

# Índice general

<b>1. Análisis Factorial</b>	<b>2</b>
1.1. Introducción: El modelo teórico básico de Análisis Factorial. Generalidades . . . . .	2
1.1.1. Modelo de Análisis Factorial no-restringido y aleatorio . . . . .	2
1.1.2. Algunos resultados sobre el Modelo . . . . .	3
1.1.3. Análisis Factorial, oblicuo, ortogonal y normal (Lawley) . . . . .	5
1.2. Problemas básicos inherentes al modelo de Análisis Factorial . . . . .	6
1.2.1. El problema de la identificación en el Análisis Factorial: existencia y unicidad de solución . . . . .	6
1.2.2. Un problema de “ambigüedad” en el Análisis Factorial. Rotación de factores e interpretación en el Análisis Factorial . . . . .	9
1.2.3. Cambio de unidades de medida . . . . .	10
1.3. Métodos de estimación en el Análisis Factorial . . . . .	11
1.3.1. Introducción . . . . .	12
1.3.2. El método de las componentes principales . . . . .	12
1.3.3. La “Principal Factor Solution” como modificación del método de las Componentes Principales en estimación en Análisis Factorial . . . . .	16
1.3.4. El método de Máxima Verosimilitud en el Modelo de Análisis Factorial (Modelo de Lawley-Maxwell) . . . . .	17
1.3.5. Ejemplo Stock-prices . . . . .	20
1.4. La rotación de factores en el Análisis Factorial. Criterios gráficos y analíticos (Kaiser-Varimax) . . . . .	23
1.4.1. Métodos gráficos en la rotación de factores . . . . .	24
1.4.2. Métodos analíticos: el criterio de Kaiser (Criterio Varimax) . . . . .	25
1.5. Rotación de factores. Criterios analíticos generales . . . . .	26
1.5.1. Métodos de rotación ortogonal . . . . .	27
1.5.2. Ejemplo Stock-prices . . . . .	28
1.6. Puntuaciones en los factores (Factor Scores) . . . . .	29
1.6.1. El método de mínimos cuadrados ponderados . . . . .	29
1.6.2. El método de regresión . . . . .	31
1.6.3. Comparación de los “factor scores” dados por los dos métodos . . . . .	32
1.6.4. Los “scatterplots” bivariantes de los factor scores . . . . .	33
1.6.5. Ejemplo Stock-prices . . . . .	34
1.7. Contrastos de Hipótesis en Análisis Factorial . . . . .	36
1.7.1. Test asintótico (Bartlett) sobre la adecuación de un modelo factorial ortogonal con m factores comunes . . . . .	36
1.7.2. Ejemplo Stock-prices . . . . .	38
1.7.3. Criterios basados en la información estadística. El criterio de Akaike (AIC) para la elección de un modelo factorial con r-factores comunes . . . . .	38
<b>2. Aplicación mediante R</b>	<b>41</b>
2.1. Paquete Factanal . . . . .	41
2.2. Paquete prcomp . . . . .	42
2.3. Ejemplo de aplicación . . . . .	43

# Tema 1

## Análisis Factorial

### 1.1. Introducción: El modelo teórico básico de Análisis Factorial. Generalidades.

En primer lugar vamos a definir el modelo teórico básico que vamos a considerar a lo largo del desarrollo del tema. Hay muchos tipos de modelo de Análisis Factorial, según las hipótesis adoptados sobre los elementos con los que se define el Modelo u otras circunstancias, como el tipo de datos (continuos, discretos, incorrelados o correlados). El modelo inicial del que partiremos es un Modelo **no-restringido** (“unrestricted”) y **aleatorio**. El significado de estos términos, como el de otros como Análisis Factorial *oblicuo* u *ortogonal*; Análisis Factorial *exploratorio* o *confirmatorio*; Análisis Factorial *normal*, etc., irá apareciendo secuencialmente a la medida que se vaya desarrollando el tema.

#### 1.1.1. Modelo de Análisis Factorial no-restringido y aleatorio

##### i) Elementos del Modelo:

- $X$  el vector ( $p \times 1$ ) de variables  $X_i$   $i = 1, \dots, p$ , de la **población**.
- $F$  designa un vector ( $m \times 1$ ) constituido por las variables latentes o “factores comunes”,  $F_i$ ,  $i = 1, \dots, m$ ,  $m < p$ .  $F$  es el **vector de factores comunes**.
- $\epsilon$ , vector ( $p \times 1$ ), con componentes  $\epsilon_i$ ,  $i = 1, \dots, p$ , que designa los “errores” asociados a cada variable  $X_i$ . Es el llamado **“vector de errores”** o “vector de factores específicos”  $e_i$ .
- Una matriz  $L$ , ( $p \times m$ ), de rango  $m$ , llamada **“matriz de factor loadings”**. Sus elementos,  $l_{ij}$ , son los “loadings” de la variable  $X_i$  respecto del factor común  $F_j$ .

##### ii) Características de los elementos del Modelo.

Los vectores  $X$ ,  $F$  y  $\epsilon$  son **aleatorios** y la matriz  $L$  es **no aleatoria** (modelo de Análisis Factorial aleatorio). Hay modelos de Análisis Factorial que no asignan a  $F$  un carácter aleatorio, interpretando el vector  $F$  como un vector de componentes  $F_i$  no aleatorias, de meros parámetros que cambian “individuo a individuo”. Esta consideración propia especialmente del Análisis Factorial en la Sicometría, hace no-aleatorio al correspondiente Modelo de Análisis Factorial. Es posible adaptar a este caso los principales resultados del correspondiente modelo aleatorio (véase, por ejemplo, Basilewsky (1994), sección 6.8.2).

Por otra parte no se impone ninguna restricción a los elementos de  $L$ . Por ejemplo no se asigna valor 0 a determinados factor-loadings  $l_{ij}$ . En este sentido se dice que el modelo es **no restringido** (“unrestricted”). Más adelante introduciremos los Análisis Factorial **exploratorio** y **confirmatorio** lo que está en relación con el carácter no-restringido o restringido antes formulado.

Finalmente,  $X$  es **observable** y  $F$  y  $\epsilon$  son no **observables**.

iii) **Hipótesis probabilísticas sobre los elementos del Modelo.**

Supondremos que los elementos aleatorios antes citados, verifican:

- $E[X] = \mu$  vector media poblacional ( $p \times 1$ ).
- $Cov[X] = \Sigma$ ,  $p \times p$ , matriz definida positiva de covarianzas de la población.
- $E[F] = 0$ ;  $E[\epsilon] = 0$ ;  $Cov[F] = E[FF'] = \Phi$ ,  $m \times m$ .
- $Cov[\epsilon] = \Psi$ ,  $p \times p$  con  $\Psi = diag(\psi_1, \dots, \psi_p)$ , siendo  $\psi_i, i = 1, \dots, p$ , las varianzas específicas o **unicidades**.
- $Cov[F; \epsilon] = E[F\epsilon'] = 0$ .

Es decir, que los vectores  $F$  y  $\epsilon$  son incorrelados, centrados y tienen, respectivamente, matrices de covarianzas  $\Phi$  y  $\Psi$ , ésta última diagonal, lo que implica que los errores  $\epsilon_i$  son heterocedásticos e incorrelados. En principio,  $\Phi = Cov(F)$  es una matriz no necesariamente diagonal, es decir, que en general, los  $F_i$  son correlados.

iv) Finalmente, los elementos  $X$ ,  $F$ ,  $\epsilon$  y  $L$  con las características e hipótesis probabilísticas antes indicadas se integran en un **modelo** lineal del tipo

$$X = \mu + LF + \epsilon$$

modelo que es el adoptado en este Tema.

**Nota 1.** *El modelo de Análisis Factorial es pues un conjunto de relaciones lineales del tipo:*

$$\begin{cases} X_1 = l_{11}F_1 + l_{12}F_2 + \dots + l_{1m}F_m + \epsilon_1 \\ \vdots \\ X_p = l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pm}F_m + \epsilon_p \end{cases}$$

en el que los factores comunes ( $F_1, \dots, F_m$ ) explican conjuntamente una parte de todas las variables  $X_i, i = 1, \dots, p$ , y la parte no explicada de cada  $X_i$  mediante los factores es explicada por los  $\epsilon_i$ , que están asociados cada  $\epsilon_i$  a la respectiva  $X_i$ . Los  $\epsilon_i$  son incorrelados e inobservables y los  $F_j$ , que son aleatorios y no observables, son correlados entre si, en general.

Formalmente, el modelo de Análisis Factorial como conjunto de ecuaciones lineales antes indicado podría ser considerado, por ejemplo, como un modelo de regresión lineal múltiple multivariante (multiple por haber  $m$  "regresores"  $F_j$ ,  $j = 1, \dots, m$ ; multivariante al haber  $p$  variables  $X_i$ ,  $i = 1, \dots, p$  "dependientes"). Pero hay una diferencia sustancial: los  $F_j$  son aleatorios y no observables, lo que diferencia de raíz ambos tipos de modelos (Análisis Factorial y Regresión Lineal). •

### 1.1.2. Algunos resultados sobre el Modelo

■ **Ecuación estructural de  $\Sigma$  en el Modelo de Análisis Factorial**

$$\begin{aligned} E[(X - \mu)(X - \mu)'] &= \Sigma = E[(LF + \epsilon)(LF + \epsilon)'] = E[LFF'L' + LF\epsilon' + \epsilon F'L' + \epsilon\epsilon'] = \\ LE[FF']L + LE[F\epsilon'] + E[\epsilon F']L' + E[\epsilon\epsilon'] &= \text{dadas las condiciones de i)-iv)} = \end{aligned}$$

$$L\Phi L' + \Psi$$

Así pues,  $\Sigma = L\Phi L' + \Psi$ , es decir, que el modelo estadístico Factorial supuesto, estructura la matriz de covarianzas de la población,  $\Sigma$ , en términos del triple  $(L; \Phi; \Psi)$ , mediante dicha ecuación.

Por tanto, técnicamente hablando, modelizar la población  $X = (X_1, \dots, X_p)$ , mediante en Análisis Factorial, es equivalente a encontrar las matrices  $L$ ,  $\Phi$  y  $\Psi$  con las condiciones exigidas en el Modelo, de tal forma que la  $\Sigma$  poblacional se “explica” mediante dicha ecuación.

En este sentido cabe clarificar al Análisis Factorial como técnica estadística multivariante del tipo “modelos de estructura de covarianzas” tipo integrado en la literatura estadística actual por técnicas como Análisis de Componentes Principales; modelos lineales estructurales (LISREL) y otras más específicas.

- Por otra parte la matriz  $(p \times m)$  de covarianzas entre la población y los factores es:

$$\text{Cov}[X; F] = E[(X - \mu)F'] = E[(LF + \epsilon)F'] = E[LFF'] + E[\epsilon F'] =$$

$$\text{por las condiciones i)-iv)} = LE[FF'] = L\Phi$$

- También se verifica lo siguiente. Supongamos que en efecto  $\Sigma$  se estructura en la forma:

$$\Sigma = L\Phi L' + \Psi$$

mediante el triple  $(L; \Phi; \Psi)$ . Es claro que si  $C$  es una matriz cualquiera  $m \times m$ , no singular, es cierto que el triple  $(LC; C^{-1}\Phi C^{-1}')$ , también estructura  $\Sigma$  de la manera equivalente. En efecto:

$$\Sigma = L\Phi L' + \Psi = LCC^{-1}\Phi C^{-1}'C'L' + \Psi = (LC)(C^{-1}\Phi C^{-1}')(C'L') + \Psi =$$

$$(LC)(C^{-1}\Phi C^{-1}')(LC)' + \Psi = L^*\Phi^*L^{*\prime} + \Psi$$

Además,  $LC$  y  $C^{-1}\Phi C^{-1}'$  cumplen todas las condiciones exigidas a  $L$  y  $\Phi$ . Por ejemplo:  $(LC)$  es  $p \times m$ ;  $C^{-1}\Phi C^{-1}'$  es  $m \times m$  y es la matriz de covarianzas de  $F^* = C^{-1}F$ ,  $m \times m$ .

En efecto,  $E[C^{-1}F] = C^{-1}E[F] = 0$ ; y  $\text{Cov}[C^{-1}F] = E[C^{-1}FF'C^{-1}'] = C^{-1}E[FF']C^{-1}' = C^{-1}\Phi C^{-1}'$

Lo anterior es equivalente a decir, en términos de la ecuación del Modelo que en lugar de tomar el Modelo

$$X - \mu = LF + \epsilon; \quad (L; F; \epsilon)$$

podemos siempre considerar uno equivalente

$$X - \mu = LCC^{-1}F + \epsilon = (LC)(C^{-1}F) + \epsilon$$

y los “nuevos factor-loadings”  $(LC)$  y los “nuevos factores”  $(C^{-1}F)$ , factorizan a  $\Sigma$  de manera equivalente.

De hecho pues, si existe un modelo factorial con  $(L; F)$  cualquier otro definido en la forma  $(L^* = LC; F^* = C^{-1}F; \epsilon)$  estructura  $\Sigma$  de manera equivalente, es decir,  $(L^*; F^*; \epsilon)$  también sería una solución equivalente.

### 1.1.3. Análisis Factorial, oblicuo, ortogonal y normal (Lawley)

El modelo de Análisis Factorial antes definido, tiene dos casos particulares especialmente importantes. En primer lugar, consideramos la situación particular definida por la condición  $Cov[F] = I_{m \times m}$ . Cuando el vector de factores comunes,  $F$  es tal que además de estar centrado ( $E[F] = 0$ ), tiene una matriz de covarianzas

$$Cov(F) = E[FF'] = I_{m \times m}$$

se habla de **Análisis Factorial Ortogonal**, dado que los factores son ortogonales en términos estadístico-geométricos.

Para un **Análisis Factorial ortogonal** se particularizan algunos resultados vistos antes. Por ejemplo, la **ecuación estructural** se simplifica, siendo  $\Sigma = LL' + \Psi$ , en la base al par  $(L; \Psi)$ . Y por otro lado,  $Cov[X; F] = L$ , lo que permite hacer una interpretación en este caso muy clara de la matriz de factor-loadings: Sus elementos son las covarianzas de  $X$  (población) con  $F$  (factores), de tal manera que los  $l_{ij} = cov[X_i; F_j]$ ;  $i = 1, \dots, p; j = 1, \dots, m$ . En un Análisis Factorial ortogonal, además, se tiene sin mas que tener en cuenta la ecuación estructural correspondiente:

1.  $\sigma_{ii}$  = varianza de  $X_i$  = elemento  $(i, i)$  de  $LL'$  más elemento  $(i, i)$  de  $\Psi = (l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2) + \psi_i$ . Así pues, la parte de la  $Var(X_i)$  explicada por los factores comunes es  $(l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2)$  que se denomina **comunalidad de  $X_i$** ; el resto,  $\psi_i$ , es la parte de la varianza específica, propia de cada  $X_i$ , es decir la **unicidad de  $X_i$** .
2.  $Cov[X_i; X_k]$  = covarianza de  $X_i$  con  $X_k$ , es decir  $\sigma_{ik}$  = elemento  $(i, k)$  de  $LL'$  + elemento  $(i, k)$  de  $\Psi = l_{i1} + l_{k1} + l_{i2} + l_{k2} + \dots + l_{im} + l_{km}$ .

A su vez a lo largo de este tema (en extracción o estimación del modelo; en los giros de factores y en los “factor loadings”) consideraremos el “**modelo normal de Lawley**” de **Análisis Factorial**. Este tipo de Análisis Factorial, está caracterizado por la hipótesis de distribución normal multivariante de  $X$  (población);  $F$  (factores) y  $\epsilon$  (errores). Este tipo especial de Análisis Factorial introducido por Lawley and Maxwell (1971), tiene implicaciones en los resultados vistos antes, derivados de la teoría de la Normal multivariante. Por ejemplo analicemos la distribución condicionada de  $X$ , dado  $F$ , es decir,  $X/F$ . Para ello, consideremos que  $F \sim N(0, I)$ , de acuerdo con las exigencias del modelo,  $E[F] = 0$  y  $Cov[F] = I$ , y veamos cual es la distribución conjunta de  $(X; F)$ ; como  $X \sim \mu; \Sigma$ , se tendrá:

$$(X; F) \sim N_{p+m} \left[ \begin{pmatrix} \mu \\ 0 \end{pmatrix}; \begin{pmatrix} Cov(X; X) & Cov(X; F) \\ Cov(F; X) & Cov(F; F) \end{pmatrix} \right] = \left( \frac{\Sigma}{L'} \middle| \frac{L}{I} \right) = \left( \frac{LL' + \Psi}{L'} \middle| \frac{L}{I} \right)$$

Por tanto (véase Gutiérrez and González (1991), proposición 1.5), la distribución condicionada será:

$$(X/F) \sim N_p [LF; \Sigma - LL'] \equiv N_p [LF; \Psi]$$

y como  $\Psi$  es diagonal, ello es N. y S. para que  $X$  y  $F$  sean condicionalmente independientes.

