la matriz de correlación indicando que sí existen relaciones significativas entre las variables. Observando las medidas de adecuación muestrales se observa que toman valores bajos en especial para el ratio AT/RP señalando a dicha variable como una posible candidata a ser eliminada del análisis si su comunalidad no es muy alta.

5.- EXTRACCIÓN DE FACTORES

Como ya hemos comentado, el objetivo del Análisis Factorial consiste en determinar un número reducido de factores que puedan representar a las variables originales. Por tanto, una vez que se ha determinado que el Análisis Factorial es una técnica apropiada para analizar los datos, debe seleccionarse el método adecuado para la extracción de los factores. Existen diversos métodos cada uno de ellos con sus ventajas e inconvenientes.

El modelo factorial en forma matricial viene dado por **X** = **FA**' **+ U** y el problema consiste en cuantificar la matriz A de cargas factoriales que explica **X** en función de los factores. A partir de esta expresión se deduce la llamada *identidad fundamental del Análisis Factorial:*

$$\mathbf{R}_\rho = \mathbf{AA'} + \psi \qquad (2)$$

donde $\mathbf{R}_\rho$ es la matriz de correlación poblacional de las variables $X_1, ..., X_p$ y $\mathbf{\Psi} = \text{diag}(\psi_i)$ es la matriz diagonal de las especificidades.

En este sentido surgen dos problemas:

a) ***Problema de los Grados de Libertad***: igualando cada elemento de la matriz $\mathbf{R}_\rho$ con la combinación lineal correspondiente al 2º miembro de la ecuación (2) resultan pxp ecuaciones, que es el número de elementos de R. Ahora bien, la matriz $\mathbf{R}_\rho$ es simétrica y, consecuentemente, está integrada por p(p+1)/2 elementos distintos, que es el número real de ecuaciones de que disponemos. En el segundo miembro los parámetros a estimar son los pxk elementos de la matriz A y los p elementos de la matriz $\mathbf{\Psi}$. En consecuencia, para que el proceso de estimación pueda efectuarse se requiere que el número de ecuaciones sea mayor o igual que el número de parámetros a estimar ($p(p+1)/2 \geq p(k+1)$) o equivalentemente $k \leq (p-1)/2$.

b) Problema de la No Unicidad de la Solucion.- Aun cuando no se presente el problema anterior hay que tener en cuenta que las soluciones dadas para la matriz A no son únicas, puesto que cualquier transformación ortogonal de A es también una solución. Si T es una matriz ortogonal, entonces $TT' = T'T = I$, al aplicar una transformación ortogonal a A se obtiene una solución distinta al sistema anterior. Esta es la base de los métodos de rotación de factores. Por tanto, si T es una matriz ortogonal, entonces $A^* = AT$ es solución, definimos $F^* = FT$ otros factores (F^* es el vector F rotado por la matriz ortogonal T). Se comprueba que X y R_ρ siguen verificando las ecuaciones del modelo, es decir:

$$R_\rho = A^*A^{*'} + \Psi = (AT)(T'A') + \Psi = AA' + \Psi$$

$$X = F^*A^{*'} + U = (FT)(T'A') + U = FA' + U$$

Por lo tanto, el modelo es único salvo rotaciones ortogonales, es decir, se pueden realizar rotaciones de la matriz de ponderaciones o cargas factoriales sin alterar el modelo.

Ejemplo 1 (continuación)

Si definimos los factores

$$F_1' = \frac{1}{\sqrt{2}}F_1 + \frac{1}{\sqrt{2}}F_2 \text{ y } F_2' = -\frac{1}{\sqrt{2}}F_1 + \frac{1}{\sqrt{2}}F_2$$

se tiene que

$Mat = 0.71\,F_1' - 0.42\,F_2' + u_{Mat}$ $Ing = 0.71\,F_1' + 0.42\,F_2' + u_{Ing}$

$\text{Fca} = 0.71\,F_1' - 0.28\,F_2' + u_{Fca}$ $\qquad$ $\text{His} = 0.69\,F_1' + 0.47\,F_2' + u_{His}$

$\text{Qca} = 0.64\,F_1' - 0.21\,F_2' + u_{Qca}$ $\qquad$ $\text{Fra} = 0.78\,F_1' + 0.42\,F_2' + u_{Fra}$

verificándose, además, que $\text{Cov}(F_1',\ F_2') = 0$ por lo que las nuevas cargas factoriales serán las correlaciones de los nuevos factores con las variables originales. Las comunalidades, especificidades y matrices de correlación permanecen idénticas.

En este caso, la matriz ortogonal T viene dada por:

$$T = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

5.1 Métodos de extracción de factores

Existen muchos métodos para obtener los factores comunes. En esta lección daremos una breve referencia de algunos de ellos; más concretamente de los implementados en SPSS.

5.1.1.- Método de las Componentes Principales

El método consiste en estimar las puntuaciones factoriales mediante las puntuaciones tipificadas de las k primeras componentes principales y la matriz de cargas factoriales mediante las correlaciones de las variables originales con dichas componentes. Este método tiene la ventaja de que siempre proporciona una solución. Tiene el inconveniente, sin embargo, de que al no estar basado en el modelo de Análisis Factorial puede llevar a estimadores muy sesgados de la matriz de cargas factoriales, particularmente, si existen variables con comunalidades bajas.

5.1.2.- Método de los Ejes Principales

Este método está basado en la identidad fundamental del Análisis Factorial (2) sustituyendo la matriz de correlaciones poblacionales $\mathbf{R}_\rho$ por la de correlaciones muestrales $\mathbf{R}$. Se sigue de (2) que

$$\mathbf{R}^* = \mathbf{R}\text{-}\boldsymbol{\Psi} = \mathbf{AA'} \qquad (3)$$

El método es iterativo y consiste en alternar una estimación de la matriz de especificidades $\boldsymbol{\Psi}$, con una estimación de la matriz de cargas factoriales $\mathbf{A}$ respetando la identidad (3). Se parte de una estimación inicial de la matriz $\boldsymbol{\Psi}$, $\boldsymbol{\Psi}^{(0)}$ y en el paso i-ésimo del algoritmo se verifica que:

$$\mathbf{R}\text{-}\boldsymbol{\Psi}^{(i)} = \mathbf{A}^{(i)}\mathbf{A}^{(i)'} \qquad (4)$$

La estimación $\mathbf{A}^{(i)}$ se obtiene aplicando el método de componentes principales a la matriz $\mathbf{R}\text{-}\boldsymbol{\Psi}^{(i-1)}$. Posteriormente se calcula $\boldsymbol{\Psi}^{(i)}$ a partir de la identidad (4) y se itera hasta que los valores de dichas estimaciones apenas cambien. Este método tiene la ventaja de estar basado en el modelo del Análisis Factorial por lo que suele proporcionar mejores estimaciones que el método anterior. Sin embargo, no está garantizada su convergencia, sobre todo en muestras pequeñas.

5.1.3.- Método de la Máxima Verosimilitud

Este método está basado en el modelo (1) adoptando, además, la hipótesis de normalidad multivariante y consiste en aplicar el método de la máxima verosimilitud. El método tiene la ventaja sobre los dos anteriores de que las estimaciones obtenidas no dependen de la escala de medida de las variables. Además, al estar basado en el método de la máxima verosimilitud, tiene todas las propiedades estadísticas de éste y, en particular, es asintóticamente insesgada, eficiente y normal si las hipótesis del modelo factorial son ciertas. Permite, además, seleccionar el número de factores mediante contrastes de hipótesis. Este método también se puede utilizar en el Análisis Factorial Confirmatorio, donde el investigador puede plantear hipótesis como que algunas cargas factoriales son nulas, que algunos factores están correlacionados con determinados factores, etc. y aplicar tests estadísticos para determinar si los datos confirman las restricciones asumidas. Su principal inconveniente radica en que, al realizarse la optimización de la función de verosimilitud por métodos iterativos, si las variables originales no son normales, puede haber problemas de convergencia sobre todo en muestras finitas.

5.1.4.- Otros métodos de extracción

El *Método de mínimos cuadrados no ponderados* que produce, para un número fijo de factores, una matriz de coeficientes que minimiza la suma de las diferencias al cuadrado entre las matrices de correlación observada, **R,** y la reproducida, $\tilde{\mathbf{R}} = \tilde{\mathbf{A}}\tilde{\mathbf{A}}'$ eliminando, en dichas diferencias, los elementos de la diagonal.

El *Método de mínimos cuadrados generalizados* que minimiza el mismo criterio pero ponderando las correlaciones inversamente por la varianza del factor específico. Este método permite, además, aplicar contrastes de hipótesis para determinar el número de factores.