Por tanto en este modelo ortogonal normal, el vector  $F$  de factores, con sus factor loadings  $L$ , reproduce todas las covarianzas de  $X$ , excepto en lo que se refiere a las varianzas de  $X$ , que tan sólo reproduce una parte de ellas, la parte restante lo será por la matriz  $\Psi$  de unicidades.

Esto marca una diferencia sustancial entre el Análisis Factorial y el Análisis de Componentes Principales, porque al contrario, el Análisis de Componentes Principales explica (sus  $p$  componentes principales) las varianzas de la población pero nada de sus covarianzas.

Más adelante, al analizar el tema de los “factor-scoring”, consideraremos en el contexto anterior, la distribución condicionada de  $F/X$ .

**Ejemplo 1.** Sea  $\Sigma = \begin{pmatrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{pmatrix}$ ; puede comprobarse que

$$\Sigma = \begin{pmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{pmatrix} \begin{pmatrix} 4 & 7 & -1 & 8 \\ 1 & 2 & 6 & 8 \end{pmatrix} + \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}; (\Sigma = LL' + \Psi)$$

Por tanto:

1. Existen dos factores ( $m = 2$ ), ortogonales, no observables.
2. La communalidad de  $X_1$ , por ejemplo, es  $h_1^2 = l_{11}^2 + l_{12}^2 = 4^2 + 1^2 = 17$ .
3.  $Var[X_1] = 19 = 4^2 + 1^2 + 2^2$  (igualmente se puede hacer para  $X_2$ ).

•

## 1.2. Problemas básicos inherentes al modelo de Análisis Factorial

### 1.2.1. El problema de la identificación en el Análisis Factorial: existencia y unicidad de solución

Consideramos en primer lugar un Análisis Factorial ortogonal. En este modelo se supone que las  $p + \frac{p(p-1)}{2} = \frac{p(p+1)}{2}$  varianzas y covarianzas del vector  $X$  pueden ser reproducidas a partir de los ( $pm$  “factor loadings”  $l_{ij}$  y las  $p$  varianzas específicas  $\psi_i$ ).

Si  $m = p$ , entonces sabemos que  $\Sigma$  es factorizable exactamente como  $LL'$  y entonces  $\Psi$  se puede tomar como una matriz nula. Este es el caso del Análisis de Componentes Principales, previamente estudiado.

Sin embargo, en general, en el Análisis Factorial,  $m < p$ , entre otras cosas porque cuando ello ocurre la utilidad del Análisis Factorial es mayor. En realidad, que  $m < p$  es sustancial para que se tenga un Análisis Factorial.

Entonces las cosas ocurren de la siguiente manera: Por ejemplo, si  $X$  contiene  $p = 12$  variables, y hubiera un modelo de Análisis Factorial con  $m = 2$  factores, entonces, los  $\frac{12 \cdot 13}{2} = 78$  elementos de  $\Sigma$  son descritos en términos de  $m \cdot p + p = 12 \cdot 2 + 12 = 36$  parámetros  $l_{ij}$  y  $\psi_i$  del Análisis Factorial.

Desgraciadamente, no toda matriz de covarianzas puede ser factorizada en la forma  $LL' + \Psi$ , con un número menor de factores que la dimensión  $p$ , e incluso, aunque a veces ello sea técnicamente posible, puede ocurrir que la solución no cumple los requisitos probabilísticos-estadísticos que exige el modelo Factorial estadístico.

**Ejemplo 2.** Sea  $p = 3$  y  $m = 1$ , en un vector aleatorio  $(X_1, X_2, X_3)$  con  $\Sigma = \begin{pmatrix} 1 & 0,9 & 0,7 \\ 0,9 & 1 & 0,4 \\ 0,7 & 0,4 & 1 \end{pmatrix} > 0$ ,

entonces: 
$$\left. \begin{array}{l} X_1 - \mu = l_{11}f_1 + \epsilon_1 \\ X_2 - \mu = l_{21}f_1 + \epsilon_2 \\ X_3 - \mu = l_{31}f_1 + \epsilon_3 \end{array} \right\} \text{(como } m = 1 \Rightarrow F_{1 \times 1} = f_1\text{); } \Sigma = LL' + \Psi \text{ se tiene que cumplir que:}$$

$$\begin{aligned} 1 &= l_{11}^2 + \psi_1; & 1 &= l_{21}^2 + \psi_2; & 1 &= l_{31}^2 + \psi_3 \\ 0,9 &= l_{11} \cdot l_{21}; & 0,4 &= l_{21} \cdot l_{31} \\ 0,7 &= l_{11} \cdot l_{31} \end{aligned}$$

de donde 
$$\left. \begin{array}{l} 0,7 = l_{11} \cdot l_{31} \\ 0,4 = l_{21} \cdot l_{31} \end{array} \right\} \mapsto \left. \begin{array}{l} l_{21} = \frac{0,4}{0,7}l_{11} \\ 0,9 = l_{11}l_{21} \end{array} \right\} \mapsto l_{11}^2 = 1,575 \mapsto l_{11} = \pm 1,255.$$

Ahora bien,  $Var[f_1] = 1$ ;  $Var[X_1] = 1$ ;  $l_{11} = Cov[X_1; F_1] = Corr[X_1F_1]$  y como quiera que el coeficiente de correlación ha de ser en valor absoluto menor o igual que 1, la solución técnica a la que hemos llegado que es  $|l_{11}| = 1,255$  no es estadísticamente aceptable.

Por otra parte, como  $1 = l_{11}^2 + \psi_1 \Rightarrow \psi_1 = 1 - l_{11}^2 = 1 - 1,575 = -0,575$ , nos conduce a otro contrasentido ya que  $\psi_1$  es una varianza que no puede ser negativa.

Así que es posible obtener, para este ejemplo con  $m = 1$ , una solución numérica única de la ecuación  $\Sigma = LL' + \Psi$  pero esta solución no es consistente con la interpretación estadística requerida, es decir, no es propia. •

A continuación analizamos con más profundidad este problema de la existencia y unicidad de solución factorial. Nos situamos en un modelo Factorial oblícuo, en general.

Dada la matriz  $\Sigma$  y un cierto número “m” de factores comunes, sabemos que en nuestro modelo teórico:

1.  $\Sigma = L\Phi L' + \Psi$ , ecuación estructural.
2.  $\Psi$  es diagonal y definida positiva,  $p \times p$ .
3.  $\Phi$  es definida no negativa en general,  $m \times m$ .
4.  $L$  es una matriz  $m \times p$ .

La cuestión puede formularse así:

¿Cuando existirán  $L$ ,  $\Phi$  y  $\Psi$  de tal manera que se cumpla la igualdad estructural? ¿Cuando habrá una solución única?

Para enfocar la solución a estas cuestiones, es decir, al problema de la identificación del Análisis Factorial (existencia y unicidad), hay que recordar en primer lugar que los triples

$$(L; \Phi; \Psi); (L^* = LC; \Phi^* = C^{-1}\Phi C'^{-1}; \Psi), \quad C \text{ } m \times m \text{ no singular.}$$

conducen a modelos equivalentes (verifican ambos la ecuación estructural). Por tanto para que desaparezca la indeterminación habrá que exigir  $m^2$  condiciones sobre  $\Phi$  y  $L$  (ya que  $C$  es una matriz  $m \times m$  no singular). Como  $\Sigma$  es observable, y es simétrica y tiene  $\frac{1}{2}p(p+1)$  elementos diferentes, que junto a las  $m^2$  condiciones antes dichas para la unicidad, hacen un total de  $[\frac{1}{2}p(p+1) + m^2]$  cantidades observadas. Por otra parte  $L$ ,  $\Phi$  y  $\Psi$  tienen un total de  $[pm + \frac{1}{2}m(m+1) + p]$  parámetros. Luego el signo de la diferencia:

$$\frac{1}{2} [(p-m)^2 - p - m]$$

marcará la situación sobre la existencia y unicidad; el hecho que sea mayor o menor que 0 nos indicará la posibilidad de existencia y unicidad de solución. Si la diferencia es positiva puede esperarse un problema de existencia de solución factorial aunque hubiera unicidad en el sentido de que existiera un conjunto de parámetros. Si es negativa, puede esperarse la existencia de solución pero posiblemente no haya unicidad en la determinación de los parámetros en  $L$  y  $\Psi$ . Si esa cantidad es cero puede esperarse la identificación del modelo de Análisis Factorial (existencia y unicidad), o al menos la existencia de un número finito de soluciones.

En general, el problema de la existencia de solución, se puede enunciar en términos de matrices así:

¿Cuando existe una matriz diagonal  $\Psi$  con diagonal de elementos no negativos tal que  $(\Sigma - \Psi)$  es una matriz,  $L\Phi L'$ ,  $p \times p$  semidefinida positiva de rango  $m$ ?

Para este problema de existencia de solución, interpretado en términos algebráicos, se han dado diversos resultados. Anderson and Rubin (1956) recogieron y aportaron resultados clásicos al respecto (entre ellos el primer resultado que se demostró por Reiersol (1950)). Un compendio de estos resultados, junto con otros más

recientes, por ejemplo los obtenidos por Takeuchi, Yanai, and Mukherjee (1982), entre otros, puede verse en Basilewsky (1994) pp 356-359.

Si existe una solución y es única el modelo se dice **identificado**. Ya se ha dicho que han de ponerse  $m^2$  condiciones sobre  $L$  y  $\Phi$  para eliminar la indeterminación dada por la matriz  $C$  tal que  $L^* = LC$  y  $\Phi^* = C^{-1}\Phi C'^{-1}$ . Algunas soluciones parciales son las siguientes:

- Una solución así es la que exigimos en el caso ortogonal  $\Phi = I$ , que obliga a que  $C$  sea ortogonal. Obsérvese que al exigir  $\Phi = I$ , obligamos a que se cumplan  $\frac{m(m+1)}{2}$  ecuaciones.
- Para otras cuestiones, por ejemplo en el método de máxima verosimilitud asociado al modelo Lawley-Maxwell, se obliga además a que se cumpla la restricción que  $\Gamma = L'\Psi^{-1}L$  sea diagonal. En este caso, si los elementos de la diagonal  $\Gamma$ , son distintos y ordenados  $[\gamma_{11} > \gamma_{22} > \dots > \gamma_{mm}]$ , entonces  $L$  está únicamente determinada (ver sección posterior 1.3.4).

Hay otros tipos de restricciones (véase Anderson (1984) pp. 554-556) para procurar la identificación y unicidad, por ejemplo imponiendo “ceros” en la matriz  $L$  de factor-loadings. Veamos los casos de Estructura Simple y de ceros en posiciones específicas.

#### ■ Estructura simple.

Hay condiciones propuestas por Thurstone (1945), pp 335, para escoger una matriz fuera de las  $LC$  que tienen un particular interés en Sicometría. Si  $\lambda_{i\alpha} = 0$ , entonces el factor  $\alpha$ -ésimo no afecta al  $i$ -ésimo test. La idea de la “estructura simple” es que algunos test no dependen de todos los factores cuando estos tengan un significado psicológico real. Esto sugiere que dada una  $L$  consideraríamos todas las rotaciones posibles, es decir, todas las matrices  $LC$  con  $C$  ortogonal, y elegiríamos la que tenga un mayor número de coeficientes cero. Esta matriz puede ser considerada como la que proporciona la estructura más simple; y, presumiblemente la de mayor significado en una interpretación sicométrica (ver apartado 1.2.2 posterior sobre rotaciones en un Análisis Factorial ortogonal).

Las posiciones de los ceros no se escogen de antemano, pero las rotaciones  $C$  se aplican hasta encontrar un  $L$  que satisfaga esas condiciones en la medida de que tengan el mayor número posible de “ceros”.

#### ■ Elementos cero en posiciones específicas. Análisis Factorial confirmatorio y Análisis Factorial exploratorio.

Hay otra forma de abordar la existencia de ceros en  $L$ . Consideraremos un conjunto de condiciones que requiera del investigador una mayor información a priori. Éste debe conocer que algunos test particulares no dependen de algunos factores específicos. En tal caso no asumimos que  $E[ff'] = I$ . Esas condiciones son similares a algunas usadas en modelos econométricos. Los coeficientes de la  $\alpha$ -ésima columna están identificados, excepto por la multiplicación por un factor de escala si hay al menos  $m - 1$  elementos cero en esa columna y si el rango de  $L^\alpha$  es  $m - 1$ , donde  $L^\alpha$  es la matriz compuesta por las filas que contienen los ceros asignados en la  $\alpha$ -ésima columna, con esos ceros asignados borrados. La multiplicación de una columna por una escala constante puede ser eliminada por una normalización, tal como  $\phi_{\alpha\alpha} = 1$  o  $\lambda_{i\alpha} = 1$  para algún  $i$  para cada  $\alpha$ . Si  $\phi_{\alpha\alpha} = 1$ ,  $\alpha = 1, \dots, m$ , entonces  $\Phi$  es una matriz de correlaciones.

Veremos que hay  $m$  normalizaciones y un mínimo de  $m(m - 1)$  condiciones cero. Esto es igual al número de elementos de  $C$ . Si hay más de  $m - 1$  elementos cero especificados en una o más columnas de  $L$ , entonces puede haber más condiciones que las requeridas para acabar con la indeterminación en  $LC$  en cuyo caso las condiciones pueden restringir  $L\Phi L'$ .

Por ejemplo, podemos considerar el siguiente modelo:

$$X = \mu + \begin{pmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & \lambda_{32} \\ 0 & \lambda_{42} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v \\ a \end{pmatrix} + U = \mu + \begin{pmatrix} v & 0 \\ \lambda_{21}v & 0 \\ \lambda_{31}v & \lambda_{32}a \\ \lambda_{42}a & a \end{pmatrix} + U$$

para las puntuaciones de 5 test, donde  $v$  y  $a$  son medidas de habilidad verbal y numérica. Es decir,  $X$  es el vector de puntuaciones en cinco test y suponemos un modelo factorial con  $m = 2$  factores ( $v$  y  $a$ ).

Los dos primeros test dependen únicamente de la habilidad verbal, mientras que los dos últimos lo hacen solo de la numérica. Las normalizaciones ponen la verbal en la escala del primer test y la aritmética en la escala del quinto test. Reiersol (1950) y Howe (1955) sugirieron el uso de ceros preasignados para la identificación y desarrollaron la estimación bajo normalidad para ese caso. Jreskog (1969) llamó al Análisis Factorial bajo esas condiciones de identificación “**Análisis Factorial Confirmatorio**”; bajo condiciones arbitrarias o con “rotación de estructura simple” lo denominó “**Análisis Factorial Exploratorio**”.

### 1.2.2. Un problema de “ambigüedad” en el Análisis Factorial. Rotación de factores e interpretación en el Análisis Factorial

Por otra parte, existe siempre una ambigüedad asociada al Análisis Factorial ortogonal. Esta es la siguiente: Sea  $T$  una matriz ortogonal ( $m \times m$ ) cualquiera, es claro que:

$$X - \mu = LF + \epsilon = L'TT'F + \epsilon = L^*F^* + \epsilon$$

con  $L^* = LT$  y  $F^* = T'F$ . Entonces, puesto que:

$$\begin{cases} E[F^*] = T'E[F] = 0 \\ Cov[F^*] = T'Cov[F]T = TT' = I_{m \times m} \end{cases}$$

es imposible sobre la base de las observaciones de  $X$  distinguir los factores loadings  $L$  y  $L^*$ . O dicho de otra manera, los factores  $F$  y  $F^*$  tienen las mismas propiedades estadísticas y aunque sus factor loadings respectivos  $L$  y  $L^*$  son en general diferentes, generan ambos la misma matriz de covarianzas  $\Sigma$  del  $X$ , ya que en efecto:

$$\Sigma = LL' + \Psi = LTT'L' + \Psi = L^*L^{*\prime} + \Psi.$$

Nótese que las matrices ortogonales corresponden a rotaciones y reflexiones del sistema de ejes coordenados para el  $X$ .

Es fácil probar también que las comunidades, dadas por los elementos diagonales de  $LL'$  no cambian ya que  $LL' = L^*L^{*\prime}$ .

La conclusión que se obtiene de lo anterior es que, aunque exista una solución  $(L; \Psi)$  para el Análisis Factorial ortogonal, ésta no es única.

Como ya se ha indicado, los “factor loadings” obtenidos por cualquier transformación ortogonal a partir de unos iniciales, reproducen la covarianza (o matriz de correlaciones). Como técnicamente hablando, una transformación ortogonal equivale a una rotación rígida (o reflexión) de los ejes coordinados, se habla de “rotación de factores”.

Supongamos que por algún método (máxima verosimilitud, componentes principales, etc.) que en la sección 1.3 posterior se estudiarán, se han obtenido estimaciones  $\hat{L}_{p \times m}$  y  $\hat{\Psi}$ . Entonces:

$$\hat{L}^* = \hat{L}T \quad \text{con} \quad T'T = TT' = I$$

es una matriz de “loading rotados”. Además a:

$$\hat{L}\hat{L}' + \hat{\Psi}$$

que es la estimación de  $\Sigma$ , le ocurre que:

$$\hat{\Sigma} = \hat{L}\hat{L}' + \hat{\Psi} = \hat{L}TT'\hat{L}' + \hat{\Psi} = \hat{L}^*\hat{L}^{*''} + \hat{\Psi}$$

es decir, que la estimación de  $\Sigma$ , no se altera cuando “se giran los factores”.

Además si  $S_n$  es la estimación “directa” de  $\Sigma$  en base a la muestra;  $\hat{L}\hat{L}' + \hat{\Psi}$  es la estimación de  $\Sigma$  a través del Análisis Factorial; y  $S_n - \hat{L}\hat{L}' - \hat{\Psi}$  es la matriz residual, que desempeñará un importante papel posteriormente en el método de estimación de las componentes principales, ocurre que dicha matriz residual es igual a  $S_n - \hat{L}^*\hat{L}^{*''} - \hat{\Psi}$ . Tampoco cambian las comunidades  $\hat{h}_i^2$  estimadas, ni las varianzas específicas  $\hat{\psi}_i$ .

Precisamente, esta rotación de factores, se utiliza en la práctica para obtener a partir de unos “factor loadings” iniciales, otros rotados para los que la interpretación sea de una estructura más simple. La situación podría ser comparable a lo que ocurre al mirar por un microscopio cuando se enfoca de distintas maneras hasta conseguir ver con más detalle.

¿Cuál es la situación ideal que podríamos tener? ¿Qué podríamos entender por una “estructura simple” de fácil interpretación? Pues por ejemplo, que cada variable estuviera ligada, “pesada”, con un solo factor fuertemente y con los demás moderadamente pesada o poco pesada.

Conseguir por rotaciones esta estructura simple, no siempre es posible. En la sección posterior (1.4), analizaremos más pormenorizadamente esta cuestión de rotación de factores.

### 1.2.3. Cambio de unidades de medida

1. En primer lugar nos planteamos la siguiente cuestión: ¿qué ocurre en el modelo de Análisis Factorial ortogonal u oblicuo, si se efectúa un cambio en las unidades de medida? Obsérvese que esta primer cuestión es previa y más general que el problema de estimación o extracción de factores por algún método (ver más adelante, sección 1.3, distintos métodos para la estimación).

Un cambio en las unidades de medida en que se miden las variables  $X_i$  de  $X$ , equivalente a multiplicarlas por una cantidad constante. La cuestión que queremos analizar es si esta transformación alterará la matriz de “factor loadings”. Veamos:

$$X = LF + \epsilon + \mu \mapsto \mathbb{D}X = \mathbb{D}LF + \mathbb{D}\epsilon + \mathbb{D}\mu$$

siendo  $\mathbb{D}$  una matriz diagonal con elementos positivos en la diagonal principal. Es claro que si

$$\mathbb{D}L = L^*; \quad \mathbb{D}\epsilon = \epsilon^*; \quad \mathbb{D}\mu = \mu^*$$

el modelo de Análisis Factorial inicial se transforma en este

$$X^* = L^*F + \epsilon^* + \mu^*$$

de modo que los “factor loadings” quedan multiplicados por  $\mathbb{D}$ .

Por otra parte la ecuación estructural de  $\Sigma$  se transforma así:

$$\begin{aligned} \Sigma &= L\Phi L' + \Psi \rightarrow E[(X^* - \mu^*)(X^* - \mu^*)'] = E[(L^*F + \epsilon)(L^*F + \epsilon)] = \\ &= L^*\Phi L^{*''} + \Psi^* = \Sigma^* \end{aligned}$$

en donde  $\Psi^* = \mathbb{D}\Psi\mathbb{D}' = \mathbb{D}\Psi\mathbb{D}$  y  $\Phi = E[FF']$  (en el caso oblicuo  $\Phi \neq I$ ).

Por otra parte  $\Sigma^* = Cov[X^*] = \mathbb{D}\Sigma\mathbb{D}$ , por lo que

$$\Sigma^* = \mathbb{D}\Sigma\mathbb{D} = L^*\Phi L' + \Psi^*.$$

Por lo tanto la conclusión es clara, un cambio en las unidades, mantiene los factores comunes, pero altera los respectivos “factor loadings”.

También es digno de notarse una propiedad que goza el cambio de unidades analizado, que es que supuesto que hayamos exigido condiciones específicas para la identificación-unicidad de estimadores, es decir, por ejemplo

$$\begin{cases} \Phi = I(\text{caso ortogonal}) \\ L'\Psi^{-1}L = \Delta \text{ matriz diagonal} \end{cases}$$

¿seguirá siendo identificado-único el modelo de Análisis Factorial transformado por  $\mathbb{D}$ ? La respuesta es si. En primer lugar, hemos visto que un cambio tal no altera los factores, luego si el modelo es ortogonal, lo es el transformado. Respecto a la unicidad de los estimadores, sigue siendo válida. En efecto:

$$L^{*'}\Psi^{*-1}L^* = L'\mathbb{D}(\mathbb{D}\Psi\mathbb{D})^{-1}\mathbb{D}L = L'\Psi^{-1}L$$

luego si  $L'\Psi^{-1}L$  es diagonal, así lo es también  $L^{*'}\Psi^{*-1}L^*$ .

Puede comprobarse que si se utilizan otras formas de identificación (por ejemplo haciendo ceros algunos términos específicos de  $L$  y  $\Phi$  es una matriz de correlaciones) entonces  $L^* = \mathbb{D}L$  esta igualmente identificada.

2. El cambio dado por  $\mathbb{D}$ , tiene un caso particular notable en la práctica. Sea  $d_{ii} = \frac{1}{\sqrt{\sigma_{ii}}}$ , es decir  $\mathbb{D} = V^{-1/2}$  en donde  $V$  es una matriz diagonal constituida por las  $\sigma_{ii}$  de  $\Sigma$ . En este caso:

$$X^* - \mu^* = \mathbb{D}X - \mathbb{D}\mu = L^*F + \epsilon^* = \mathbb{D}LF + \mathbb{D}\epsilon$$

en donde  $\Sigma^* = \mathbb{D}\Sigma\mathbb{D} = \mathbb{R}$ , con  $\mathbb{R}$  la matriz de correlaciones de  $X$ .

Por tanto

$$\Sigma^* = \mathbb{D}\Sigma\mathbb{D} = \mathbb{R} = L^*\Phi L' + \Psi^* \quad \text{con } \Psi^* = \mathbb{D}\Psi\mathbb{D}$$

es decir

$$\Sigma^* = \mathbb{R} = V^{-1/2}L\Phi L'V^{-1/2} + V^{-1/2}\Psi V^{-1/2}.$$

### 1.3. Métodos de estimación en el Análisis Factorial

Hay distintos métodos para estimar (extraer) los factores comunes en un modelo de Análisis Factorial ortogonal, es decir para estimar los “factor loadings” ( $L$ ) y la matriz  $\Psi$  de unicidades.

En el desarrollo del presente tema, hemos elegido dos métodos para su estudio detallado, que son paradigmas de la metodología de estimación según se suponga o no se suponga normalidad en el modelo. Cuando no se supone explícitamente normalidad de los elementos aleatorios del modelo básico  $X$ ,  $F$  y  $\epsilon$ , la metodología de “extracción por Componentes Principales” (y del factor principal asociado) es típica al respecto. En cambio cuando se supone normalidad multivariante (modelo factorial de Lawley-Maxwell) el método de “estimación por Máxima Verosimilitud” está ligado de manera natural a la situación y es el método idóneo al respecto.

Analizamos a continuación estos dos métodos. Hay por supuesto otros métodos de estimación, por ejemplo, los métodos, del centroide, o de mínimos cuadrados que también pueden ser útiles en la práctica en determinados enfoques de análisis factoriales propios en ciertos campos de aplicación (ver por ejemplo Basilewsky (1994); Harman (1967))

### 1.3.1. Introducción

La cuestión que se plantea en la práctica y que se intenta resolver por la técnica del Análisis Factorial es la siguiente: Tenemos observaciones  $X_1, \dots, X_N$  de  $p$  variables, correlacionadas en general, y se quiere representar adecuadamente esos datos usando un modelo del tipo  $X = \mu + LF + \epsilon$  (con las condiciones exigidas sobre él, antes especificadas), de tal manera que este modelo “reproducza” los datos con un número pequeño de factores. En términos estructurales y con una visión “exploratoria” se trata de establecer las relaciones propias del Análisis Factorial ortogonal:

1.  $Cov[X] = LL' + \Psi$ , es decir:

$$\begin{aligned} Var[X_i] &= l_{i1}^2 + \dots + l_{im}^2 + \psi_i \\ Cov[X_i; X_j] &= l_{i1}l_{j1} + \dots + l_{im}l_{jm}. \end{aligned}$$

2.  $Cov[X; F] = L$ , es decir,  $Cov[X_i; F_j] = l_{ij}$

en términos estadísticos.

Así pues, es claro que la versión muestral del Análisis Factorial ha de basarse en la matriz de covarianzas muestrales  $S = \frac{A}{N}$ , como estimador de la matriz de covarianzas teóricas poblacionales  $\Sigma$ . Es obvio, por otra parte que si  $S$  es de tal naturaleza que “tiende” a ser una matriz diagonal (o lo que es lo mismo la matriz de correlaciones muestrales  $R$  “tiende” a ser la identidad), entonces las variables son incorreladas; en este caso el Análisis Factorial es irrelevante, siendo los “factores específicos” dominantes en la explicación de los datos.

Por contra, si  $\Sigma$  es claramente una matriz no diagonal, entonces en el Análisis Factorial hay que, inicialmente, estimar los “factor loadings” y las “varianza específicas” (los  $l_{ij}$  y  $\psi_i$  respectivamente).

Ahora bien, ¿qué métodos se emplean para conseguir estas estimaciones? Hay diversos métodos al respecto, dentro de los cuales, los más utilizados son:

- a) El método de las componentes principales (y del factor principal).
- b) El método de máxima verosimilitud (Lawley-Maxwell).

Como advertencia general y a la vista de lo antes comentado sobre la determinación de la matriz de factor loadings, es claro que las soluciones de estos métodos, pueden rotarse de modo que se consiga una mejor interpretación de los factores. Como recomiendan los autores, es aconsejable en todos los casos prácticos, emplear más de un método, porque del análisis de la coherencia de los resultados proporcionados por ellos, en vista de su consistencia, podremos deducir el grado de adecuación de un modelo de Análisis Factorial al caso en cuestión.

### 1.3.2. El método de las componentes principales

#### Aspectos teóricos

Es sabido que dada una matriz de covarianzas  $\Sigma$ , teórica, con autovalores  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$  y autovectores  $e_1, e_2, \dots, e_p$  es posible establecer la descomposición espectral de  $\Sigma$  de tal manera que

$$\Sigma_{p \times p} = \lambda_1 e_1 e_1' + \dots + \lambda_p e_p e_p' = \left( \sqrt{\lambda_1} e_1; \dots; \sqrt{\lambda_p} e_p \right)_{p \times p} \begin{pmatrix} \sqrt{\lambda_1} e_1' \\ \vdots \\ \sqrt{\lambda_p} e_p' \end{pmatrix}_{p \times p} = LL'.$$

Obsérvese que en el Análisis de Componentes Principales,  $\Sigma = \mathbb{P}\Lambda\mathbb{P}'$ , con  $\mathbb{P}\mathbb{P}' = I$  y por lo tanto  $\Sigma = \mathbb{P}\Lambda^{1/2}\Lambda^{1/2}\mathbb{P}' = LL'$ , con  $L = \mathbb{P}\Lambda^{1/2}$ ,  $\Lambda = diag(\lambda_1, \dots, \lambda_p)$  y  $\mathbb{P} = [e_1; \dots, e_p]$ .

Lo cual significa que la estructura buscada para factorizar  $\Sigma$  se consigue de la forma:

$$\Sigma = LL' + 0 = LL'$$

de tal manera que existen tantos factores como variables ( $m = p$ ), y se obtiene un “Análisis Factorial” con las  $\psi_i = 0$ ,  $\forall i = 1, \dots, p$ . En este caso los factor loadings son tales que valen  $\sqrt{\lambda_j}e_j$ , los situados en la columna  $j$ -ésima de la matriz de loadings  $L$ . En definitiva, salvo el factor  $\sqrt{\lambda_j}$ , los factor loadings del  $j$ -ésimo factor son los coeficientes de la componente principal,  $j$ -ésima, dado que, en efecto, los coeficientes de la componente principal  $j$ -ésima es el vector  $e_j$  (ver tema Análisis Componentes Principales).

Así pues, cuando  $m = p$ , se obtiene una representación exacta  $\Sigma = LL'$  de tipo factorial, con tantos factores como variables. Pero esto no es en general el objeto del Análisis Factorial, en el que se pretende fundamentalmente explicar la correlación entre variables con un número significativamente menor de factores que de variables. Obsérvese que además, la representación exacta de  $\Sigma = LL'$ , hace desaparecer los factores específicos o errores  $\epsilon$  en el modelo básico  $X = \mu + LF + \epsilon$  de Análisis Factorial.

Ante esta situación, cabe adoptar la estrategia siguiente: Despreciar las  $p - m$  contribuciones debidas a los  $p - m$  “últimos” autovalores  $\lambda_{m+1} \geq \lambda_{m+2} \geq \dots \geq \lambda_p$ , es decir, despreciar la contribución de  $\lambda_{m+1}e_{m+1}e'_{m+1} + \dots + \lambda_pe_pe'_p$  a la explicación estructural de la matriz  $\Sigma$ . Por tanto:

$$\Sigma \doteq \left( \sqrt{\lambda_1}e_1; \dots; \sqrt{\lambda_m}e_m \right)_{p \times m} \begin{pmatrix} \sqrt{\lambda_1}e'_1 \\ \vdots \\ \sqrt{\lambda_m}e'_m \end{pmatrix}_{m \times p} = L_{p \times m}L'_{m \times p}$$

es una aproximación de la representación exacta de  $\Sigma$ , que en principio sigue ignorando el factor específico  $\epsilon$  en cuanto a la estructuración factorial de  $\Sigma$ . Y decimos en principio porque desde luego cabe la posibilidad de perfeccionar esta aproximación incluyendo las unicidades  $\psi_i$ . Así:

$$\Sigma_{p \times p} = LL' + \Psi = \left( \sqrt{\lambda_1}e_1; \dots; \sqrt{\lambda_m}e_m \right)_{p \times m} \begin{pmatrix} \sqrt{\lambda_1}e'_1 \\ \vdots \\ \sqrt{\lambda_m}e'_m \end{pmatrix}_{m \times p} + \begin{pmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & & 0 \\ 0 & \dots & \ddots & \ddots \\ \vdots & 0 & 0 & \psi_p \end{pmatrix}_{p \times p}$$

siendo obviamente  $\psi_i = \sigma_{ii} - \sum_{j=1}^m l_{ij}^2$ ;  $i = 1, \dots, p$ .

En definitiva el Análisis Factorial es aproximado mediante la metodología del Análisis de Componentes Principales.

**Nota 2.** Recuérdese que en el Análisis Factorial, las comunidades (para un  $i$ , valen  $\sum_{j=1}^m l_{ij}^2$ ) son los elementos de la diagonal principal de  $LL'$ , siendo, en este caso  $L$  dado por  $(\sqrt{\lambda_1}e_1; \dots; \sqrt{\lambda_m}e_m)$ . •

### Aspectos muestrales

Veamos a continuación como se aplica la aproximación descrita en términos teóricos, a datos. Sean  $X_1, \dots, X_N$  los datos correspondientes a  $p$  variables, que tipificados son en definitiva:

$$Z_1, \dots, Z_N, \text{ con } Z_j = \begin{pmatrix} \frac{X_{1j} - \bar{X}_1}{\sqrt{s_{11}}} \\ \vdots \\ \frac{X_{pj} - \bar{X}_p}{\sqrt{s_{pp}}} \end{pmatrix}; j = 1, \dots, N$$

en donde  $\bar{X}_i$  son las medias en cada variable  $i = 1, \dots, p$ ; y  $s_{ii}$  son los elementos de la diagonal principal de  $S$ , es decir, las varianzas muestrales.