El *Método de factorización por imágenes* que consiste en aplicar el método del eje principal a la matriz de correlaciones $\tilde{\mathbf{R}}$ obtenida a partir de las partes predichas de las diversas regresiones lineales de cada una de las variables sobre las demás (dicha parte recibe el nombre de imagen de la variable).

El *Método Alfa* que maximiza el alfa de Cronbach para los factores

5.1.5.- Comparación entre los distintos métodos

1) Cuando las comunalidades son altas (mayores que 0.6) todos los procedimientos tienden a dar la misma solución.

2) Cuando las comunalidades son bajas para algunas de las variables el método de componentes principales tiende a dar soluciones muy diferentes del resto de los métodos, con cargas factoriales mayores.

3) Si el número de variables es alto (mayor que 30), las estimaciones de la comunalidad tienen menos influencia en la solución obtenida y todos los métodos tienden a dar el mismo resultado.

4) Si el número de variables es bajo todo depende del método utilizado para estimar las comunalidades y de si éstas son altas más que del método utilizado para estimarlas.

5) Es más robusto, por lo tanto, utilizar un método para el modelo de factores comunes. Su único problema puede ser la falta de convergencia del método utilizado.

6.- DETERMINACIÓN DEL NÚMERO DE FACTORES

La matriz factorial puede presentar un número de factores superior al necesario para explicar la estructura de los datos originales. Generalmente, hay un conjunto reducido de factores, los primeros, que contienen casi toda la información. Los otros factores suelen contribuir relativamente poco. Uno de los problemas que se plantean consiste en determinar el número de factores que conviene conservar puesto que de lo que se trata es de cumplir el *principio de parsimonia*. Se han dado diversas reglas y criterios para determinar el número de factores a conservar. A continuación, listamos algunos de los más utilizados.

6.1.- Determinación "a priori"

Este es el criterio más fiable si los datos y las variables están bien elegidos y el investigador conoce a fondo el terreno que pisa puesto que, como ya comentamos anteriormente, lo ideal es plantear el Análisis Factorial con una idea previa de cuántos factores hay y cuáles son.

6.2.- Regla de Kaiser

Consiste en calcular los valores propios de la matriz de correlaciones **R** y tomar como número de factores el número de valores propios superiores a la unidad. Este criterio es una reminiscencia del Análisis de Componentes Principales y se ha comprobado en simulaciones que, generalmente, tiende a infraestimar el número de factores por lo que se recomienda su uso para establecer un límite inferior. Un límite superior se calcularía aplicando este mismo criterio pero tomando como límite 0.7.

6.3.- Criterio del porcentaje de la varianza

También es una reminiscencia del Análisis de Componentes Principales y consiste en tomar como número de factores el número mínimo necesario para que el porcentaje acumulado de la varianza explicado alcance un nivel satisfactorio que suele ser del 75% o el 80%. Tiene la ventaja de poderse aplicar también cuando la matriz analizada es la de varianzas y covarianzas pero no tiene ninguna justificación teórica ni práctica.

6.4.- Gráfico de Sedimentación

Consiste en una representación gráfica donde los factores están en el eje de abscisas y los valores propios en el de ordenadas. Los factores con varianzas altas se suelen distinguir de los factores con varianzas bajas. El punto de distinción viene representado por un punto de inflexión en la gráfica. Se pueden conservar los factores situados antes de este punto de inflexión. En simulaciones este criterio ha funcionado bien pero tiene el inconveniente de que depende del "ojo" del analista.

6.5.- Criterio de división a la mitad

La muestra se divide en dos partes iguales tomadas al azar y se realiza el Análisis Factorial en cada una de ellas. Sólo se conservan los factores que tienen alta correspondencia de cargas de factores en las dos muestras. Es conveniente, sin embargo, antes de aplicarlo comprobar que no existen diferencias significativas entre las dos muestras en lo que a las variables estudiadas se refiere

6.6.- Pruebas de significación

Ya se ha comentado en la sección anterior y consiste en aplicar contrastes de hipótesis de modelos anidados para seleccionar dicho número. Este criterio se puede calcular si el método utilizado para estimar los factores es el de máxima verosimilitud En la mayoría de los estudios exploratorios k no puede ser especificado por adelantado y, por tanto, se utilizan procedimientos secuenciales para determinar k. Se comienza con un valor pequeño para k (usualmente 1), los parámetros en el modelo factorial son estimados utilizando el método de máxima verosimilitud. Si el estadístico del test no es significativo, aceptamos el modelo con este número de factores, en otro caso, se aumenta k en uno y se repite el proceso hasta que se alcance una solución aceptable. El principal inconveniente de este método es que está basado en resultados asintóticos y que, si el tamaño de la muestra es grande, se corre el riesgo de tomar k excesivamente grande puesto que el test detecta cualquier factor por pequeño que sea su poder explicativo.

7.- INTERPRETACIÓN DE LOS FACTORES

La interpretación de los factores se basa en las correlaciones estimadas de los mismos con las variables originales del problema. Observar que, si el modelo de Análisis Factorial es cierto, se tiene que:

$$\mathrm{Corr}(X_i,F_\ell) = \mathrm{Cov}(X_i,F_\ell) = \sum_{j=1}^{k} a_{ij}\mathrm{Cov}\left(F_j, F_\ell\right) \quad \forall i=1,\ldots,p;\ \ell=1,\ldots,k$$

y, en particular, si los factores son ortogonales

$$\mathrm{Corr}(X_i,F_\ell) = a_{i\ell} \quad \forall i=1,\ldots,p;\ \ell=1,\ldots,k$$

Vemos, por lo tanto, que la matriz de cargas factoriales, **A**, juega un papel fundamental en dicha interpretación. Además, las cargas factoriales al cuadrado $a_{i\ell}^2$ indican, si los factores son ortogonales, qué porcentaje de la varianza de la variable original $\mathbf{X}_i$ es explicado por el factor F_ℓ.

En la fase de interpretación juega un papel preponderante la teoría existente sobre el tema. A efectos prácticos, en la interpretación de los factores se sugieren los dos pasos siguientes:

1) Identificar las variables cuyas correlaciones con el factor son las más elevadas en valor absoluto
2) Intentar dar un nombre a los factores. El nombre debe asignarse de acuerdo con la estructura de sus correlaciones con las variables. Si dicha correlación es positiva (resp. negativa) la relación entre el factor y dicha variable es directa (resp. inversa). Analizando con qué variables tiene una relación

fuerte es posible, en muchos casos, hacerse una idea más o menos clara de cuál es el significado de un factor.

Una ayuda en la interpretación de los factores puede ser representar gráficamente los resultados obtenidos. La representación se hace tomando los factores dos a dos. Cada factor representa un eje de coordenadas. A estos ejes se les denomina *ejes factoriales*. Sobre estos ejes se proyectan las variables originales. Las coordenadas vienen dadas por los respectivos coeficientes de correlación entre la variable y el factor de forma que las variables saturadas en un mismo factor aparecen agrupadas. Esto puede ayudar a descubrir la estructura latente de este factor. Las variables al final de un eje son aquellas que tienen correlaciones elevadas sólo en ese factor y, por consiguiente, lo describen. Las variables cerca del origen tienen correlaciones reducidas en ambos factores. Las variables que no están cerca de ninguno de los ejes se relacionan con ambos factores.

Dos estrategias más pueden ayudar a interpretar los factores: a) *ordenarlos* y b) *eliminar las cargas bajas*. Se puede ordenar la matriz factorial de tal forma que las variables con cargas altas para el mismo factor aparezcan juntas. La eliminación de las cargas factoriales bajas también facilita la interpretación de los resultados, al suprimir información redundante. El investigador debe decidir a partir de qué valor deben eliminarse las cargas factoriales. Ambas posibilidades pueden utilizarse conjuntamente de cara a una mayor facilidad interpretativa. En general, y como consejo, tomaremos como significativas cargas factoriales superiores a 0.5 en valor absoluto. Sin embargo, conforme el factor es más tardío o el número de variables es mayor elevaremos el valor mínimo de la carga factorial significativa.

Ejemplo 1 (continuación)

Se tiene que

$$\mathrm{Corr}(Mat,F_1) = \mathrm{Cov}(Mat,F_1) = \mathrm{Cov}(0.8F_1+0.2F_2+u_{Mat},F_1) =$$

$$= 0.8\ \mathrm{Var}(F_1)+0.2\ \mathrm{Cov}(F_2,F_1) + \mathrm{Cov}(u_{Mat},F_1) = 0.8$$

y, en general, dado que los factores F_1 y F_2 son ortogonales, las correlaciones de las calificaciones de los test con dichos factores vendrán dadas por las cargas factoriales.

Observando el patrón de dichas cargas se aprecia que el factor F_1 está muy relacionado con las variables Mat, Fca y Qca pero poco relacionado con Ing, His y Fra. Por su parte el factor F_2 está muy relacionado con las variables Ing, His y Fra y poco con el resto.

Por tanto, podemos interpretar que el factor F_1 mide la aptitud científica del alumno mientras que el factor F_2 mide su aptitud verbal.

Razonando de manera similar, si analizamos la matriz de cargas factoriales correspondientes a los factores $F_1^{'}$ y $F_2^{'}$, se observa que el factor $F_1^{'}$ está muy relacionado con todas las variables de forma directa y, por tanto, podría interpretarse como un factor de inteligencia general mientras que $F_2^{'}$ realiza un contraste entre la aptitud verbal y la aptitud científica de un estudiante al estar relacionado de forma directa con las variables Ing, His y Fra y de forma inversa con Mat, Fca y Qca.

¿Cuál de las dos interpretaciones es más correcta?. Todo dependerá de la teoría subyacente al problema que llevará al analista a hacer más hincapié en una interpretación que en otra. Deberá, sin embargo, validar el modelo elegido siguiendo algunas de los procedimientos vistos en la sección 10 de esta lección

Ejemplo 2 (continuación)

En la Tabla 4 y en la Figura 1 se muestran los valores propios de la matriz de correlaciones y el gráfico de sedimentación. Tanto el criterio del valor propio mayor que 1 o 0.7 como el gráfico de sedimentación sugieren la presencia de 3 factores que explicarían el 88.44% de la variación total de los datos. Por lo tanto optamos por extraer 3 factores.

Tabla 4: Autovalores de la matriz de correlación y porcentajes de varianza explicada

Varianza total explicada

Factor	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	4.416	49.066	49.066	4.301	47.792	47.792
2	2.149	23.879	72.945	1.925	21.384	69.176
3	1.395	15.496	88.440	1.307	14.521	83.697
4	.496	5.510	93.950			
5	.330	3.670	97.620			
6	.114	1.272	98.892			
7	8.809E-02	.979	99.871			
8	6.178E-03	6.865E-02	99.939			
9	5.454E-03	6.060E-02	100.000			

Método de extracción: Mínimos cuadrados no ponderados.

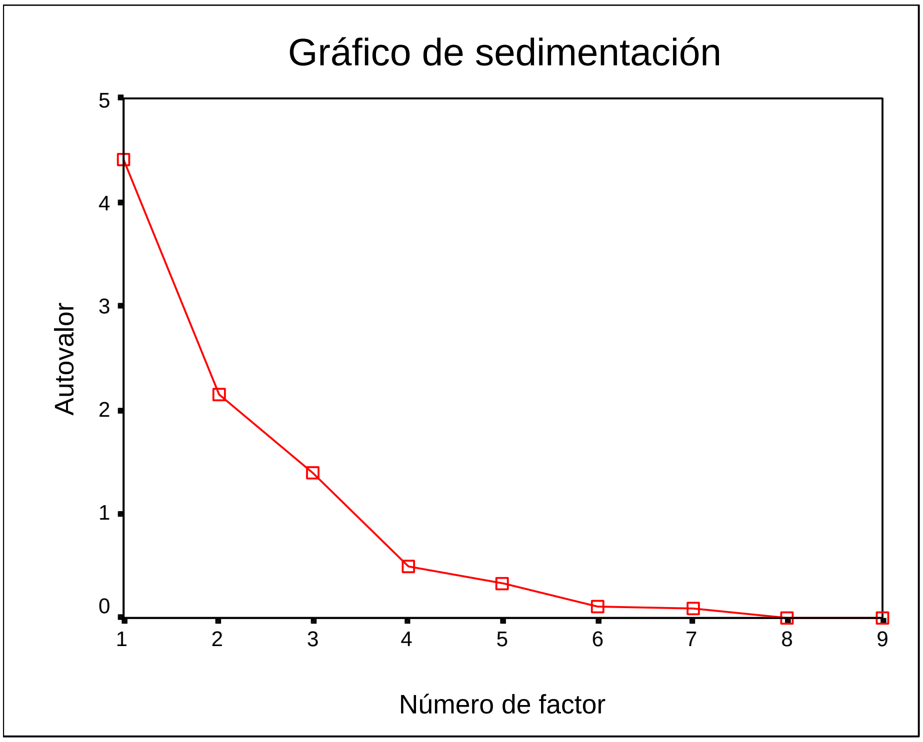


Figura 1: Gráfico de sedimentación

En la Tabla 5 se muestra la matriz factorial y las comunalidades estimadas por el método de los mínimos cuadrados no ponderados que fue el que mejor se ajustó a los datos en términos de las correlaciones reproducidas, las cuales se presentan en la Tabla 6. Se observa que en todos los casos las comunalidades estimadas superan el 50% y que tan sólo un 5% de los residuos superan un valor absoluto de 0.05 aunque la solución debe interpretarse con precaución dado que en las iteraciones realizadas por el método algunas comunalidades superaron el 100%.

Tabla 5
Matriz factorial y comunalidades estimadas

Matriz factorial[a]

	Factor		
	1	2	3
DLP/KP	.937		
DLP/AT	.897		
DT/RP	.788		.583
PC/DT	-.785		.511
CF/DLP	-.731		
BN/RP	-.518	.833	
BAIT/RP		.731	
CF/RP		.676	
AT/RP	.607		.685

Método de extracción: Mínimos cuadrados no ponderados.
a. 3 factores extraídos. Requeridas 7 iteraciones.

Comunalidades[a]

	Inicial	Extracción
AT/RP	.971	.862
BAIT/RP	.671	.586
BN/RP	.842	.999
CF/DLP	.812	.565
CF/RP	.745	.638
DLP/AT	.991	.985
DLP/KP	.985	.955
DT/RP	.988	.999
PC/DT	.983	.913

Método de extracción: Mínimos cuadrados no ponderados.
a. Se han encontrado una o más estimaciones de la comunalidad mayores que 1,0 durante las iteraciones. La solución resultante debe interpretarse con precaución.

Tabla 6
Matriz de correlaciones reproducidas y residuos

Correlaciones reproducidas

		AT/RP	BAIT/RP	BN/RP	CF/DLP	CF/RP	DLP/AT	DLP/KP	DT/RP	PC/DT
Correlación reproducida	AT/RP	.862[b]	8.282E-02	-.341	-.321	-9.06E-02	.342	.539	.914	-.156
	BAIT/RP	8.282E-02	.586[b]	.679	.190	.584	-3.69E-02	-3.13E-03	9.273E-02	7.512E-02
	BN/RP	-.341	.679	1.000[b]	.375	.760	-.184	-.249	-.347	.134
	CF/DLP	-.321	.190	.375	.565[b]	.347	-.705	-.692	-.467	.653
	CF/RP	-9.06E-02	.584	.760	.347	.638[b]	-.240	-.226	-.122	.243
	DLP/AT	.342	-3.69E-02	-.184	-.705	-.240	.985[b]	.938	.559	-.928
	DLP/KP	.539	-3.13E-03	-.249	-.692	-.226	.938	.955[b]	.738	-.836
	DT/RP	.914	9.273E-02	-.347	-.467	-.122	.559	.738	1.000[b]	-.364
	PC/DT	-.156	7.512E-02	.134	.653	.243	-.928	-.836	-.364	.913[b]
Residual[a]	AT/RP		1.317E-02	-1.70E-02	3.657E-02	-2.48E-03	1.320E-02	-1.82E-02	6.662E-03	-2.47E-02
	BAIT/RP	1.317E-02		3.176E-02	-2.76E-02	-3.08E-02	-1.02E-02	-2.79E-03	-6.11E-03	1.016E-02
	BN/RP	-1.70E-02	3.176E-02		-5.18E-02	1.903E-03	-1.77E-02	-1.07E-02	3.721E-03	6.967E-03
	CF/DLP	3.657E-02	-2.76E-02	-5.18E-02		9.326E-02	3.637E-02	-2.45E-02	-4.32E-02	-1.14E-02
	CF/RP	-2.48E-03	-3.08E-02	1.903E-03	9.326E-02		2.218E-02	2.242E-02	1.435E-02	-1.61E-02
	DLP/AT	1.320E-02	-1.02E-02	-1.77E-02	3.637E-02	2.218E-02		1.560E-03	-8.89E-03	-2.84E-02
	DLP/KP	-1.82E-02	-2.79E-03	-1.07E-02	-2.45E-02	2.242E-02	1.560E-03		9.124E-03	1.549E-02
	DT/RP	6.662E-03	-6.11E-03	3.721E-03	-4.32E-02	1.435E-02	-8.89E-03	9.124E-03		2.090E-02
	PC/DT	-2.47E-02	1.016E-02	6.967E-03	-1.14E-02	-1.61E-02	-2.84E-02	1.549E-02	2.090E-02	

Método de extracción: Mínimos cuadrados no ponderados.

a. Los residuos se calculan entre las correlaciones observadas y reproducidas. Hay 2 (5.0%) residuos no redundantes con valores absolutos > 0,05.

b. Comunalidades reproducidas

Si analizamos la matriz factorial estimada (en la que se han eliminado las cargas factoriales cuyo valor absoluto es menor que 0.5) no se observa una interpretación clara de los factores dada la gran cantidad de cargas factoriales con valores intermedios y debido a que el primer factor está relacionado con muchas variables. Para obtener una solución más inteligible es necesario recurrir a métodos de rotación de factores que se explican a continuación.