Este conjunto  $(Z_1, \dots, Z_N)$  tiene como matriz de covarianzas la matriz de correlaciones muestrales  $R$ . Entonces se puede aplicar la representación dada por la segunda aproximación de  $\Sigma$  vista antes, a la matriz  $S$  o a la  $R$ , originándose lo que se llama la solución de componentes principales, que se llama así porque, como antes decíamos, los factor loadings son, salvo un factor de escala, los coeficientes de las componentes principales primeras muestrales.

En definitiva, este procedimiento de estimación de Análisis Factorial, denominado de componentes principales, sigue los pasos siguientes:

1. Dada la matriz de covarianzas muestrales  $S$ , se calculan los autovalores-autovectores muestrales, como soluciones, respectivamente, de  $|S - \lambda I| = 0$  y  $(S - \lambda I)e = 0$ . Se obtienen así los pares  $(\hat{\lambda}_i; \hat{e}_i)$ , con  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$ ,  $i = 1, \dots, p$ .
2. Se toman, cuando  $m < p$ , los primeros  $\hat{\lambda}_1, \dots, \hat{\lambda}_m$  y se construye:

$$\hat{L} = \left[ \sqrt{\hat{\lambda}_1} \hat{e}_1, \dots, \sqrt{\hat{\lambda}_m} \hat{e}_m \right]_{p \times m}.$$

3. Las varianzas específicas muestrales, estimaciones de las  $\psi_i$ , se construyen mediante las relaciones

$$\hat{\psi}_i = s_{ii} - \sum_{j=1}^m \hat{l}_{ij}^2; \quad i = 1, \dots, p;$$

es decir, la diagonal principal de la matriz  $S - \hat{L}\hat{L}'$ .

Por tanto  $\hat{\Psi} = \text{diag}(\psi_1, \dots, \psi_p)$ , es la matriz estimada de unicidades.

4. Las communalidades estimadas  $\hat{h}_i^2$  se construyen entonces como:

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \dots + \hat{l}_{im}^2.$$

**Nota 3.** Si se actúa con la matriz  $R$ , se realiza todo igual actuando con  $R$  en lugar de  $S$ .•

Este procedimiento de solución del problema de estimación en Análisis Factorial, tiene las siguientes características:

- i) Si el número de factores ( $m$ ) crece, los factor loadings correspondientes a factores ya utilizados, no cambian. Sólo hay que ir añadiendo los correspondientes a los nuevos factores adicionales.
- ii) Por otro lado, del procedimiento empleado se deduce que los elementos diagonales de  $S$  deben ser iguales a los diagonales de  $\hat{L}\hat{L}' + \hat{\Psi}$ .

Con los elementos que no están en la diagonal de  $S$ , lo normal en la práctica es que  $\hat{L}\hat{L}' + \hat{\Psi}$  no reproduzca esos elementos no diagonales. Entonces se plantea la siguiente cuestión, ¿cómo seleccionar el número de factores para conseguir una adecuada representación que reproduzca la matriz  $S$  con el menor “error” posible?

La manera de actuar es análoga a la que se emplea en las componentes principales:

Sea la “matriz residual”:  $S - (\hat{L}\hat{L}' + \hat{\Psi})$ , que resulta de aproximar  $S$  por la solución de componentes principales dada por los pasos 1 y 2 anteriores.

Se sabe que si consideramos la suma de cuadrados “entries”, SCE, de  $S - (\hat{L}\hat{L}' + \hat{\Psi}) = \text{tr}[(S - \hat{L}\hat{L}' - \hat{\Psi})(S - \hat{L}\hat{L}' - \hat{\Psi})']$  se verifica que  $\text{SCE}[S - (\hat{L}\hat{L}' + \hat{\Psi})] \leq \hat{\lambda}_{m+1}^2 + \dots + \hat{\lambda}_p^2$  de tal manera que si  $\hat{\lambda}_{m+1}^2 + \dots + \hat{\lambda}_p^2$  es pequeño entonces la suma de cuadrados de los errores es pequeña.

En efecto, sea la matriz residual  $S - (\hat{L}\hat{L}' + \hat{\Psi})$  que resulta de aproximar  $S$  por la solución de componente principal. Se sabe que, puesto que  $S - \hat{L}\hat{L}' - \hat{\Psi}$  tiene ceros en la diagonal principal, la suma de cuadrados “entries” de  $S - \hat{L}\hat{L}' - \hat{\Psi}$  es menor o igual a la suma de cuadrados de “entries” de  $S - \hat{L}\hat{L}'$ . Pero,  $S - \hat{L}\hat{L}' = \hat{\lambda}_{m+1}\hat{e}_{m+1}\hat{e}'_{m+1} + \dots + \hat{\lambda}_p\hat{e}_p\hat{e}'_p = \hat{P}_{(2)}\hat{L}_{(2)}\hat{P}'_{(2)}$ , donde  $\hat{P}_{(2)} = (\hat{e}_{m+1}, \dots, \hat{e}_p)$  y  $\hat{L}_{(2)} = \text{diag}(\hat{\lambda}_{m+1} \dots \hat{\lambda}_p)$ . Ahora bien: la suma de cuadrados “entries” de  $A$  es igual a  $\text{tr}(AA')$  de donde  $\text{tr}[\hat{P}_{(2)}\hat{L}_{(2)}\hat{P}'_{(2)}\hat{P}_{(2)}\hat{L}'_{(2)}\hat{P}'_{(2)}] = \text{tr}[\hat{P}_{(2)}\hat{L}_{(2)}\hat{L}'_{(2)}\hat{P}'_{(2)}] = \text{tr}[\hat{L}_{(2)}\hat{P}'_{(2)}\hat{P}_{(2)}\hat{L}_{(2)}] = \text{tr}[\hat{L}_{(2)}\hat{L}_{(2)}]$ .

De manera ideal, si la contribución de unos pocos primeros factores a la varianza total es grande, el Análisis Factorial “funcionara” adecuadamente.

La contribución del primer factor a la varianza muestral es  $\hat{l}_{i1}^2$ . Dada la varianza total

$$s_{11} + \dots + s_{pp} = \text{tr}(S)$$

la contribución a ella del primer factor será:

$$\hat{l}_{i1}^2 + \dots + \hat{l}_{ip}^2 = \left( \sqrt{\hat{\lambda}_1} \hat{e}_1 \right)'_{1 \times p} \left( \sqrt{\hat{\lambda}_1} \hat{e}_1 \right)_{1 \times p} = \hat{\lambda}_1.$$

Y en general, la proporción de la varianza total debida al factor  $j$ -ésimo será  $\frac{\hat{\lambda}_j}{s_{11} + \dots + s_{pp}}$  (sobre  $S$ ) o  $\frac{\hat{\lambda}_j}{p}$  (sobre  $R$ ).

Y en base a esta expresión se decide heurísticamente el número de factores a usar, haciendo el error lo suficientemente pequeño, es decir, contentándonos con una proporción de varianza explicada suficientemente alta o lo que es equivalente, tomando un modelo de Análisis Factorial con un número de factores  $m$  en la metodología de Análisis de Componentes Principales utilizada.

**Nota 4.** Otra convención que suele hacerse, es tomar  $m$  igual al número de autovalores de  $R$  mayores que 1; o igual al número de autovalores de  $S$  que son positivos. Pero no debe aplicarse indiscriminadamente, pues, por ejemplo, para grandes muestras, es esperable que todos los autovalores de  $S$  sean positivos y entonces  $m$  sería  $p$  y no hay en tal caso, Análisis Factorial en sentido estricto. •

**Ejemplo 3.** Sea  $R = \begin{pmatrix} 1 & 0,02 & 0,96 & 0,42 & 0,01 \\ 0,02 & 1 & 0,13 & 0,71 & 0,85 \\ 0,96 & 0,13 & 1 & 0,5 & 0,11 \\ 0,42 & 0,71 & 0,5 & 1 & 0,79 \\ 0,01 & 0,85 & 0,11 & 0,79 & 1 \end{pmatrix}$ .

Los dos únicos autovalores de  $R$  mayores de 1, son  $\hat{\lambda}_1 = 2,85$  y  $\hat{\lambda}_2 = 1,81$ . Entonces,  $\frac{\hat{\lambda}_1 + \hat{\lambda}_2}{p} = 0,93$ .

Variable	Estimación de los factor loading		$\hat{h}_i^2$	$\text{Varianzas específicas}$ $\hat{\psi}_i = 1 - \hat{h}_i^2$
	$F_1$	$F_2$		
1	0.56	0.82	0.98	0.02
2	0.78	-0.53	0.88	0.12
3	0.65	0.75	0.98	0.02
4	0.94	-0.11	0.89	0.11
5	0.80	-0.54	0.93	0.07
Autovalores			2.85	1.81
Proporción acumulada de la varianza total			0.571	0.932

$$\begin{aligned}
Y & \\
\hat{L}\hat{L}' + \hat{\Psi} &= \begin{pmatrix} 0,56 & 0,82 \\ 0,78 & -0,53 \\ 0,65 & 0,75 \\ 0,94 & -0,11 \\ 0,80 & -0,54 \end{pmatrix} \begin{pmatrix} 0,56 & 0,78 & 0,65 & 0,94 & 0,80 \\ 0,82 & -0,53 & 0,75 & -0,11 & -0,54 \end{pmatrix} + \\
& \begin{pmatrix} 0,02 & 0 & 0 & 0 & 0 \\ 0 & 0,12 & 0 & 0 & 0 \\ 0 & 0 & 0,02 & 0 & 0 \\ 0 & 0 & 0 & 0,11 & 0 \\ 0 & 0 & 0 & 0 & 0,07 \end{pmatrix} = \begin{pmatrix} 1 & 0,01 & 0,97 & 0,44 & 0,00 \\ 0,01 & 1 & 0,11 & 0,79 & 0,91 \\ 0,97 & 0,11 & 1 & 0,53 & 0,11 \\ 0,44 & 0,79 & 0,53 & 1 & 0,81 \\ 0,00 & 0,91 & 0,11 & 0,81 & 1 \end{pmatrix} \bullet
\end{aligned}$$

### 1.3.3. La “Principal Factor Solution” como modificación del método de las Componentes Principales en estimación en Análisis Factorial

Describimos el método en términos de  $R$ , matriz de correlaciones muestrales, aunque se puede hacer igualmente a partir de la matriz  $S$ . Como ya se ha dicho en el método teórico del método de Componentes Principales, la matriz de correlación teórica  $\rho$  queda estructurada de la forma:

$$\rho = LL' + \Psi.$$

Si la especificación del Modelo es correcta, los  $m$  factores harán que:

$$\rho_{ij} = 1 = h_i^2 + \psi_i$$

siendo  $h_i^2$ ,  $i = 1, \dots, p$ , las comunalidades de las variables  $X_i$ .

Por tanto

$$\rho - \Psi = LL'.$$

Supongamos que se tienen estimaciones iniciales  $\psi_i^*$  de las varianzas específicas. Entonces reemplazando los elementos de la diagonal de  $R$  por  $h_i^{*2} = 1 - \psi_i^*$ , obtenemos una matriz de correlaciones muestrales llamada reducida:

$$R_r = \begin{pmatrix} h_1^{*2} & r_{12} & \cdots & \cdots & r_{1p} \\ r_{12} & h_2^{*2} & r_{13} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{1p} & \cdots & \cdots & \cdots & h_p^{*2} \end{pmatrix}.$$

Entonces se aplica el método de componentes principales a:

$$R_r = L_r^* L_r^{*'} \quad (1)$$

en donde  $L_r^* = (l_{ij}^*)$ , con  $l_{ij}^*$  los factor loadings estimados, es decir:

$$\begin{aligned}
L_r^* &= \left[ \sqrt{\hat{\lambda}_1^*} \hat{e}_1^*; \sqrt{\hat{\lambda}_2^*} \hat{e}_2^*, \dots, \sqrt{\hat{\lambda}_m^*} \hat{e}_m^* \right] \\
\psi_i^* &= 1 - \sum_{j=1}^m l_{ij}^{*2}
\end{aligned}$$

siendo  $(\hat{\lambda}_i^*, \hat{e}_i^*)$ ,  $i = 1, \dots, m$ ; los  $m$  mayores autovalores-autovectores de  $R_r$ . Se puede entonces, re-estimar las comunalidades

$$\hat{h}_i^* = \sum_{j=1}^m l_{ij}^{*2}$$

que puede utilizarse, reiteradamente, como una estimación no inicial, en lugar de  $h_i^{*2}$ .

Igual que en el método anterior de Componentes Principales, las consideraciones sobre  $\hat{\lambda}_1^*, \dots, \hat{\lambda}_p^*$ , sirven para determinar los factores comunes a retener.

Hay una complicación en este método: algunos autovalores pueden ser negativos debido al uso inicial de una estimación de las comunidades. Idealmente, se toman tantos factores como rango tenga  $R_r$ , pero en todo caso debe procederse con precaución.

Por otra parte cabe hacer algunas consideraciones sobre como hacer la elección de las estimaciones iniciales de las comunidades. La forma más popular de hacer esta elección es tomar las estimaciones iniciales citadas así:  $\psi_i^* = \frac{1}{r^{ii}}$ , con  $r^{ii}$  el elemento  $(i, i)$  de  $R^{-1}$ , es decir:

$$h_i^{*2} = 1 - \psi_i^* = 1 - \frac{1}{r^{ii}}.$$

Obsérvese que esto equivale a tomar  $h_i^{*2}$  como el cuadrado del coeficiente de correlación múltiple entre  $X_i$  y las  $p - 1$  variables restantes. Obsérvese también, que el método de las Componentes Principales para  $R$  puede ser considerado como un método de factor principal con estimaciones iniciales de comunidades igual a uno o varianzas específicas iguales a cero. Ambas aproximaciones son filosófica y geométricamente diferentes (el método de componentes principales y el método del factor principal). En la práctica, ambos procedimientos conducen a factor loadings similares, si el número de variables es grande y el número de factores pequeño.

### 1.3.4. El método de Máxima Verosimilitud en el Modelo de Análisis Factorial (Modelo de Lawley-Maxwell)

Revisemos los aspectos generales que el método de Máxima Verosimilitud tiene en cuanto a la estimación del Modelo. Otros aspectos, como los contrastes hipótesis útiles en la Inferencia sobre el Modelo de Análisis Factorial que estamos considerando, serán estudiados en una sección posterior (1.7).

#### Estimación de parámetros del Modelo de Análisis Factorial (ortogonal y oblicuo)

##### ■ Modelo de Análisis Factorial ortogonal

Supongamos el Modelo de Factores ortogonales, en primer lugar. En este caso la estructuración de la matriz de covarianza es:

$$X = \mu + LF + \epsilon; \quad \Sigma = LL' + \Psi$$

de tal manera que  $X$  se supone distribuida Normal Multivariante, así como el vector de error  $\epsilon$ . Por tanto,

$$X_{p \times 1} \sim \mathcal{N}_p [\mu; \Sigma = LL' + \Psi]$$

Supongamos que disponemos de observaciones:

$$\{X_\alpha = (X_{\alpha 1}, \dots, X_{\alpha p})'; \alpha = 1, \dots, N\}$$

que constituyen una muestra aleatoria independiente del vector poblacional  $X$ .

Supongamos los estadísticos muestrales:

$$\bar{X} = \hat{\mu} = \frac{1}{N} \sum_{\alpha=1}^N X_\alpha \quad \frac{s}{N} = \hat{\Sigma} = \frac{1}{N} \sum_{\alpha=1}^N (X_\alpha - \bar{X})(X_\alpha - \bar{X})' \equiv \frac{A}{N}$$

que, en las hipótesis de normalidad tomadas, son los estimadores máximo-verosímiles de los parámetros  $\mu$  y  $\Sigma$  de la Normal Multivariante con la que se distribuye el vector  $X$ .

Obviamente, esta hipótesis de normalidad equivale a la normalidad del vector  $F$  y del  $\epsilon$ , que según las hipótesis generales del Modelo de Análisis Factorial, por otra parte, son incorreladas y, en virtud de la

normalidad, independientes.

Recuérdese también, que en el Análisis Factorial ortogonal, la matriz  $Cov(F) = I$ . En estas circunstancias, la función de verosimilitud adopta la siguiente forma:

$$\mathbb{L}(L; \Psi; \mu) = (2\pi)^{-\frac{Np}{2}} |LL' + \Psi|^{-\frac{N}{2}} \exp\left(-\frac{1}{2} \text{tr}\left[(LL' + \Psi)^{-1} (s + N(\bar{X} - \mu)(\bar{X} - \mu)')\right]\right)$$

(véase por ejemplo Gutiérrez and González (1991)).

Dada esta verosimilitud, lo primero que cabe observar es que si se cambia  $L$  por  $LO$  (siendo  $O$  una matriz ortogonal  $p \times p$ ) dicha verosimilitud no cambia. Ello implica que si  $\hat{L}$  es la matriz de estimación máximo-verosímil de  $L$ , también es una matriz estimadora máximo-verosímil de  $L$  la matriz  $\hat{L}O$ . Se impone pues exigir la unicidad de la matriz estimadora, para lo que es habitual exigir que:

$$L'\Psi^{-1}L = \Delta \text{ con } \Delta \text{ una matriz } (m \times m) \text{ diagonal}$$

(ver Anexo ??, para la justificación teórica).

En segundo lugar, la función de verosimilitud anterior puede escribirse así:

$$\begin{aligned} \mathbb{L}(L; \Psi; \mu) &= (2\pi)^{-\frac{(N-1)p}{2}} |LL' + \Psi|^{-\frac{(N-1)}{2}} \exp\left(-\frac{1}{2} \text{tr}\left[(LL' + \Psi)^{-1}s\right]\right) \times \\ &\quad (2\pi)^{-\frac{p}{2}} |LL' + \Psi|^{-\frac{1}{2}} \exp\left(-\frac{N}{2}(\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu)\right) \end{aligned}$$

El último factor puede escribirse además así:

$$(2\pi)^{-\frac{p}{2}} |LL' + \Psi|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \text{tr}\left[N(LL' + \Psi)^{-1}(\bar{X} - \mu)(\bar{X} - \mu)'\right]\right)$$

y este factor, al maximizar  $\mathbb{L}$ , y hacerse cero, es claro que permite concluir que  $\hat{\mu} = \bar{X}$ . Por tanto el problema queda reducido a

$$\underset{L, \Psi}{\text{Max}} \mathbb{L}(L; \Psi; \mu) = \underset{L, \Psi}{\text{Max}} \mathbb{L}(L; \Psi; \bar{X})$$

o bien (ver Gutiérrez and González (1991)),  $\underset{L, \Psi}{\text{Max}} \ln \mathbb{L}(L; \Psi; \mu)$ .

Puede demostrarse (utilizando el cálculo de Neudecker) que esta maximización conduce al siguiente resultado:

**Teorema 1.** Los estimadores de máxima-verosimilitud de  $L$  y  $\Psi$ , denotados  $\hat{L}$  y  $\hat{\Psi}$ , en el modelo de Análisis Factorial ortogonal supuesto, con la condición de unicidad, obedecen al siguiente sistema:

$$\text{Diag}(\hat{L}\hat{L}' + \hat{\Psi}) = \text{diag}\left(\frac{s}{N}\right)$$

$$\frac{s}{N} \hat{\Psi}^{-1} \hat{L} = \hat{L}(I + \hat{L}' \hat{\Psi}^{-1} \hat{L})$$

(ver Anexo ?? para la demostración detallada).

#### ■ Modelo Análisis Factorial oblicuo

En este caso la  $Cov(F) = \Phi$  en lugar de  $Cov[F] = I$ . Existe pues una correlación entre los factores. En este caso la ecuación estructural toma el aspecto siguiente:

$$\Sigma = L\Phi L' + \Psi.$$

Puede demostrarse que en este caso, el método de Máxima-Verosimilitud conduce al siguiente resultado:

**Teorema 2.** Los estimadores de máxima verosimilitud de  $L$ ,  $\Phi$  y  $\Psi$ , vienen dados, en el modelo de Análisis Factorial oblicuo, como soluciones del sistema siguiente:

$$\begin{aligned}\hat{\Psi} &= \text{diag} \left[ \frac{s}{N} - \hat{L}\hat{\Phi}\hat{L}' \right] \\ \hat{\Phi}\hat{L}\hat{\Psi}^{-1}\hat{L}' + I &= \left( \hat{L}\hat{\Psi}^{-1}\hat{L}' \right)^{-1} \left( \hat{L}'\hat{\Psi}^{-1}\frac{s}{N}\hat{\Psi}^{-1}\hat{L} \right) \\ \hat{\Phi}\hat{L} \left( \hat{L}\hat{L}' + \hat{\Psi}^{-1} \left( I - \frac{s}{N}(\hat{L}\hat{L}' + \hat{\Psi})^{-1} \right) \right) &= \hat{\Phi}\hat{L}' \left[ I - (\hat{L}\hat{L}' + \hat{\Psi})^{-1}\frac{s}{N} \right] \hat{\Psi}^{-1}.\end{aligned}$$

### Obtención de los estimadores

La resolución de los sistemas dados para el caso ortogonal y oblicuo, se hace por medio de métodos numéricos. Son en definitiva la solución que maximiza  $\mathbb{L}(L, \Psi, \mu)$  o  $\mathbb{L}(L, \Psi, \Phi, \mu)$ , sujeto a la restricción de unicidad antes impuesta: que  $\hat{L}'\hat{\Psi}^{-1}\hat{L}$  sea diagonal. Obsérvese que obtenidos los estimadores  $\hat{L}$ ,  $\hat{\Psi}$  y  $\hat{\mu}$ , en el caso ortogonal por ejemplo, las comunidades (en virtud del teorema de Zehna o principio de invarianza de los estimadores de máxima verosimilitud) serán estimadas máximo-verosimilmente mediante la correspondiente función de los estimadores, es decir:

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \dots + \hat{l}_{im}^2; \quad i = 1, 2, \dots, p$$

de modo que la estimación de la proporción de la varianza total “explicada” o “debida” al factor  $j$ -ésimo será:

$$\frac{\hat{l}_{1j}^2 + \hat{l}_{2j}^2 + \dots + \hat{l}_{pj}^2}{s_{11} + s_{22} + \dots + s_{pp}}.$$

### Notas sobre la Estimación Máximo-Verosímil en el Análisis Factorial ortogonal

Analizamos a continuación la repercusión en el método de Máxima Verosimilitud de lo visto en la sección 1.2.3 anterior. En particular veamos que ocurre con el método de máxima-verosimilitud de estimación del modelo de Análisis Factorial ortogonal cuando se efectúa un cambio de unidades dado por una matriz  $\mathbb{D}$ , como la antes definida (y en particular con un  $\mathbb{D} = V^{-1/2}$ ).

Sea en efecto un cambio de tipificación poblacional, de tal manera que el vector  $X$  se transforma en el  $Z = V^{-1/2}(X - \mu)$  (como caso particular de  $\mathbb{D}(X - \mu) = \mathbb{D}X - \mathbb{D}\mu$ ).

Como ya hemos estudiado (ver sección 1.2.3), el nuevo vector  $Z$  tiene entonces la covarianza  $\mathbb{R}$

$$\mathbb{R} = \rho = V^{-1/2}\Sigma V^{-1/2} = V^{-1/2}LL'V^{-1/2} + V^{-1/2}\Psi V^{-1/2}$$

lo que implica que  $Z$  es tal que su matriz de covarianzas  $\mathbb{R}$  se factoriza de la forma:

$$\mathbb{R} = L_Z L'_Z + \Psi_Z \quad \text{con } L_Z = V^{-1/2}L; \quad \Psi_Z = V^{-1/2}\Psi V^{-1/2}$$

Entonces, aplicando de nuevo el principio de invarianza de Zehna, si  $\hat{L}$  y  $\hat{\Psi}$  son estimadores de máxima-verosimilitud de  $L$  y  $\Psi$  en el modelo  $X = \mu + LF + \epsilon$ , entonces  $\hat{L}_Z = \hat{V}^{-1/2}\hat{L}$ ;  $\hat{\Psi}_Z = \hat{V}^{-1/2}\hat{\Psi}\hat{V}^{-1/2}$  son estimadores de máxima verosimilitud de  $L_Z$  y  $\Psi_Z$  respectivamente, es decir, el estimador máximo verosímil de  $\mathbb{R}$ ,  $\hat{\mathbb{R}}$ , es

$$\hat{\mathbb{R}} = \hat{L}_Z \hat{L}'_Z + \hat{\Psi}_Z.$$

En consecuencia las comunidades son estimadas por

$$\hat{h}_i^{*2} = \hat{l}_{i1}^{*2} + \dots + \hat{l}_{im}^{*2}, \quad i = 1, \dots, p$$

denotando  $\hat{L}_z = (\hat{l}_{ij}^*)$ .

La proporción de varianza muestral (estandarizada) correspondiente (o explicada) por el  $j$ -ésimo factor viene dada por:

$$\frac{\hat{l}_{1j}^{*2} + \dots + \hat{l}_{pj}^{*2}}{p}$$

en donde  $\hat{l}_{ij}^*$  designa los elementos de  $\hat{L}_z$ .

**Nota 5.** En la práctica generalmente se actúa sobre los datos tipificados, es decir, que se hace Análisis Factorial sobre la matriz de correlaciones. Puede reconstruirse la estimación máximo-verosímil para  $R$  en lugar de para  $\Sigma$ . Se actúa pues con la matriz de correlaciones muestrales  $R$  que se “inserta” en  $\mathbb{L}(L, \Psi, \mu)$  mediante la relación:

$$S = \frac{1}{N-1} \sum_{\alpha=1}^N (X_\alpha - \bar{X})(X_\alpha - \bar{X})'; \quad \frac{(N-1)S}{N} = \frac{s}{N} = \frac{A}{N}.$$

### 1.3.5. Ejemplo Stock-prices

Consideramos un ejemplo recogido por Johnson and Wichern (1998) que va a ir resolviendo secuencialmente en cada sección del desarrollo teórico del tema, mostrándose la aplicación de cada método que se propone (extracción de factores, giros, contrastes de hipótesis, etc.).

Se considera la tasa de rendimiento semanal para las acciones de cinco compañías (Allied Chemical, DuPont, Union Carbide, Exxon y Texaco) de la bolsa de Nueva York, estudiada desde Enero de 1975 hasta Diciembre de 1976. La tasa de rendimiento semanal se define como el cociente de la diferencia del precio (cotización) al cierre en el viernes actual menos el precio al cierre el viernes anterior, entre el precio al cierre el viernes anterior. Las observaciones en 100 semanas sucesivas, aparentemente, deben estar distribuidas independientemente, pero la tasa de rendimiento están correlacionadas, como es sabido, las tendencias de las acciones se mueven junto con las condiciones económicas generales.

El problema es 5-dimensional, con las variables aleatorias  $X_i$ ;  $i = 1, \dots, 5$  que significan respectivamente, el índice de rendimiento semanal de las 5 empresas citadas. Y sean  $(x_1, \dots, x_5)'_\alpha$ ;  $\alpha = 1, \dots, 100$ ; las 100 observaciones ( $N = 100$ ) de la tasa de rendimiento semanal para las acciones de Allied Chemical, DuPont, Union Carbide, Exxon y Texaco respectivamente. Las tres primeras pertenecen al sector industrial químico y las dos últimas al sector petrolífero. El vector de medias muestrales es:

$$\bar{x} = [0,0054; 0,0048; 0,0057; 0,0063; 0,0037]'$$

Para homogeneizar los datos, se consideraran los datos tipificados, dados por:

$$z = \left( z_1 = \frac{x_1 - \bar{x}_1}{\sqrt{s_{11}}}, z_2 = \frac{x_2 - \bar{x}_2}{\sqrt{s_{22}}}, \dots, z_5 = \frac{x_5 - \bar{x}_5}{\sqrt{s_{55}}} \right)'$$

La matriz de covarianzas muestrales de los datos tipificados, es decir, la matriz de correlaciones muestrales  $R$ , tiene la forma

$$R = \begin{pmatrix} 1 & 0,577 & 0,509 & 0,387 & 0,462 \\ 0,577 & 1 & 0,599 & 0,389 & 0,322 \\ 0,509 & 0,599 & 1 & 0,436 & 0,426 \\ 0,387 & 0,389 & 0,436 & 1 & 0,523 \\ 0,462 & 0,322 & 0,426 & 0,523 & 1 \end{pmatrix}.$$

Los autovalores y autovectores normalizados correspondientes de  $R$  son:

$$\hat{\lambda}_1 = 2,857; \quad \hat{e}'_1 = [0,464; 0,457; 0,470; 0,421; 0,421].$$

$$\hat{\lambda}_2 = 0,809; \quad \hat{e}'_2 = [-0,240; -0,590; -0,260; 0,526; 0,582].$$

$$\hat{\lambda}_3 = 0,540; \quad \hat{e}'_3 = [-0,612; 0,178; 0,335; 0,541; -0,435].$$

$$\hat{\lambda}_4 = 0,452; \quad \hat{e}'_4 = [0,387; 0,206; -0,662; 0,472; -0,382].$$

Cuadro 1.1: Análisis factorial por componentes principales

Variable	Solución con un factor		Solución con dos factores		
	Estimación del factor loading $F_1$	Varianzas específicas $\hat{\psi}_i = 1 - \hat{h}_i^2$	Estimación de los factor loadings $F_1$	$F_2$	Varianzas específicas $\hat{\psi}_i = 1 - \hat{h}_i^2$
Allied Chemical	0,783	0,39	0,783	-0,217	0,34
DuPont	0,773	0,40	0,773	-0,458	0,199
Union Carbide	0,794	0,37	0,794	-0,234	0,31
Exxon	0,713	0,49	0,713	0,472	0,27
Texaco	0,712	0,49	0,712	0,524	0,22
Proporción acumulada del total de la varianza explicada			0,571	0,733	
0,571					

$$\hat{\lambda}_5 = 0,343; \quad \hat{e}'_5 = [-0,451; 0,676; -0,400; -0,176; 0,385].$$

Como hemos visto el ejemplo contiene la tasa de rendimiento semanal 100 semanas ( $n=100$ ) para 5 empresas distintas ( $p=5$ ). En primer lugar presentamos los factor loadings obtenidos por el método de componentes principales. Tomaremos  $m=1$  y  $m=2$  para obtener la **solución por componentes principales para el modelo factorial ortogonal**. Específicamente, las estimaciones de los factor loadings son los coeficientes de las componentes principales (autovectores de  $R$ ) escalados por la raíz cuadrada de correspondientes autovalor. Los factor loadings estimados, comunidades, varianzas específicas y la proporción real de varianza explicada por cada factor para  $m = 1$  y  $m = 2$  están en la tabla 1.1. Por ejemplo, para  $m = 2$ , la comunidad asociada a la primera variable (Allied Chemical),  $\hat{h}_1^2$ , es:  $\hat{h}_1^2 = \hat{l}_{11}^2 + \hat{l}_{12}^2 = 0,783^2 + (-0,217)^2 = 0,66$ .

La matriz residual de este ejemplo, para  $m = 2$ , es:

$$R - \hat{L}\hat{L}' - \hat{\Psi} = \begin{pmatrix} 0 & -0,127 & -0,164 & -0,069 & 0,017 \\ -0,127 & 0 & -0,122 & -0,055 & 0,012 \\ -0,164 & -0,122 & 0 & -0,019 & -0,017 \\ -0,069 & 0,055 & -0,019 & 0 & -0,232 \\ 0,017 & 0,012 & -0,017 & -0,232 & 0 \end{pmatrix}$$

Evidentemente, la proporción de varianza total explicada por el modelo con dos factores es bastante mayor que el modelo con un factor, 0,733 frente a 0,571. Sin embargo, para  $m = 2$ ,  $\hat{L}\hat{L}'$  produce valores relativamente grandes de algunos elementos de la matriz residual, especialmente ( $r_{45} = -0,232$ ).

Parece bastante claro que el primer factor,  $F_1$  representa las condiciones económicas generales y podría ser llamado el “factor de mercado”. Todas las empresas tienen cargas altas en este factor y relativamente parecidas. El segundo factor, contrasta las empresas químicas de las petrolíferas (las químicas están relativamente cargadas negativamente y las petrolíferas positivamente). Por ello,  $F_2$  parece diferenciar el tipo de industria, sería un “factor industrial”.

A continuación vamos a estimar (extraer) los factores por el **método de máxima verosimilitud**. En este caso vamos a realizarlo únicamente para  $m = 2$ . Los resultados se recogen en tabla 1.2.