8.- ROTACIÓN DE FACTORES

Como ya se ha visto en la sección anterior, la matriz de cargas factoriales juega un papel destacado a la hora de interpretar el significado de los factores y, si éstos son ortogonales, cuantifican el grado y tipo de la relación entre éstos y las variables originales. Sin embargo, rara vez los métodos de extracción de factores vistos en la sección 5 proporcionan matrices de cargas factoriales adecuadas para la interpretación.

Para resolver este problema están los procedimientos de ***Rotación de Factores*** que, basándose en la falta de identificabilidad del modelo por rotaciones vista en la sección 5, buscan obtener, a partir de la solución inicial, unos factores cuya matriz de cargas factoriales los haga más fácilmente interpretables.

Dichos métodos intentan aproximar la solución obtenida ***al Principio de Estructura Simple*** (Thurstone, 1935) según el cual la matriz de cargas factoriales debe reunir las siguientes características:

1) cada factor debe tener unos pocos pesos altos y los otros próximos a cero;
2) cada variable no debe estar saturada más que en un factor;
3) no deben existir factores con la misma distribución, es decir, dos factores distintos deben presentar distribuciones diferentes de cargas altas y bajas.

De esta forma, y dado que hay más variables que factores comunes, cada factor tendrá una correlación alta con un grupo de variables y baja con el resto de variables. Examinando las características de las variables de un grupo asociado a un determinado factor se pueden encontrar rasgos comunes que permitan identificar el factor y darle una denominación que responda a esos rasgos comunes.

Si se consiguen identificar claramente estos rasgos, se habrá dado un paso importante, ya que con los factores comunes no sólo se *reducirá la dimensión* del problema, sino que también se conseguirá desvelar la *naturaleza de las interrelaciones* existentes entre las variables originales.

Existen dos formas básicas de realizar la rotación de factores: ***la Rotación Ortogonal*** y ***la Rotación Oblicua*** según que los factores rotados sigan siendo ortogonales o no. Conviene advertir que tanto en la rotación ortogonal, como en la rotación oblicua la comunalidad de cada variable no se modifica, es decir, la rotación no afecta a la bondad de ajuste de la solución factorial: aunque cambie la matriz factorial, las especificidades no cambian y por tanto, las comunalidades permanecen inalteradas. Sin embargo, cambia la varianza explicada por cada factor, luego los nuevos factores rotados no están ordenados de acuerdo con la información que contienen, cuantificada a través de su varianza.

8.1.- Rotación Ortogonal

En la rotación ortogonal, los ejes se rotan de forma que quede preservada la incorrelación entre los factores. Dicho de otra forma, los nuevos ejes, o ejes rotados, son perpendiculares de igual forma que lo son los factores sin rotar.

Como ya se ha dicho dicha rotación se apoya en el problema de la falta de identificabilidad de los factores obtenidos por rotaciones ortogonales de forma que si **T** es una matriz ortogonal con **TT' = T'T = I**, entonces:

$$\mathbf{X = FA' + U = FTT'A' + U = GB' + U}$$

La matriz G *geométricamente* es una rotación de F y verifica las mismas hipótesis que ésta. Lo que realmente se realiza es un giro de ejes, de manera que cambian las cargas factoriales y los factores. Se trata de buscar una matriz T tal que la nueva matriz de cargas factoriales B tenga muchos valores nulos o casi nulos, y unos pocos valores cercanos a la unidad de acuerdo con el principio de estructura simple descrito anteriormente.

Ejemplo 1 (continuación)

Los factores $F_1^{'}$ y $F_2^{'}$ se han obtenido a partir de los factores F_1 y F_2 aplicando la rotación ortogonal dibujada en la Figura 2. En dicha Figura se representan las variables originales en el espacio factorial definido por los ejes factoriales F_1 y F_2 y en el definido por los ejes $F_1^{'}$ y $F_2^{'}$. En particular, se muestra cuál es la relación existente entre las cargas factoriales de la variable Qca en ambos espacios.

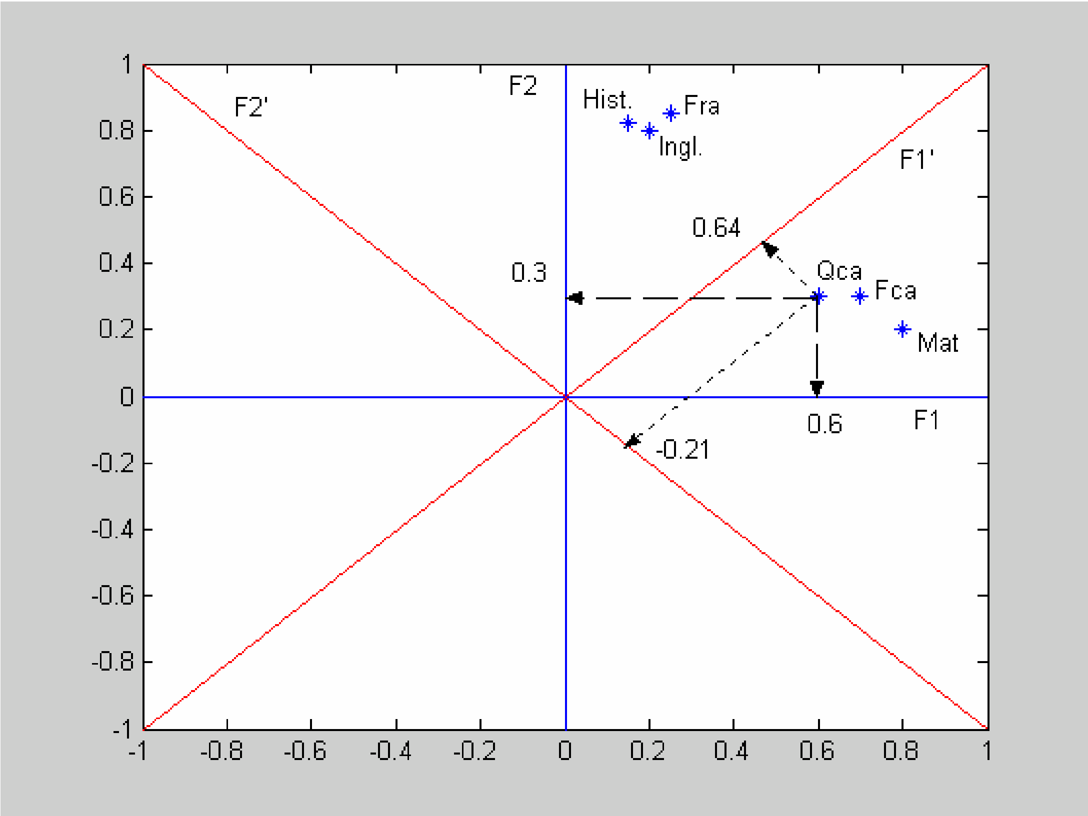


Figura 2: Rotación ortogonal de los factores F_1 y F_2

En este caso la matriz de rotación $\mathbf{T} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$ y la nueva matriz de cargas factoriales será $\mathbf{B} = \begin{pmatrix} 0.71 & -0.42 \\ 0.71 & -0.28 \\ 0.64 & -0.21 \\ 0.71 & 0.42 \\ 0.69 & 0.47 \\ 0.78 & 0.42 \end{pmatrix}$

La forma de calcular estas matrices da lugar a los distintos métodos de rotación ortogonales de los cuáles los más utilizados son los siguientes:

8.1.1- Método Varimax

Se trata de un método de rotación que minimiza el número de variables con cargas altas en un factor, mejorando así la capacidad de interpretación de factores. Este método considera que si se logra aumentar la *varianza de las cargas factoriales al cuadrado de cada factor* consiguiendo que algunas de sus cargas factoriales tiendan a acercarse a uno mientras que otras se acerquen a cero, lo que se obtiene es una pertenencia más clara e inteligible de cada variable a ese factor. Los nuevos ejes se obtienen maximizando la suma para los k factores retenidos de las varianzas de las cargas factoriales al cuadrado dentro de cada factor. Para evitar que las variables con mayores comunalidades tengan más peso en la solución final, suele efectuarse la normalización de Kaiser consistente en dividir cada carga factorial al cuadrado por la comunalidad de la variable correspondiente. En consecuencia, el *método varimax* determina la matriz B de forma que se maximice la suma de las varianzas:

$$V = p\sum_{i=1}^{k}\sum_{j=1}^{p}\left(\frac{b_{ij}}{h_j}\right)^4 - \sum_{i=1}^{k}\left(\sum_{j=1}^{p}\frac{b_{ij}^2}{h_j^2}\right)^2$$

8.1.2- Método Quartimax

El objetivo de este método es que cada variable tenga correlaciones elevadas con un pequeño número de factores. Para ello busca maximizar la *varianza de las cargas factoriales al cuadrado de cada variable* en los factores, es decir, el método trata de maximizar la función:

$$S = \sum_{i=1}^{p}\sum_{j=1}^{k}\left(b_{ij}^2 - \overline{b}_i^2\right)^2 \quad \text{donde } \overline{b}_i^2 = \frac{1}{k}\sum_{j=1}^{k} b_{ij}^2$$

Con ello, se logra que cada variable concentre su pertenencia en un determinado factor, es decir, presente una carga factorial alta mientras que, en los demás factores, sus cargas factoriales tiendan a ser bajas. La interpretación así gana en claridad por cuanto la comunalidad total de cada variable permanece constante, quedando más evidente de este modo hacia qué factor se inclina con más fuerza cada variable. El método es tanto más clarificador cuanto mayor número de factores se hayan calculado.

Este método tiende a producir un primer factor general que se le suele dar el nombre de *tamaño* y el resto de factores presentan ponderaciones menores que las dadas por el método *Varimax*.

8.1.3.- Método Equamax

Este método busca maximizar la media de los criterios anteriores. Suele tener un comportamiento similar a uno de lo dos métodos anteriores.

Ejemplo 2 (continuación)

En la Tabla 7 se muestran los resultados de aplicar una rotación Varimax (los resultados obtenidos al aplicar los otros dos métodos son similares y no se muestran por brevedad).

En primer lugar, se muestra la matriz **B** de cargas factoriales rotadas y, a continuación, la matriz de rotación **T.** Se observa que la interpretabilidad de los factores obtenidos ha mejorado sustancialmente debido a que, en este caso, cada variable tiende a relacionarse con un solo factor.

Se observa que el factor 1 rotado está muy relacionado con los ratios que reflejan el nivel de endeudamiento a largo plazo de la empresa estando relacionado directamente con aquellos ratios que tienen dicha partida en el numerador (DLP/AT y DLP/KP) e inversamente con los que la tienen en el denomiador (PC/DT y CF/DLP). Es, por lo tanto, un factor que refleja el ENDEUDAMIENTO A LARGO PLAZO.

El factor 2 por su parte, está relacionado directamente con ratios que reflejan el nivel de beneficios de la empresa (BN/RP, CF/RP y BAIT/RP) por lo que puede interpretarse como un factor de RENTABILIDAD.

Finalmente, el factor 3 está relacionado con ratios que reflejan el peso de los recursos ajenos en la gestión de las empresas (DT/RP y AT/RP) por lo que podría interpretarse como un factor de RECURSOS AJENOS.

Tabla 7
Matriz de cargas factoriales rotada y matriz de rotación

Matriz de factores rotados[a]

	Factor		
	1	2	3
DLP/AT	.967		
PC/DT	-.953		
DLP/KP	.867		
CF/DLP	-.662		
BN/RP		.956	
CF/RP		.774	
BAIT/RP		.753	
DT/RP			.939
AT/RP			.916

Método de extracción: Mínimos cuadrados no ponderados.

Método de rotación: Normalización Varimax con Kaiser.

a. La rotación ha convergido en 4 iteraciones.

Matriz de transformación de los factores

Factor	1	2	3
1	.792	-.319	.520
2	.264	.948	.179
3	-.550	-.005	.835

Método de extracción: Mínimos cuadrados no ponderados.

Método de rotación: Normalización Varimax con Kaiser.

8.2.- Rotación oblicua

Se diferencia de la rotación ortogonal en que a la matriz **T** de rotación no se les exige ser ortogonal sino únicamente no singular. De esta forma los factores rotados no tienen por qué ser ortogonales y tener, por lo tanto, correlaciones distintas de cero entre sí.

La rotación oblicua puede utilizarse cuando es probable que los factores en la población tengan una correlación muy fuerte. Insistimos en que hay que ir con mucho cuidado en la interpretación de las rotaciones oblicuas, ya que la superposición de factores puede confundir la significación de los mismos. De esta forma el análisis gana más flexibilidad y realismo pero a riesgo de perder robustez por lo que conviene aplicar estos métodos si el número de observaciones por factor es elevada.

Ejemplo 1 (continuación)

Si definimos $F_1^{''} = \frac{4}{\sqrt{17}} F_1 + \frac{1}{\sqrt{17}} F_2$ y $F_2^{''} = \frac{1}{\sqrt{17}} F_1 + \frac{4}{\sqrt{17}} F_2$ entonces $\mathrm{Corr}\left(F_1^{''}, F_2^{''}\right) = \frac{8}{17}$ = 0.47 ≠ 0 por lo que los nuevos factores estarán correlacionados. En la Figura 3 se muestra la rotación geométrica asociada a dicha transformación que viene determinada por las variables Mat e Ing que tienen una carga factorial 0 en los nuevos factores.

En este caso se tiene que la matriz de rotación vendrá dada por **T** = $\begin{pmatrix} \frac{4\sqrt{17}}{\sqrt{15}} & -\frac{\sqrt{17}}{\sqrt{15}} \\ -\frac{\sqrt{17}}{\sqrt{15}} & \frac{4\sqrt{17}}{\sqrt{15}} \end{pmatrix}$ y la matriz de cargas factoriales vendrá dada por B = $\begin{pmatrix} 0.82 & 0.00 \\ 0.69 & 0.14 \\ 0.58 & 0.17 \\ 0.00 & 0.82 \\ -0.06 & 0.86 \\ 0.04 & 0.87 \end{pmatrix}$. Dicha matriz recibe el nombre de **matriz de configuración.**

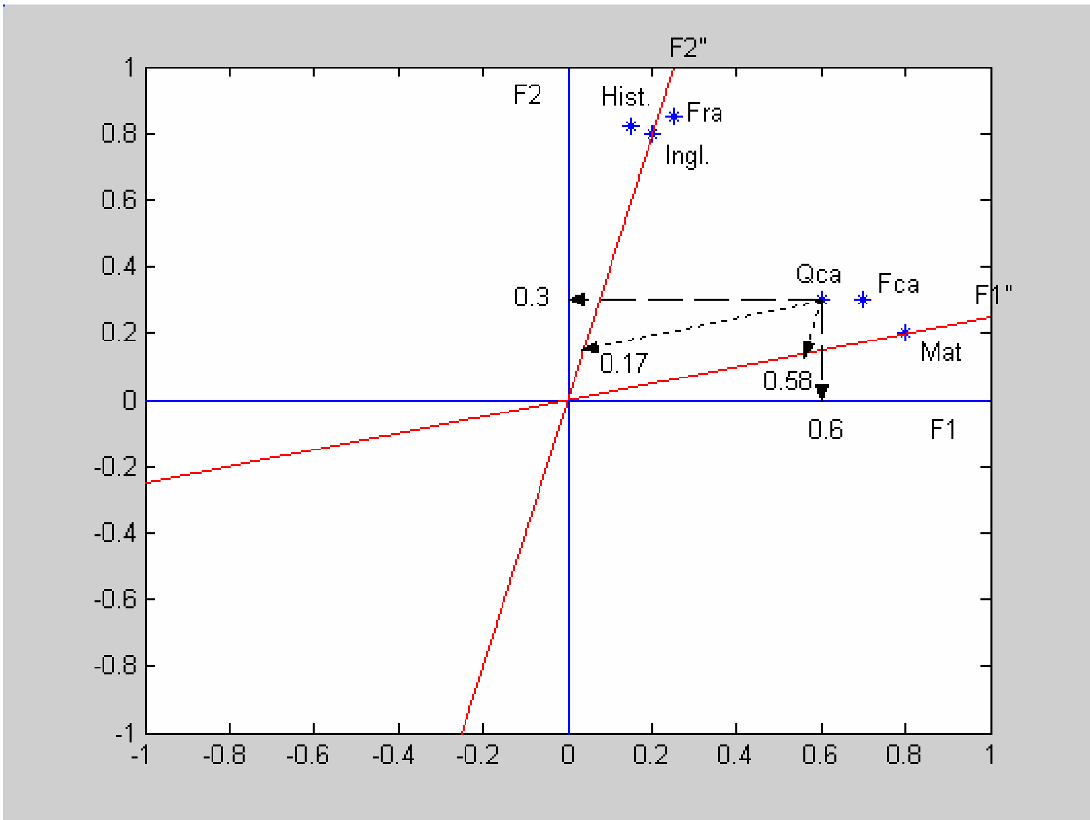


Figura 3: Rotación oblicua de los factores F_1 y F_2

La matriz que contiene las correlaciones de las variables originales con los nuevos factores recibe el nombre de **matriz de estructura** y vendrá dada en este caso por:

$$\mathbf{B}\begin{pmatrix} 1 & 8/17 \\ 8/17 & 1 \end{pmatrix} = \begin{pmatrix} 0.82 & 0.39 \\ 0.76 & 0.46 \\ 0.66 & 0.44 \\ 0.39 & 0.82 \\ 0.34 & 0.83 \\ 0.45 & 0.89 \end{pmatrix}$$

Se observa que $F_1^{"}$ puede interpretarse, de nuevo, como un factor de aptitud científica mientras que $F_2^{"}$ como un factor de aptitud verbal pero, a diferencia de los factores F_1 y F_2 que son incorrelados, en esta nueva estimación ambos factores tienen una correlación positiva significativa lo cual proporciona más realismo al análisis realizado.

8.2.1.- Método Oblimin

Busca minimizar la siguiente expresión:

$$\sum_{s<q=1}^{k}\left[\alpha\sum_{i=1}^{p} b_{is}^{2} b_{iq}^{2} + (1-\alpha)\sum_{i=1}^{p}\left(b_{is}^{2} - \overline{b}_{s}^{2}\right)\left(b_{is}^{2} - \overline{b}_{s}^{2}\right)\right]$$

donde:

$\sum_{s<q=1}^{k}\sum_{i=1}^{p} b_{is}^{2} b_{iq}^{2}$ controla la interpretabilidad de los factores

$\sum_{s<q=1}^{k}\sum_{i=1}^{p}\left(b_{is}^{2} - \overline{b}_{s}^{2}\right)\left(b_{is}^{2} - \overline{b}_{s}^{2}\right)$ controla la ortogonalidad de los factores

Para un valor $\alpha = 1$ se alcanza el máximo grado de oblicuidad y cuanto más cerca de 0 toma sus valores, más ortogonales son los factores.

Como se ha visto en la sección anterior, en la rotación oblicua las cargas factoriales no coinciden con las correlaciones entre el factor y la variable, puesto que los factores están correlacionados entre sí.

Por eso, los paquetes estadísticos calculan dos matrices: la *matriz de cargas factoriales* que muestra la contribución única de cada variable al factor y la *matriz de estructura factorial* que muestra las correlaciones entre los factores y las variables y que contiene información acerca de la contribución única y de las correlaciones entre factores. Además de estas dos matrices, es interesante analizar también la matriz de correlaciones entre factores. Si las correlaciones entre los factores son muy pequeñas es más robusto aplicar rotaciones ortogonales; por otro lado si dos factores están muy correlacionados puede ser señal de que estén midiendo el mismo concepto y que, en consecuencia, haya que reducir el número de factores.

8.2.2. Método Promax

Consiste en alterar los resultados de una rotación ortogonal hasta crear una solución con cargas factoriales lo más próximas a la estructura ideal. Dicha estructura se supone que se obtiene elevando las cargas factoriales obtenidas en una rotación ortogonal, a una potencia que suele estar entre 2 y 4. Cuanto mayor es esta potencia más oblicua es la solución obtenida.

Si H es la matriz de cargas buscada el método promax busca una matriz T tal que $\mathbf{AT = H}$. Multipiplicando ambos miembros por la matriz $\mathbf{(A'A)^{-1}A'}$ se tiene que $\mathbf{T} = \mathbf{(A'A)^{-1}A'H}$.

Ejemplo 2 (continuación)

Con el fin de analizar hasta qué punto los factores extraídos con el método Varimax son ortogonales realizamos, a continuación, una rotación oblicua. En la Tabla 8 se muestran los resultados obtenidos al aplicar el método oblimin con $\alpha = 1$ que es el que permite el mayor grado de oblicuidad (los resultados obtenidos con el método promax son similares y no se muestran por brevedad)

En este caso se muestra la matriz de cargas factoriales rotada **B** que recibe el nombre de *matriz de configuración* así como *la matriz de estructura*, que contiene las correlaciones entre los factores y las variables originales. Finalmente, también se muestra la matriz de correlaciones entre factores.

Se observa que la interpretación dada a los factores se mantiene en lineas generales pero que dichos factores no son ortogonales, existiendo una correlación significativa positiva entre el factor de Endeudamiento y el de Recursos Ajenos igual a 0.375. Dicha correlación se debe a la presencia de una relación directa de los ratios DLP/KP y DT/RP con ambos factores.

Tabla 8
Resultados obtenidos al realizar una rotación oblícua

Matriz de configuración[a]

	Factor		
	1	2	3
PC/DT	-1.012		
DLP/AT	.986		
DLP/KP	.830		
CF/DLP	-.635		
BN/RP		.958	
BAIT/RP		.766	
CF/RP		.766	
AT/RP			.941
DT/RP			.923

Método de extracción: Mínimos cuadrados no ponderados.

Metodo de rotación: Normalización Oblimin con Kaiser.

a. La rotación ha convergido en 5 iteraciones.

Matriz de estructura

	Factor		
	1	2	3
DLP/AT	.991		
PC/DT	-.941		
DLP/KP	.935		.610
CF/DLP	-.717		
BN/RP		.969	
CF/RP		.790	
BAIT/RP		.745	
DT/RP	.533		.992
AT/RP			.928

Método de extracción: Mínimos cuadrados no ponderados.

Metodo de rotación: Normalización Oblimin con Kaiser.

Matriz de correlaciones entre los factores

Factor	1	2	3
1	1.000	-.207	.375
2	-.207	1.000	-.103
3	.375	-.103	1.000

Método de extracción: Mínimos cuadrados no ponderados.

Metodo de rotación: Normalización Oblimin con Kaiser.

9.- CÁLCULO DE PUNTUACIONES FACTORIALES

Una vez determinados los factores rotados el siguiente paso es calcular las *matriz de puntuaciones factoriales* **F**. Las posibilidades de analizar las puntuaciones factoriales de los sujetos son muy variadas según lo que se pretenda:

- conocer qué sujetos son los más raros o extremos, es decir, la representación gráfica de las puntuaciones factoriales para cada par de ejes factoriales puede ayudar a detectar casos atípicos;
- conocer dónde se ubican ciertos grupos o subcolectivos de la muestra (los jóvenes frente a los mayores, los de clase alta frente a los de baja, los más católicos frente a los no católicos, los de una provincia frente a los de otras provincias, etc);
- conocer en qué factor sobresalen unos sujetos y en qué factor no, etc.
- explicar, analizando las informaciones anteriores, por qué han aparecido dichos factores en el análisis realizado

El Análisis Factorial es en otras ocasiones un paso previo a otros análisis, como por ejemplo, Regresión Múltiple o Análisis Cluster, en los que se sustituye el conjunto de variables originales por los factores obtenidos. Por ello, es necesario conocer los valores que toman los factores en cada observación.

9.1. Métodos de cálculo de las puntuaciones

Existen diversos métodos de estimación de la matriz **F.** Las propiedades que sería deseable cumpliesen los factores estimados son:

- cada factor estimado tenga correlación alta con el verdadero factor.
- cada factor estimado tenga correlación nula con los demás factores verdaderos.
- los factores estimados sean incorrelacionados dos a dos, es decir, mutuamente ortogonales si son ortogonales
- los factores estimados sean estimadores insesgados de los verdaderos factores.

Sin embargo, por la propia naturaleza de los factores comunes, el problema de su estimación es complejo. Se puede demostrar que los factores no son, en general, combinación lineal de las variables originales. Además, en la mayoría de las situaciones, no existirá una solución exacta ni siquiera será única.