Cuadro 1.2: Análisis factorial por máxima verosimilitud

Variables	Estimación de los factor loadings		Varianzas específicas $\hat{\psi}_i = 1 - \hat{h}_i^2$
	$F_1$	$F_2$	
Allied Chemical	0,684	0,189	0,50
DuPont	0,694	0,517	0,25
Union Carbide	0,681	0,248	0,47
Exxon	0,621	-0,073	0,61
Texaco	0,792	-0,442	0,18
Proporción acumulada del total de la varianza explicada	0,485	0,598	

La matriz residual correspondiente, es:

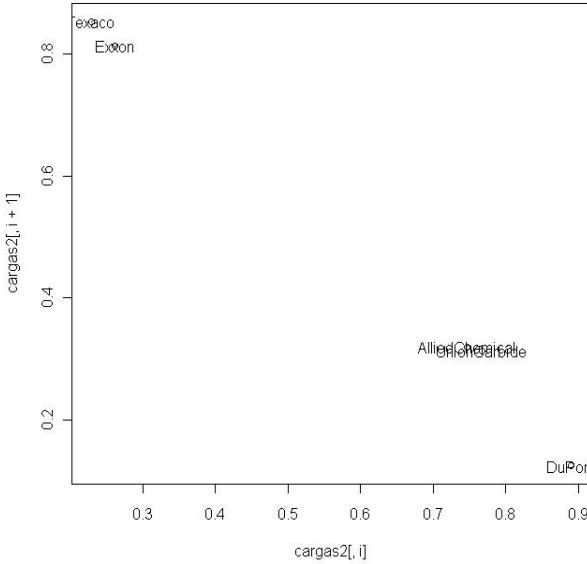
$$R - \hat{L}\hat{L}' - \hat{\Psi} = \begin{pmatrix} 0 & 0,005 & -0,004 & -0,024 & -0,004 \\ 0,005 & 0 & -0,003 & -0,004 & 0,000 \\ -0,004 & -0,003 & 0 & 0,031 & -0,004 \\ -0,024 & -0,004 & 0,031 & 0 & -0,000 \\ -0,004 & 0,000 & -0,004 & -0,000 & 0 \end{pmatrix}.$$

En este caso los elementos de  $R - \hat{L}\hat{L}' - \hat{\Psi}$  son mucho más pequeños que los mismos de la matriz correspondiente al método de componentes principales. Por esto, preferiremos el método de máxima verosimilitud.

Obsérvese que la proporción acumulada del total de la varianza explicada por dos factores, es mayor en el método de componentes principales (0,733) que en el de máxima verosimilitud (0,598). Esto no es una sorpresa, ya que este criterio suele ser favorable en componentes principales. Este hecho no es nada extraño, porque sabemos que si se actúa en Análisis Factorial mediante la metodología de Componentes Principales, se está primando la “explicación” de las varianzas sobre la explicación de todas las varianzas-covarianzas que es objetivo más propio del Análisis Factorial, que es mejor recogida en el método de máxima-verosimilitud.

Prestando atención a la solución de máxima verosimilitud, vemos que todas las variables tienen cargas positivas en el primer factor  $F_1$ . Al igual que en el caso anterior, interpretaremos este factor como el “factor de mercado”. La interpretación del segundo factor, sin embargo, no es tan clara como en la solución del método de componentes principales; parece que indica un factor de “tipo de industria”, pero en este caso las magnitudes de los factores son pequeñas y se podría identificar este factor como comparación (enfrentadas) de DuPont frente a Texaco. Por tanto, para obtener una interpretación más clara de este segundo factor tendremos que rotarlos. Recuérdese además, que cuando se utiliza el método de máxima-verosimilitud, hemos tenido que imponer la restricción de ser diagonal la matriz  $\hat{L}'\hat{\Psi}^{-1}\hat{L}$ . Esta restricción hace que la solución inicial obtenida por máxima-verosimilitud, en la práctica “enmascare” los factores, por lo que es prácticamente obligatoria rotarlos para poner en evidencia factores rotados que conduzcan a una interpretación más clara (ver posteriormente la sección de rotaciones).

Figura 1.1: Rotación



#### 1.4. La rotación de factores en el Análisis Factorial. Criterios gráficos y analíticos (Kaiser-Varimax)

La “propiedad” del Análisis Factorial, consistente en el hecho de que una matriz de factor loadings  $L$  y una matriz de factor loading “rotada” mediante una matriz ortogonal cualquiera  $T$ , reproducen la misma estructura factorial, con dos representaciones estructurales de  $\Sigma$  indistinguibles, va a ser utilizada, paradójicamente como elemento práctico de gran interés a la hora de “aclarar” la estructura factorial de un conjunto de datos.

En efecto, cuando  $m > 1$ , si  $T$  es tal que  $TT' = T'T = I$ ,

$$X - \mu = LF + \epsilon = LTT'F + \epsilon = L^*F^* + \epsilon, \text{ con } L^* = LT; F^* = T'F$$

y se tiene que:

$$E[F^*] = T'E[F] = 0 \text{ y } Cov[F^*] = T'Cov[F]T = T'T = I$$

de modo que, en efecto es imposible sobre la base de las observaciones del  $X$  distinguir  $L$  de  $L^*$ . O dicho de otra manera, los factores  $F$  y  $F^*$  son estadísticamente iguales (en sus propiedades) y estructuran o generan la misma matriz de covarianzas  $\Sigma$  ya que:

$$\Sigma = LL' + \Psi = LTT'L + \Psi = L^*L^{*''} + \Psi.$$

Ahora bien, ¿qué ocurre con las estimaciones cuando se giran los factores, es decir, cuando la matriz de factor loading se rota mediante una matriz ortogonal  $T$ ? Si, por cualquier método (componentes principales, máxima verosimilitud, etc.) se ha estimado la matriz de factor loading  $\hat{L}$ , entonces  $\hat{L}T = \hat{L}^*$  con  $TT' = I$ , es la matriz rotada de factor loadings. Cuando esta operación se efectúa permanecen invariantes las herramientas fundamentales del Análisis Factorial, resumidas en la siguiente tabla:

Matriz estimada de covarianza	$\underbrace{\hat{L}\hat{L}'}_{\hat{\Sigma}} + \hat{\Psi} \xrightarrow{\hat{L} \rightarrow \hat{L}T} \underbrace{\hat{L}^*\hat{L}^{*'}}_{\hat{\Sigma}} + \hat{\Psi}$
Matriz residual	$S_n - \hat{L}\hat{L}' - \hat{\Psi} \xrightarrow{\hat{L} \rightarrow \hat{L}T} S_n - \hat{L}^*\hat{L}^{*'} - \hat{\Psi}$
Varianza específicas	$\hat{\Psi}_i \xrightarrow{\hat{L} \rightarrow \hat{L}T} \hat{\Psi}_i$
Comunalidades	$\hat{h}_i^2 \xrightarrow[\hat{L} \rightarrow \hat{L}T]{\hat{L}\hat{L}'=\hat{L}T\hat{T}'\hat{L}'=\hat{L}^*\hat{L}^{*'}} \hat{h}_i^{*2} = \hat{h}_i^2$

Estas “cualidades” de la rotación de los factores y de las matrices de los loadings, son aprovechables para obtener una “estructura de factores más simple” que sea más fácilmente interpretable.

¿Qué podemos entender por una estructura más simple o un modelo simple de loadings (pattern-loadings)? Pues una estructura de factores que tienda a parecerse a la situación ideal en la que cada variable esté muy pesada con un solo factor y con los restantes está poco o moderadamente ponderada.

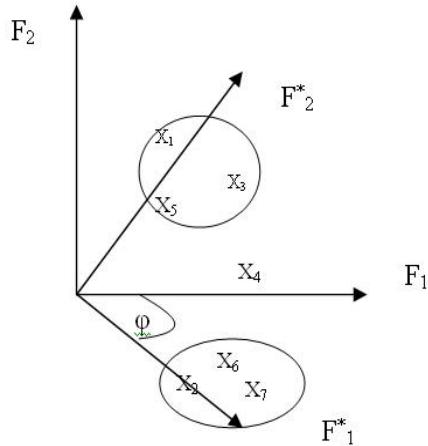
El problema que surge entonces es el siguiente ¿cómo determinar el giro que conduce a una estructura factorial, con un modelo de loadings asociado, lo más simple posible en cada caso concreto? Pues hay métodos gráficos y analíticos para resolver esta cuestión.

#### 1.4.1. Métodos gráficos en la rotación de factores

Cuando  $m = 2$  o los factores comunes se consideran dos a dos, la transformación ortogonal más adecuada se puede obtener gráficamente. Si interpretamos los factores comunes incorrelados como vectores unitarios a lo largo de ejes perpendiculares, y dibujamos los factor loadings respecto de un factor  $(\hat{l}_{i1}; \hat{l}_{i2})$ , obtendremos p puntos, y cada uno de estos puntos corresponde a una variable

$$X_i \longrightarrow (\hat{l}_{i1}; \hat{l}_{i2})$$

Figura 1.2: Variables vs factores



Entonces, en muchos casos es posible elegir un ángulo  $\phi$  de rotación adecuado de modo que los nuevos factor loadings  $\hat{l}_{ij}^*$  se obtendrán así:  $\hat{L}_{p \times 2}^* = \hat{L}_{p \times 2} T_{2 \times 2}$  con  $T$  de la forma:

$$T = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix}, \text{(sentido del reloj)}$$

$$T = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}, \text{(sentido contrario al del reloj.)}$$

Estos métodos gráficos, con el cálculo de  $T$ , no suelen considerarse. Se actúa “a ojo” normalmente, definiendo agrupamientos (cluster) de variables que tengan claramente un factor común.

Si  $m > 2$ , por otra parte, el cálculo de  $\phi$  no es fácil de visualizarse.

Hay casos en que las propias gráficas sugieren cambios de ejes que pasen a ejes oblicuos (no perpendiculares) de tal manera que si esto se hiciera, la interpretación, cabe esperar, que fuese mucho más clara. En Anexo ?? se analizan más detalladamente los criterios analíticos para determinar los giros oblicuos a partir de unos factores

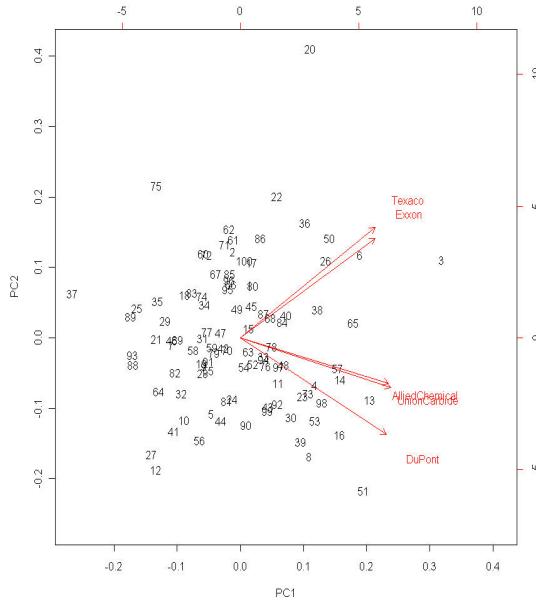
loadings previamente estimados.

En general, dados  $m$  factores, el punto de coordenadas

$$(\hat{l}_{i1}, \hat{l}_{i2}, \dots, \hat{l}_{im})$$

representa la posición de la  $i$ -ésima variable en un espacio de factores. Definidos cluster de variables, no superpuestos, es posible que un giro no rígido, oblicuo, hiciera pasar por los cluster a los ejes factoriales, consiguiéndose, previsiblemente, una mejor interpretación de los factores resultantes.

Figura 1.3: Factor scores y variables



#### 1.4.2. Métodos analíticos: el criterio de Kaiser (Criterio Varimax)

Veamos un primer ejemplo de criterio para encontrar analíticamente una matriz  $T$  ortogonal adecuada para obtener unos factores interpretables. Posteriormente, en sección 1.5 se tratarán pormenorizadamente otros criterios analíticos de búsqueda del giro adecuado. Suponemos pues, giros ortogonales (es decir,  $T$  ortogonal).

Un método sugerido por Kaiser (1958), para elegir una matriz  $T$  con criterios analíticos se basa en lo siguiente. Definimos:

$$\tilde{l}_{ij}^* = \frac{\hat{l}_{ij}^*}{\hat{h}_i}$$

como los factor loading rotados, escalados por las raíces cuadradas de las comunidades. Y se construye la cantidad:

$$V = \frac{1}{p} \sum_{j=1}^m \left[ \sum_{i=1}^p \tilde{l}_{ij}^{*4} - \frac{\left( \sum_{i=1}^p \tilde{l}_{ij}^{*2} \right)^2}{p} \right]$$

que según Kaiser (criterio Kaiser-Varimax) debe ser maximizada por los factor loadings girados.

Salvo constante de proporcionalidad,  $V$  admite una interpretación clara:  $V$  es proporcional a la suma desde  $j=1$  hasta  $m$  de la varianza de los loadings al cuadrado, para el factor  $j$ ; por tanto, maximizar  $V$  consiste en “separar” lo más posible los cuadrados de los loadings, de tal manera que la estructura final rotada haga máxima dicha “separación”, de manera que en cada columna de la matriz de loadings exista una separación “grande” entre algunos loadings de esa columna, con otros de la misma.

**Nota 6.** El “escalamiento” previo se hace (se parte de  $\tilde{l}_{ij}^* = \frac{\hat{l}_{ij}^*}{\hat{h}_i}$ ) con objeto de dar a las variables con menores communalidades un peso relativo mayor en la determinación de la estructura final. Una vez que se ha determinado  $T$ , los loadings  $\tilde{l}_{ij}^*$  son multiplicados por las communalidades  $\hat{h}_i$  de modo que las communalidades originales se mantienen. •

**Nota 7.** Como cabe esperar, los factor loadings obtenidos por el criterio Kaiser-Varimax aplicado a factor loadings obtenidos por métodos diferentes de estimación (componentes principales, máxima verosimilitud ...) no coinciden en general. También ocurre que el modelo de loadings rotado cambia considerablemente cuando se añade un factor en la rotación. Por ejemplo suele ocurrir que existe un factor dominante simple el cual será “obscurecido” en cualquier rotación ortogonal; en este caso se le puede dejar fijo y rotar los restantes factores. •

**Nota 8.** Se recomienda especialmente la rotación de factores cuando se ha utilizado el método de máxima verosimilitud en la estimación, puesto que se habrá utilizado la condición de unicidad  $\hat{L}'\hat{\Psi}^{-1}\hat{L} = \Delta$  diagonal que tiene su justificación teórica (ver anexo ??) pero que probablemente conduzca a factores difícilmente interpretable. •

## 1.5. Rotación de factores. Criterios analíticos generales

En lo que sigue tratamos la cuestión de búsqueda de una rotación de factores adecuada para conseguir una estructura factorial interpretable óptimamente. Realizamos el análisis a partir de una solución inicial que denotamos  $(L, \Psi)$  obtenida por algún método de estimación (por ejemplo los vistos en las secciones 1.3.2 y 1.3.4 anteriores).

Una vez que se han estimado los factores comunes el Análisis Factorial entra en una fase de interpretación y explotación de los resultados. Como ya hemos indicado, la “indeterminación de la solución factorial” en el sentido de que si  $(L, \Psi)$  es una solución, también lo es el par  $(LT, \Psi)$ , es decir:

$$\begin{aligned} X &= \mu + LF + \epsilon \longrightarrow X = \mu + LTT'F + \epsilon = \mu + L^*F^* + \epsilon \\ \Sigma &= LL' + \Psi \xrightarrow[L \rightarrow LT]{} \Sigma = LTT'L + \Psi = L^*L^* + \Psi. \end{aligned}$$

De modo que dicho en otras palabras, podemos encontrar factores comunes “nuevos”, denotados  $F_1^{(R)}, \dots, F_m^{(R)}$  que son combinaciones lineales de  $F_1, \dots, F_m$  componentes del vector  $F = (F_1, \dots, F_m)'$ ; estos nuevos factores son ortogonales, puesto que son  $F^* = T'F$ , con  $T$  ortogonal y tienen varianza unidad (ortogonales=incorrelados).

Cuando  $T$  es ortogonal se habla de rotación ortogonal de factores, para referirse al proceso de obtención de un conjunto de nuevos factores  $F_i^{(R)}$  a partir de los antiguos  $F_i$ ,  $i = 1, \dots, m$ .

El nuevo modelo desarrollado será:

$$X = \mu + L^*F^* + \epsilon = \mu + (LT)(T'F) + \epsilon$$

$$X_i = \mu_i + \sum_{j=1}^m c_{ij} F_j^{(R)} + \epsilon_i \quad i = 1, \dots, p$$

siendo  $c_{ij}$  los elementos de una nueva matriz rotada  $L^* = LT$ . Un cambio ortogonal (rotación sólida) como el aquí hecho, no cambia las comunidades de las variables ya que

$$h_i^2 = \sum_{j=1}^m c_{ij}^2 = \sum_{j=1}^m l_{ij}^2; \quad i = 1, \dots, p$$

siendo  $l_{ij}$  los elementos de la matriz “inicial” de factor loadings.

Por otro lado, la expresión matricial de la transformación de la matriz de factor loadings

$$L \longrightarrow L^* = LT$$

se escribe desglosada y con la notación anterior, como:

$$c_{ij} = \sum_{k=1}^m l_{ik} q_{kj}, \quad i = 1, \dots, p; j = 1, \dots, m; \quad c_{ij} \in L^*; \quad l_{ik} \in L; \quad q_{kj} \in T.$$

Siendo  $q_{kj}$  números, con  $k = 1, \dots, m; j = 1, \dots, m$  y siendo la matriz ortogonal  $T$  de dimensión  $m \times m$ . A efectos de interpretación de los nuevos factores, las constantes  $q_{kj}$  se toman de manera que los loadings rotados sean “simples” respecto de algún criterio.

La siguiente cuestión, es cómo elegir la matriz  $T$ . El criterio que se utiliza es seleccionar una matriz  $T$  ortogonal, tal que los loadings  $c_{ij}$  sean “simples”, en el sentido que “el mayor número posible de ellos, sean próximos a cero y los restantes relativamente grandes”.

La estrategia pues, es buscar una matriz  $T$  ortogonal tal que se obtenga una representación de  $X_i$ ,  $i = 1, \dots, p$  (las variables respuestas), es decir, una representación del tipo  $X_i = \sum_{j=1}^m c_{ij} F_j^{(R)} + \epsilon_i$ , tal que los pesos en pocos factores sean grandes y pequeños en los restantes (Thurstone (1945)). La razón que justifica esta estrategia, es que en la medida en que, en efecto, consigamos una estructura simple, y dado que los loadings expresan, en el Análisis Factorial ortogonal, las correlaciones entre variable y factor, podremos interpretar más fácilmente la estructura factorial obtenida.

### 1.5.1. Métodos de rotación ortogonal

Se han ideado muchos métodos analíticos y gráficos para obtener, mediante rotaciones adecuadas, estructuras más simples a partir de una matriz inicial de factor loadings. Los métodos analíticos, se pueden sistematizar como lo hacen Harman (1967), Afifi and Azen (1979), entre otros, en base al siguiente criterio analítico: Minimizar la función objetivo  $G$ , dada así:

$$G = \sum_{k=1}^m \sum_{\substack{j=1 \\ j=k; j \neq K}}^m \left[ \sum_{i=1}^p c_{ij}^2 c_{ik}^2 - \frac{\gamma}{p} \left( \sum_{i=1}^p c_{ij}^2 \right) \left( \sum_{i=1}^p c_{ik}^2 \right) \right]$$

con  $\gamma \in [0, 1]$ .

1. Cuando  $\gamma = 0$  la rotación se dice **Cuartimax**. En este caso se puede comprobar que minimizar  $G$ , es equivalente a maximizar:

$$\frac{1}{pm} \sum_{j=1}^m \sum_{i=1}^p (c_{ij}^2 - \bar{c}_{..}^2); \quad \bar{c}_{..} = \frac{1}{pm} \sum_{j=1}^m \sum_{i=1}^p c_{ij}^2.$$

Obsérvese que la expresión anterior a maximizar es la varianza de los cuadrados de todos los factor loadings implicados en el problema de Análisis Factorial. Si se maximiza esa varianza quiere decir que dispersamos al máximo los valores posibles de los factor loadings, aumentando los mayores y disminuyendo los más próximos a cero, en cumplimiento de la estrategia antes formulada para llegar a “estructuras simples”.

2. Otro criterio es tomar  $\gamma = 1$ . Este criterio se denomina **Varimax** y en este caso puede comprobarse que la minimización de  $G$  es equivalente a maximizar

$$\frac{1}{p} \sum_{j=1}^m \sum_{i=1}^p (c_{ij}^2 - \bar{c}_{\cdot j}^2)^2; \quad \bar{c}_{\cdot j}^2 = \frac{1}{p} \sum_{i=1}^p c_{ij}^2; \quad j = 1, \dots, m.$$

La diferencia con el cuartimax es la siguiente: Este actúa sobre todos los loadings, en bloque. En cambio el varimax actúa por columnas de loadings ( $\bar{c}_{\cdot j}^2$  es la media de los loadings de cada columna  $j = 1, \dots, m$  correspondiente a cada factor). En el varimax se hace máxima la dispersión de los loadings para cada factor separadamente. Este es el criterio más utilizado.

**Nota 9.** A veces se utiliza la normalización de Kaiser (1958) al aplicar los criterios, en particular el varimax en el caso de rotación ortogonal. Esta normalización consiste en considerar a los loadings  $c_{ij}$ , normalizados respecto de las comunidades a través de la raíz cuadrada de ellas. Es decir, considerar:

$$c_{ij}^* = \frac{c_{ij}}{h_i}; \quad i = 1, \dots, p; \quad j = 1, \dots, m; \quad \text{con } h_i \text{ la raíz cuadrada de la communalidad } i\text{-ésima,}$$

normalizando cada  $c_{ij}$  respecto de la communalidad  $i$ -ésima de la variable  $X_i$ .

La idea que persigue esta normalización es conseguir que cada variable-respuesta  $X_i$  esté contribuida por unos loadings “commensurados” con la communalidad total de dicha variable  $X_i$ , para  $i = 1, \dots, p$ . Se origina, entonces el criterio **Varimax Normalizado** (o de Kaiser) que maximiza la expresión:

$$MaxV = Max \left[ \frac{1}{p} \sum_{j=1}^m \left[ \sum_{i=1}^p c_{ij}^{*4} - \frac{(\sum_{i=1}^p c_{ij}^*)^2}{p} \right] \right].$$

Este criterio Varimax-Kaiser, ya fue considerado en la sección 1.4.2

•

### 1.5.2. Ejemplo Stock-prices

Los métodos descritos de rotación de factores, pueden aplicarse a unos factores comunes estimados “inicialmente” por algún método de estimación (componentes principales o máxima-verosimilitud). A su vez, las rotaciones pueden realizarse con alguno de los criterios analíticos descritos antes. Por ejemplo, tomando el ejemplo que venimos utilizando (stocks prices/ rendimientos semanales), vamos a realizar una rotación de tipo ortogonal, con criterio **Varimax** para clarificar la estructura de los factor loadings en el caso del Análisis Factorial máximo verosímil y para  $m = 2$ . En este caso los factor loadings (rotados y sin rotar), las varianzas específicas y proporción acumulada de varianza explicada son recogidas en tabla 1.3.

En el modelo sin rotar, veíamos clara la existencia de un primer factor, que identificábamos con el “factor de mercado” y otro segundo factor que no estaba tan claro, ya que podía marcar diferencias entre las industrias petrolíferas y químicas o entre DuPont frente a Texaco. Precisamente por este motivo realizábamos la rotación. Los factores rotados indican que las empresas químicas (Allied Chemical, DuPont y Union Carbide) están más cargadas en el primer factor mientras que las petrolíferas (Exxon y Texaco) tienen mayores cargas en el segundo factor. Por lo tanto, los dos factores rotados nos marcarán diferencias entre los tipos de industrias. Es difícil etiquetar estos factores; el Factor 1 podría representar las fuerzas económicas que causan movimientos conjuntos en la cotización de las empresas químicas. El Factor 2 podría representar las condiciones económicas que afectan a las empresas petrolíferas.

Es importante destacar, cómo el factor  $F_1$  original, que es un factor general, dado que todas las variables están ponderadas casi por igual, tiende a “destrozarse” tras la rotación. Esta es la razón por la que en los casos en que existe un factor “general” es interesante rotar los factores manteniendo fijo uno o varios con dicho carácter general. Esto posible en el software actual sobre Análisis Factorial.

Cuadro 1.3: Rotación ortogonal

Variable	Estimación de los factor loadings Max. Verosimil		Estimación de los factor loadings rotados		Varianzas específicas $\hat{\psi}_i = 1 - \hat{h}_i^2$
	$F_1$	$F_2$	$F_1^*$	$F_2^*$	
Allied Chemical	0,684	0,189	0,601	0,377	0,50
DuPont	0,694	0,517	0,850	0,164	0,25
Union Carbide	0,681	0,248	0,643	0,335	0,47
Exxon	0,621	-0,073	0,365	0,507	0,61
Texaco	0,792	-0,442	0,208	0,883	0,18
Proporción acumulada del total de la varianza explicada	0,485	0,598	0,335	0,598	

## 1.6. Puntuaciones en los factores (Factor Scores)

El interés primordial en el Análisis Factorial es el de estimar los parámetros del Modelo. Pero una vez estimados estos, se aborda el problema de estimar valores de los factores comunes, llamados “**factor scores**”, para cada caso de la muestra observada. Estos valores se utilizan para varios pasos posteriores, por ejemplo para diagnósticos del Modelo o para “entradas” de análisis estadísticos posteriores.

En el Modelo de Análisis Factorial exploratorio que venimos analizando, los factores  $F_j$ , así como los errores  $\epsilon_i$ , son inobservables y aleatorios. Por tanto al hablar de su estimación se entiende que ésta es de naturaleza diferente a la estimación de parámetros.

De hecho por “factor scores” (puntuación en  $F$  para el caso j-ésimo de individuo) entenderemos unos valores  $\hat{f}_j$  tales que:

$$\hat{f}_j = \text{estimación del valor } f_j \text{ que toma } F_j \text{ (caso j-ésimo); } j=1,2,\dots,n$$

en donde  $\hat{f}_j$  y  $F_j$  son vectores ( $m \times 1$ ), siendo m el número de factores comunes.

La dificultad de obtener estas estimaciones reside, además, en el hecho de que las cantidades  $f_j$  y  $\epsilon_j$ , inobservables, exceden en número a las observaciones  $x_j$ .

Para obviar esta dificultad, se han ideado varios métodos como el de **Mínimos Cuadrados Ponderados (weighted least squares methods)** o el **método de Regresión**, que son en realidad aproximaciones heurísticas al problema, basadas en planteamientos razonables de tipo estadístico-técnico. Estos dos métodos, tienen dos características comunes:

1. Los factor loadings,  $\hat{l}_i$  y las varianzas específicas  $\hat{\psi}_i$ , estimados, son considerados como los verdaderos valores.
2. Son métodos “robustos” frente a transformaciones de los datos originales del tipo habitualmente utilizado en Análisis Factorial (rotaciones; tipificación, etc.).

A continuación se describen ambos métodos.

### 1.6.1. El método de mínimos cuadrados ponderados

#### ■ Versión teórica.

Supongamos el modelo de Análisis Factorial  $X - \mu = LF + \epsilon$  antes manejado, y supongamos que conocemos  $\mu$ ,  $L$  y las varianzas específicas  $\Psi$ . Sea el vector de errores  $\epsilon = (\epsilon_1, \dots, \epsilon_p)'$ , de modo que  $Var[\epsilon_i] = \psi_i$ ;  $i = 1, \dots, p$ . Bartlett (1947) sugirió estimar los factores comunes por el método de mínimos

cuadrados ponderados (“weighted least squares”) dado que no es preciso que los  $\psi_i$  sean iguales (heterocedasticidad), método que se describe a continuación.

La idea es simple: consideramos la suma de los cuadrados de los errores ponderada por el recíproco de sus varianzas respectivas:

$$\sum_{i=1}^p \frac{\epsilon_i^2}{Var[\epsilon_i]} = \epsilon' \Psi^{-1} \epsilon = (x - \mu - Lf)' \Psi^{-1} (x - \mu - Lf).$$

Bartlett propuso estimar  $f$  por un  $\hat{f}$  tal que haga mínima esa suma ponderada. La solución de este problema de minimización viene dada por:

$$\hat{f}_j = (L' \Psi^{-1} L)_{m \times m}^{-1} L'_{m \times p} \Psi_{p \times p}^{-1} (x - \mu)_{p \times 1}.$$

**Nota 10.** La solución indicada es bien conocida en la teoría del Modelo de Regresión mínimo-cuadrática ponderada. •

#### ■ Versión muestral.

Cuando no se conocen  $\mu$ ,  $L$  y  $\Psi$ , pueden ser sustituidos por sus estimaciones obtenidas a partir de las observaciones de  $X$ , de donde se obtienen los factor scores para cada caso  $j = 1, \dots, n$  (obsérvese que dado el vector de factores  $f$ , los factor scores son “estimaciones” de los valores que dicho vector de factores toma para cada elemento de la muestra  $x_1, \dots, x_n$ ). Así:

$$\hat{f}_{j_{m \times 1}} = (\hat{L}' \hat{\Psi}^{-1} \hat{L})_{m \times m}^{-1} \hat{L}'_{m \times p} \hat{\Psi}_{p \times p}^{-1} (x_j - \bar{x})_{p \times 1}; \quad j = 1, \dots, n$$

con  $\bar{x}$  el vector media muestral y  $x_j$  el elemento  $j$ -ésimo de la muestra, es decir, un vector  $(x_j)_{p \times 1}$ . Obsérvese que lo anterior equivale a tomar como verdaderos valores de  $L$ ,  $\Psi$  y  $\mu$ , sus respectivos estimadores.

El aspecto que toma  $\hat{f}_j$  varía según el método que se haya empleado para ajustar el modelo. Veamos dos casos:

- a) Si se ha empleado el **método de máxima verosimilitud**, se ha de tener en cuenta la condición de unicidad que se ha exigido en este método:  $\hat{L}' \hat{\Psi}^{-1} \hat{L} = \hat{\Delta}$  diagonal. Siendo los factores obtenidos por Mínimos Cuadrados Ponderados, a partir de estimadores máximo verosímiles:

$$\hat{f}_j = (\hat{L}' \hat{\Psi}^{-1} \hat{L})_{m \times m}^{-1} \hat{L}'_{m \times p} \hat{\Psi}_{p \times p}^{-1} (x_j - \hat{\mu}) = \hat{\Delta}^{-1} \hat{L}' \hat{\Psi}^{-1} (x_j - \bar{x}) \quad j = 1, 2, \dots, n.$$

Si se ha actuado con la matriz de correlación, entonces:

$$\hat{f}_j = (\hat{L}'_z \hat{\Psi}_z^{-1} \hat{L}_z)_{m \times m}^{-1} \hat{L}'_z \hat{\Psi}_z^{-1} (z_j) = \hat{\Delta}_z^{-1} \hat{L}'_z \hat{\Psi}_z^{-1} z_j \quad j = 1, 2, \dots, n$$

siendo  $z_j = D^{-1/2} (x_j - \bar{x})$  y  $\hat{\rho} = \hat{L}_z \hat{L}'_z + \hat{\Psi}_z$ .

**Nota 11.** En lo anterior se tiene en cuenta los resultados vistos en la sección 1.3.4, correspondientes al Análisis Factorial sobre datos tipificados, para los que  $\Sigma$  se transforma en  $\rho = R$ . •

Puede probarse fácilmente que los factor scores generados por este procedimiento, tienen media muestral cero (vector media cero), covarianzas muestrales cero y varianzas muestrales unidad. Esto quiere decir que

$$\frac{1}{n} \sum_{j=1}^n \hat{f}_j = 0; \quad \frac{1}{n-1} \sum_{j=1}^n \hat{f}_j \hat{f}'_j = I_{m \times m}.$$

Por otra parte, si se rotan los factores mediante una matriz ortogonal  $T$  de modo que tuviéramos una matriz de factor loadings rotada  $L^*$ , entonces  $L \rightarrow LT = L^* \Rightarrow \hat{L}^* = \hat{L}T$  y entonces los factor scores quedan transformados análogamente  $\hat{f}_j \rightarrow \hat{f}_j^* = T' \hat{f}_j$ .

- b) Si se hubiera empleado el **método de las componentes principales** para la extracción de los factores, las cosas son más simples. En realidad se supone que las  $\psi_i$  son la unidad o valores próximos a la unidad. Esto implica que puede emplearse unos mínimos cuadrados no ponderados (es decir, mínimos cuadrados ordinarios  $LQO$ ). En este caso el planteamiento anterior tipo Bartlett (con  $\Psi = I$ ) conduce a:

$$\hat{f}_j = (\hat{L}' \hat{L})^{-1} \hat{L}' (x_j - \bar{x})$$

o si se actúa con la matriz de correlaciones:

$$\hat{f}_j = (\hat{L}'_z \hat{L}_z)^{-1} \hat{L}'_z (z_j)$$

en donde  $z_j$  son los datos tipificados ( $z_j = D^{-1/2}(x_j - \bar{x})$ ,  $j = 1, \dots, n$ ).