Todos los métodos de obtención de puntaciones factoriales parten de la expresión:

$$\mathbf{X} = \mathbf{F}A' + \mathbf{U} \text{ con } \mathbf{E[U]=0, Var[U]} = \Psi$$

a partir de la cual buscan estimar el valor de **F.**

Tres de los métodos de estimación más utilizados son los siguientes:

9.1.1. Método de regresión

Estima **F** mediante el método de los mínimos cuadrados

$$\hat{\mathbf{F}} = (A'A)^{-1}A'\mathbf{X}$$

9.1.2. Método de Barlett

Utiliza el método de los mínimos cuadrados generalizados estimando las puntuaciones factoriales mediante:

$$\hat{\mathbf{F}} = (A'\,\Psi^{-1}\,A)^{-1}A'\Psi^{-1}\mathbf{X}$$

9.1.3 Método de Anderson-Rubin

Estima **F** mediante el método de los mínimos cuadrados generalizados pero imponiendo la condición adicional **F'F = I**

$$\hat{\mathbf{F}} = (A'\,\boldsymbol{\Psi}^{-1}\mathbf{R}\boldsymbol{\Psi}^{-1}A)^{-1}A'\boldsymbol{\Psi}^{-1}\mathbf{X}$$

9.1.4. Comparación de los tres métodos

1) El método de regresión da lugar a puntuaciones con máxima correlación con las puntuaciones teóricas. Sin embargo, el estimador no es insesgado, ni unívoco y, en el caso de que los factores sean ortogonales, puede dar lugar a puntuaciones correladas.
2) El método de Bartlett da lugar a puntuaciones correladas con las puntuaciones teóricas, insesgadas y unívocas. Sin embargo, en el caso de que los factores sean ortogonales, puede dar lugar a puntuaciones correladas.
3) El método de Anderson-Rubin da lugar a puntuaciones ortogonales que están correladas con las puntuaciones teóricas. Sin embargo, el estimador no es insesgado ni es unívoco.

Ejemplo 2 (continuación)

En la Figura 4 se muestran, en forma de diagramas de líneas, la evolución anual de las puntuaciones factoriales calculadas para cada uno de los países. Dichas puntuaciones han sido calculadas utilizando el método de Bartlett que es el más aconsejable en este caso dado la oblicuidad observada entre algunos de los factores extraídos.

Se observa que, con respecto al factor de ENDEUDAMIENTO A LARGO PLAZO los mayores niveles de endeudamiento a largo plazo han correspondido a las empresas holandesas y finlandesas con un claro descenso, en éste último caso, a partir de 1992. Por el contrario, los menores niveles de endeudamiento de este tipo corresponden a las empresas alemanas y austriacas con niveles bastante estables a lo largo del tiempo.

El factor de RENTABILIDAD refleja, por su parte, cuál ha sido la evolución de la coyuntura económica en cada país con niveles altos y bajos en todos ellos a lo largo del tiempo. Así, por ejemplo, en el caso de España se observa claramente una caida de la rentabilidad en 1985 y en 1993 periodos de crisis en la economía española, y una recuperación muy clara entre 1986 y 1991, por un lado, y a partir de 1994 por el otro.

Finalmente, el factor de RECURSOS AJENOS muestra la mayor tendencia que tienen las empresas finlandesas y holandesas a recurrir a ellos, frente a la postura de las empresas alemanas, portuguesas, belgas y austriacas que utilizan este tipo de financiación con menor intensidad. En el caso de las empresas españolas se aprecia que el peso de este tipo de recursos aumenta en las épocas de crisis y disminuye en las épocas de bonanza económica.

Finalmente, en la Figura 5 se muestra el diagrama de puntos de las puntuaciones factoriales de los factores de Endeudamiento a Largo Plazo y Recursos Ajenos en la que se aprecia la relación positiva existente entre ellos.

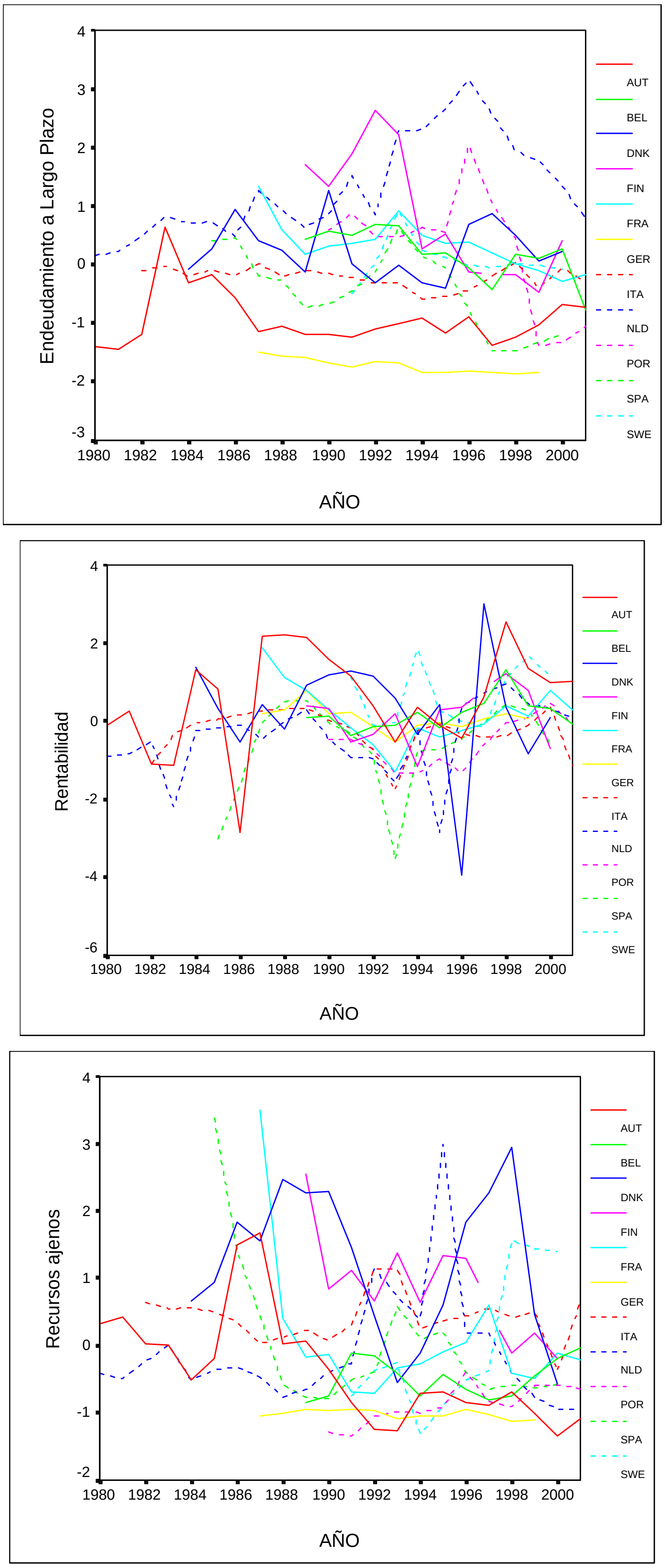


Figura 4: Evolución de las puntuaciones factoriales

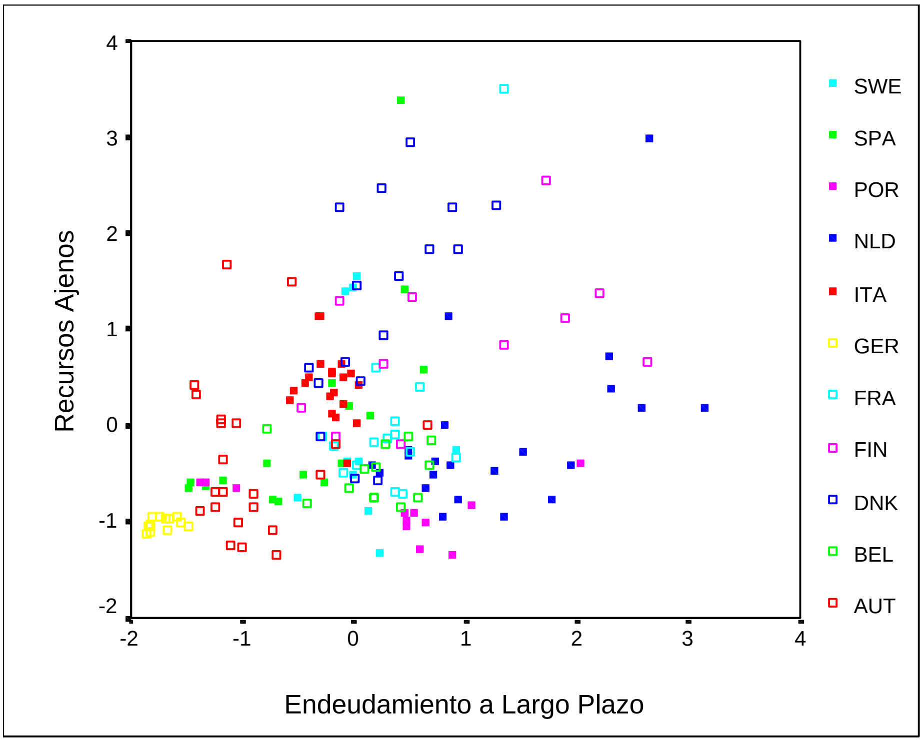


Figura 5: Endeudamiento a Largo Plazo vs Recursos Ajenos

9.2. Selección de Variables

En algunas ocasiones, el investigador desea seleccionar las variables más representativas de los factores, en lugar de calcular sus puntuaciones. Así, por ejemplo, si se utiliza el Análisis Factorial para reducir el número de datos por razones de economía es más interesante, si se quieren aplicar los resultados obtenidos a objetos diferentes de los estudiados en el análisis, seleccionar algunas de las variables originalmente medidas dada la dificultad de cálculo de las puntuaciones factoriales para las que se necesitaría medir todas las variables utilizadas en el estudio.

Una manera de llevar a cabo dicha selección es estudiar la matriz de correlaciones de las variables con los factores, seleccionando como representante de cada factor la variable con la correlación más elevada en éste, que sea más fácil de medir y que tenga más sentido desde un punto de vista teórico.

En cualquier caso conviene elegirlas de forma que una misma variable no se utilice para medir dos factores distintos. Una vez elegidas se les asigna pesos basados en su correlación con el factor, y se comprueba su validez estimando su correlación con los factores que quiere estimar mediante la fórmula R_{fs} = A'Wdiag(R_{ss}) donde R_{ss} es la matriz de correlaciones de las puntuaciones estimadas

10.- VALIDACIÓN DEL MODELO

El último paso en el Análisis Factorial es estudiar la *validez del modelo*. Dicha validación debe hacerse en dos direcciones: analizando la bondad de ajuste del mismo y la generabilidad de sus conclusiones.

10.1 Bondad de ajuste

Una suposición básica subyacente al Análisis Factorial es que la correlación observada entre las variables puede atribuirse a factores comunes. Por consiguiente, las correlaciones entre variables pueden deducirse o reproducirse a partir de las correlaciones estimadas entre las variables y los factores. A fin de determinar el ajuste del modelo, pueden estudiarse las diferencias entre las correlaciones observadas (como se dan en la matriz de correlación de entrada) y las correlaciones reproducidas (como se estiman a partir de la matriz factorial). Estas diferencias se conocen como *residuos*. Si el modelo factorial es adecuado entonces estos residuos deben ser pequeños. Si existe un porcentaje elevado de residuos superiores a una cantidad pequeña prefijada (por ejemplo, 0.05), esto será indicativo de que el modelo factorial estimado no se ajusta a los datos. Se sabe además que hay más estabilidad en los resultados si el número de casos por variable es alto.

10.2. Generalidad de los resultados

También es adecuado refrendar los resultados del primer análisis factorial realizando nuevos análisis factoriales sobre nuevas muestras extraídas de la población objeto de estudio y, caso de no ser posible esto último, sobre submuestras de la muestra original. En cada caso se estudiaría qué factores de los calculados son replicados en los distintos análisis llevados a cabo. Otra posibilidad es realizar nuevos análisis factoriales modificando las variables consideradas bien sea eliminando aquellas variables que no tienen relación con ningún factor o eliminando las variables con relaciones más fuertes tratando de descubrir cómo se comporta el resto de ellas sin su presencia.

Otro de los procedimientos metodológicos y estadísticos que complementan y profundizan las interpretaciones que se deducen del análisis factorial consiste en la realización de otros análisis factoriales en base, no al conjunto total de la muestra o población, sino referido a subcolectivos o grupos que están presentes en esa muestra y que pueden formarse utilizando las categorías de las variables primarias: sexo, clase social, tipo de centro, tipo de metodología pedagógica, tipos sociológicos, tipos de actitud, etc. Lo que se desprende de los trabajos e investigaciones que han utilizado este procedimiento es que normalmente la interpretación que se da y que es válida para el conjunto total de sujetos debe modificarse, en algunos casos sustancialmente, cuando se refiere a esos subcolectivos. Caso de ser así, la conclusión que se deriva es doble: por una parte, las variables se comportan en el Análisis Factorial de distinta forma según de qué muestra se trate y, por otra, que no existe el sujeto «tipo» sino que existen diferentes «tipos» de sujetos en la muestra global.

Finalmente se debería plantear un Análisis Factorial Confirmatorio para comprobar los resultados obtenidos en la versión exploratoria.

Ejemplo 2 (continuación)

En este caso habría que tratar de encontrar el sentido de los factores hallados utilizando otro tipo de ratios financieros así como analizar más a fondo la idiosincrasia de cada país para tratar de hallar explicaciones a los movimientos detectados a lo largo del tiempo.

Resumen

El Análisis Factorial es una técnica estadística multivariante cuya finalidad es analizar las relaciones de interdependencia existentes entre un conjunto de variables, calculando un conjunto de variables latentes, denominadas factores, que explican con un número menor de dimensiones, dichas relaciones. Por esta razón el Análisis Factorial es una técnica de reducción de datos que permite expresar la información contenida en un conjunto de datos con un número menor de variables sin distorsionar dicha información, lo cual aumenta el grado de manejabilidad e inteligibilidad de la misma.

En esta lección hemos estudiado qué es y cómo llevar a cabo un Análisis Factorial sobre un conjunto de datos cuantitativos, ilustrando los pasos a seguir con el análisis de un caso real.

Conviene insistir, finalmente, en la importancia que tiene realizar un buen planteamiento del problema tanto en la selección de las variables a analizar como en la de los objetos sobre las que deben ser medidos. Es muy conveniente tener un conocimiento previo de qué factores queremos medir y elegir las variables de acuerdo con los mismos. Actuando de esta manera el análisis gana en potencia y generalidad aumentando significativamente el grado de inteligibilidad de los resultados obtenidos.

Bibliografía

Un buen libro sobre Análisis Factorial que sigue de vigente actualidad a pesar del tiempo transcurrido desde su publicación es:

GORSUCH, R. (1983). *Factor Analysis. Second Edition.* LEA

que es, en nuestra opinión, una de las "biblias" del Análisis Factorial.

Desde un punto de vista más práctico recomendamos

AFIFI, A.A. and CLARK, V. (1996) *Computer-Aided Multivariate Analysis.* Third Edition. Texts in Statistical Science. Chapman and Hall.

EVERITT, B. And GRAHAM, D. (1991). *Applied Multivariate Data Analysis.* Arnold.

HAIR, J., ANDERSON, R., TATHAM, R. y BLACK, W. (1999). *Análisis Multivariante.* 5ª Edición. Prentice Hall.

SHARMA, S. (1998). *Applied Multivariate Techiques.* John Wiley and Sons

que contienen excelentes capítulos dedicados al tema.

Una revisión breve pero completa del Análisis Factorial Exploratorio muy orientada al SPSS puede encontrarse en

GARCIA JIMÉNEZ, E.; GIL FLORES, J. y RODRIGUEZ GOMEZ, G. (2000). *Análisis Factorial.* . Cuadernos de Estadística. Editorial La Muralla.

También contienen un capítulo dedicado al Análisis Factorial con SPSS los libros

PEREZ, César (2001). *Técnicas Estadísticas con SPSS.* Prentice-Hall

VISAUTA, B. (1998) *Análisis Estadístico con SPSS para WINDOWS (Vol II. Análisis Multivariante).* Mc-Graw Hill

Un enfoque más riguroso y matemático del tema puede encontrarse en:

JOBSON, J.D. (1992) *Applied Multivariate Data Analysis. Volume II: Categorical and Multivariate Methods.* Springer-Verlag.

MARDIA, K.V., KENT, J.T. y BIBBY, J.M. (1994). *Multivariate Analysis.* Academic Press.

SEBER, G.A.F. (1984). *Multivariate Observations*. John Wiley & Sons.

Finalmente si se está interesado en el Análisis Factorial Confirmatorio y su relación los modelos de las ecuaciones estructurales aconsejamos el libro:

KLINE, R.B. (1998). *Principles and Practice of Structural Equation Modeling.* The Guilford Press.