Ahora bien, en este método de componentes principales, como ya se vio;

$$\hat{L}_{p \times m} = \left[ \begin{array}{c|c|c|c} \sqrt{\hat{\lambda}_1} \hat{e}_1 & \dots & \sqrt{\hat{\lambda}_m} \hat{e}_m \\ \hline p \times 1 & & p \times 1 & \end{array} \right]$$

por lo que :

$$\hat{f}_j = (\hat{L}' \hat{L})^{-1} \hat{L}' (x_j - \bar{x}) = \left[ \frac{1}{\sqrt{\hat{\lambda}_1}} (x_j - \bar{x})' \hat{e}_1 \mid \dots \mid \frac{1}{\sqrt{\hat{\lambda}_m}} (x_j - \bar{x})' \hat{e}_m \right]'$$

puesto que:

$$(\hat{L}' \hat{L})^{-1} = \begin{pmatrix} \frac{1}{\hat{\lambda}_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\hat{\lambda}_2} & & 0 \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\hat{\lambda}_m} \end{pmatrix}_{m \times m}$$

dado que  $\hat{e}_1 \dots \hat{e}_m$  son los primeros autovectores ortogonales y unitarios que proporcionan las  $m$  primeras componentes principales utilizadas en el método de extracción de factores por el método de componentes principales. Pudiéndose comprobar que:

$$\frac{1}{n} \sum_{j=1}^n \hat{f}_j = 0; \quad \frac{1}{n-1} \sum_{j=1}^n \hat{f}_j \hat{f}_j' = I.$$

**Nota 12.** Obsérvese que los  $\hat{f}_j$  obtenidos, no son más que los valores de las  $m$  primeras componentes principales evaluadas en  $x_j$  con el factor de escala  $\frac{1}{\sqrt{\hat{\lambda}_i}}$ . •

### 1.6.2. El método de regresión

#### ■ Versión teórica.

Sea otra vez el modelo original de Análisis Factorial  $X - \mu = LF + \epsilon$ , con parámetros conocidos. Y coloquémonos bajo hipótesis de Normalidad. Ello quiere decir que tanto  $F$  como  $\epsilon$ , son vectores normales multivariantes, con medias y matrices de covarianzas las correspondientes. Es decir que  $F \rightsquigarrow \mathcal{N}_m(0; I)$  (caso ortogonal);  $\epsilon \rightsquigarrow \mathcal{N}_p(0; \Psi)$ . Entonces, es claro que:

$$X - \mu = LF + \epsilon \rightsquigarrow \mathcal{N}_p[0; LL' + \Psi]$$

dado que:  $LF \rightsquigarrow \mathcal{N}_m(0; LL')$  y según las hipótesis  $F$  y  $\epsilon$  son incorreladas y por tanto independientes por ser Normales ambos.

Además, la distribución conjunta de  $X - \mu$  y  $F$  es una  $\mathcal{N}_{m+p}(0; \Sigma^*)$  en donde, en virtud de las hipótesis del modelo de Análisis Factorial ortogonal que estamos considerando:

$$\Sigma^* = \left( \begin{array}{c|c} \Sigma_{p \times p} = LL' + \Psi & L_{p \times m} \\ \hline L'_{m \times p} & I_{m \times m} \end{array} \right)_{(m+p) \times (m+p)} \left( = \left( \begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array} \right) \right).$$

Podemos considerar, entonces, en este ambiente normal, la distribución condicionada de los factores al vector  $x : F/X = x$ . Y en particular la correspondiente media condicionada que será:

$$E[F/X = x] = L'\Sigma^{-1}(x - \mu) = L'(LL' + \Psi)^{-1}(x - \mu) \quad (\Sigma_{21}\Sigma_{11}^{-1}(x - \mu))$$

y como también sabemos:

$$Cov[F/X = x] = I - L'\Sigma^{-1}L = I - L'(LL' + \Psi)^{-1}L \quad (\Sigma_{22 \cdot 1}).$$

**Nota 13.** Esta última es la que al estudiar el modelo de regresión lineal múltiple llamamos  $\Sigma_{22 \cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ . Como también sabemos de la regresión múltiple, la matriz  $L'(LL' + \Psi)^{-1}$  no es otra cosa que la matriz de coeficientes de regresión en la regresión lineal múltiple de los factores  $F$  sobre las variables  $X$  (superficie de la regresión lineal múltiple de  $F$  sobre  $X = x$ ). •

#### ■ Versión muestral.

La idea entonces, a efectos de estimar los factor scores, es aprovechar estos resultados teóricos, vía regresión múltiple de  $F$  sobre las variables, de modo que si disponemos de estimadores máximo-verosímiles de  $L$  y  $\Psi$ , podemos tomarlos como verdaderos valores y entonces, podemos estimar  $L'(LL' + \Psi)^{-1}$ , mediante  $\hat{L}'(\hat{L}\hat{L}' + \hat{\Psi})^{-1}$ , y podemos tomar como factor scores los valores dados por la correspondiente regresión de  $F$  sobre  $X$  estimada, en definitiva:

$$\hat{f}_j = E[\widehat{f/X} = x] = \hat{L}'(\hat{L}\hat{L}' + \hat{\Psi})^{-1}(x_j - \bar{x}) \quad j = 1, \dots, n.$$

(hiperplano estimado de regresión lineal múltiple de  $F$  sobre  $X$ ).

#### 1.6.3. Comparación de los “factor scores” dados por los dos métodos

Los factor scores dados por estos dos métodos (LS y R) pueden compararse en general, si se tienen en cuenta lo siguiente:

$$\hat{L}'(\hat{L}\hat{L}' + \Psi)^{-1} = (I + \hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1}\hat{L}'\hat{\Psi}^{-1}$$

lo que permite simplificar el cálculo de las  $\hat{f}_j$  en el método de regresión descrito antes (sección 1.6.2). Además ello permite también comparar los factor scores generados por ambos métodos. Se tiene

$$\hat{f}_j = (\hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1}\hat{L}'\hat{\Psi}^{-1}(x_j - \bar{x}) \longrightarrow \hat{f}_j^{LS} \text{ mínimos cuadrados}$$

$$\hat{f}_j = (I + \hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1}\hat{L}\hat{\Psi}^{-1}(x_j - \bar{x}) \longrightarrow \hat{f}_j^R \text{ regresión.}$$

Es claro entonces que:

$$\hat{f}_j^{LS} = (\hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1}(I + \hat{L}'\hat{\Psi}^{-1}\hat{L})\hat{f}_j^R = (I + (\hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1})\hat{f}_j^R.$$

**Nota 14.** A su vez podemos proceder a analizar esta relación, en función del método que se hubiera seguido en el caso de mínimos cuadrados, de tal forma que si se hubiera adoptado el de máxima verosimilitud, para la estimación de factores, entonces:

$$(\hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1} = \hat{\Delta}^{-1}$$

ya que habríamos de utilizar la condición de unicidad  $L'\Psi^{-1}L = \Delta$  (matriz diagonal). De tal manera entonces que si  $\hat{\Delta}^{-1}$  es tal que sus elementos (de la diagonal principal) fueran próximos a cero, entonces:

$$\hat{f}_j^{LS} \rightarrow \hat{f}_j^R.$$

Es decir, ambos métodos darían factor scores “parecidos”. •

**Nota 15.** Se suele en la práctica del Análisis Factorial, actuar con la matriz  $S$  de covarianzas muestrales iniciales, en lugar de con  $\hat{\Sigma} = \hat{L}\hat{L}' + \hat{\Psi}$ , que es fruto de la estimación del modelo de Análisis Factorial con un determinado número de factores  $m$ . Así se pretende en la práctica, reducir el efecto que pudiera tener una posiblemente incorrecta determinación del número de factores. Se calcula entonces:

$$\hat{f}_j = \hat{L}'S^{-1}(x_j - \bar{x}) \quad j = 1, \dots, n.$$

De modo que resumiendo:

i) Los factor scores obtenidos por regresión, basados en  $S$  son:

$$\hat{f}_j = \hat{L}'S^{-1}(x_j - \bar{x}), \quad j = 1, 2, \dots, n.$$

ii) Los factor scores obtenidos por regresión y datos tipificados son:

$$\hat{f}_j = \hat{L}'_z R^{-1}(z_j); \quad z_j = D^{-1/2}(x_j - \bar{x})$$

con  $z_j = D^{1/2}(x_j - \bar{x})$  y  $R$  la matriz de correlaciones muestrales.

Es evidente, que si se rotan los factores, se rotan los factor scores:

$$\hat{f}_j \rightarrow \hat{f}_j^* = T'\hat{f}_j = T'\hat{L}'S^{-1}(x_j - \bar{x})$$

**Nota 16.** Los autores y prácticos de Análisis Factorial, indican que no existe unas superioridad uniforme de un método sobre otro (regresión y mínimos cuadrados). Cada caso de aplicación debe ser analizado “ad hoc”.

Un procedimiento práctico de efectuar la comparación entre los factor scores proporcionados por los dos métodos, es calcular numéricamente el coeficiente de correlación muestral entre scores de un mismo factor.

Normalmente, los factor scores se calculan sobre los datos transformados (tipificados) y sobre los rotados por algún método de rotación de los ejes factoriales. Por cada combinación de un método de estimación de factores y factor scores, con un método de rotación y con datos tipificados o no, aparece una solución concreta.

•

#### 1.6.4. Los “scatterplots” bivariantes de los factor scores

En diversos puntos antes desarrollados, la hipótesis de normalidad ha jugado un papel primordial (que origina en particular la posibilidad de aplicación de métodos de máxima verosimilitud sobre poblaciones normales). Por tanto, comprobaciones de normalidad o grado de normalidad, tanto para las variables  $X$  como para los factores  $F$  deben ser hechas.

Cuando el número de variables y factores es grande, ya sabemos que comprobar la normalidad es un problema difícil y teóricamente imposible de resolver exactamente. Una herramienta útil en el análisis de la posible normalidad de los factores (y de las variables también) es proceder a representaciones gráficas bivariantes marginales (bivariate scatterplots). Es obvio que contornos elípticos son la “traza” de la normalidad, por ejemplo, de los factores. Lo que se hace pues, es representar dos a dos los factor scores y analizar los contornos elípticos frente a no elípticos que se producen en las diferentes parejas.

Hay que tener en cuenta que si los factor scores, van a usarse como “datos de entrada” para otros análisis multivariantes que exigen normalidad, es de gran interés conocer la posible normalidad de los factor scores.

Por otra parte, queda dicho que estos “bivariate scatterplots” pueden hacerse con los factor scores obtenidos por dos métodos distintos (máxima verosimilitud o componentes principales).

Los scatterplots también son útiles para la detección de posibles outliers.

### 1.6.5. Ejemplo Stock-prices

A continuación, obtendremos los factor scores, por los métodos de mínimos cuadrados y regresión, para el ejemplo del stock-prices. En apartados anteriores, actuando con la matriz  $R$  y con el método máximoverosímil, obtuvimos la estimación de los factor loadings rotados,  $\hat{L}^*$ , y la estimación  $\hat{\Psi}$  que son:

$$\hat{L}^* = \begin{pmatrix} 0,601 & 0,377 \\ 0,850 & 0,164 \\ 0,643 & 0,335 \\ 0,365 & 0,507 \\ 0,208 & 0,883 \end{pmatrix} ; \quad \hat{\Psi} = \begin{pmatrix} 0,5 & 0 & 0 & 0 & 0 \\ 0 & 0,25 & 0 & 0 & 0 \\ 0 & 0 & 0,47 & 0 & 0 \\ 0 & 0 & 0 & 0,61 & 0 \\ 0 & 0 & 0 & 0 & 0,18 \end{pmatrix}.$$

Aplicamos los dos métodos descritos teóricamente antes, es decir, el método de Mínimos Cuadrados Ponderados y el método de Regresión. En este último caso, por ejemplo, la matriz  $\hat{L}'_z R^{-1}$  es

$$\begin{pmatrix} 0,187 & 0,657 & 0,222 & 0,050 & -0,210 \\ 0,037 & -0,185 & 0,013 & 0,107 & 0,864 \end{pmatrix}.$$

Por ejemplo, si se toma la observación tipificada de la segunda semana observada:

$$z' = (0,50; -1,40; -0,20; -0,70; 1,40).$$

La “puntuación” (factor score de  $z$ ) en los dos factores comunes estimados-rotados  $F_1^*$  y  $F_2^*$ , serían, por componentes principales y máxima verosimilitud, respectivamente:

$$\begin{aligned} f &= (\hat{L}'_z \hat{\Psi}_z^{-1} \hat{L}_z^*)^{-1} \hat{L}'_z \hat{\Psi}_z^{-1} z = \begin{pmatrix} -1,8 \\ 2,0 \end{pmatrix} \\ \hat{f} &= \hat{L}'_z R^{-1} z = \begin{pmatrix} 0,187 & 0,657 & 0,222 & 0,050 & -0,210 \\ 0,037 & -0,185 & 0,013 & 0,107 & 0,864 \end{pmatrix} \begin{pmatrix} 0,50 \\ -1,40 \\ -0,20 \\ -0,70 \\ 1,40 \end{pmatrix} = \begin{pmatrix} -1,2 \\ 1,4 \end{pmatrix}. \end{aligned}$$

Se puede observar que las puntuaciones son diferentes según el método de estimación empleado.

Si se partiera de una Análisis Factorial por componentes principales, se utilizarían las matrices  $\hat{L}^*$  y  $\hat{\Psi}^*$ , de factor loadings y unicidades respectivamente, obtenidas por dicho método. Concretamente, las matrices de factor loadings, original ( $\hat{L}$ ) y rotada ( $\hat{L}^*$ ) serían:

$$\hat{L} = \begin{pmatrix} 0,784 & -0,216 \\ 0,773 & -0,458 \\ 0,795 & -0,234 \\ 0,712 & 0,473 \\ 0,712 & 0,524 \end{pmatrix} \quad \text{y} \quad \hat{L}^* = \hat{L}T = \begin{pmatrix} 0,746 & 0,323 \\ 0,889 & 0,128 \\ 0,766 & 0,316 \\ 0,258 & 0,815 \\ 0,226 & 0,854 \end{pmatrix}.$$

## Factor Scores

$X - \mu = LF + \epsilon \rightarrow F_j, j = 1, \dots, n$ ; vector de factores para el caso j-ésimo con n el número de casos  $\rightarrow f_j$  valor de  $F_j$  en el caso j-ésimo e inobservable  $\rightarrow \hat{f}_j$  estimación de  $f_j$ . Donde  $F_j, f_j$  y  $\hat{f}_j$  son vectores ( $m \times 1$ ).

Métodos de calcular la “estimación”, o “puntuación”,  $f_j$  (versión teórica.)

1. Mínimos Cuadrados Ponderados (WLSM).
  - a) Método de Bartlett: Estimar  $f$  por el vector  $\hat{f}$  tal que minimice  $\sum_{i=1}^p \frac{\epsilon_i^2}{Var[\epsilon_i]} = (x - \mu - Lf)' \Psi^{-1} (x - \mu - Lf)$ .
  - b) Solución analítica: estimación del vector de factores  $F$  en la población.
$$\hat{f}_{m \times 1} = (L'_{m \times p} \Psi_{p \times p}^{-1} L_{p \times m})_{m \times m}^{-1} (L'_{m \times p} \Psi_{p \times p}^{-1} (x - \mu)_{p \times 1})_{m \times 1}.$$
2. Método de Regresión (RM), bajo Normalidad Multivariante de  $F$  y  $\epsilon$  (Maxwell-Lawley).

Caso de  $\mu, L$  y  $\Psi$  desconocidos (versión muestral.)

Hay que sustituir  $\mu, L$  y  $\Psi$  por estimaciones proporcionadas por alguno de los métodos de estimación del Análisis Factorial. Llamaremos  $\hat{\mu}, \hat{L}$  y  $\hat{\Psi}$  a estas estimaciones (independientemente del método por el que se obtengan):

$$\hat{f}_j = (\hat{L}' \hat{\Psi} \hat{L})^{-1} \hat{L}' \hat{\Psi}^{-1} (x_j - \bar{x}), \quad j = 1, \dots, n.$$

con  $\bar{x}$  el vector de medias muestrales  $j = 1, \dots, n$  y  $x_j$  el vector j-ésimo de la muestra. Entonces  $\hat{f}_j$  son los factor scores estimados para cada caso  $j = 1, \dots, n$ .

Métodos de obtención de  $\hat{\mu}, \hat{L}$  y  $\hat{\Psi}$ .

1. Máxima Verosimilitud: bajo la condición de unicidad:  $\hat{L}' \hat{\Psi}^{-1} \hat{L} = \hat{\Delta}$  diagonal. A su vez puede emplearse con matriz de covarianzas muestrales o con matriz de correlaciones.
2. Método de Componentes principales:  $\psi_i \approx 1$  (WLSM  $\rightarrow$  LSM) ( $LSM \equiv LQO \equiv \Psi = I \rightarrow \hat{f}_j = (\hat{L}' \hat{L})^{-1} \hat{L}' (x_j - \bar{x})$ ) o  $\hat{f}_j = (\hat{L}'_z \hat{L}_z)^{-1} \hat{L}'_z z_j$  con  $z_j = D^{-1/2} (x_j - \bar{x})$ , según se empleen datos no tipificados o tipificados respectivamente.

## 1.7. Contrastes de Hipótesis en Análisis Factorial

En este apartado vamos a considerar un test clásico del Análisis Factorial ortogonal de Lawley-Maxwell, es decir en un modelo factorial en el que sus elementos aleatorios  $X$ ,  $F$  y  $\epsilon$  se distribuyen según normales multivariantes, y en particular  $X \xrightarrow{d} \mathcal{N}_p(\mu; \Sigma)$ , con  $\Sigma = LL' + \Psi$ . Esta es la situación en la que en la sección 1.3.4 se ha analizado la estimación (extracción) de factores. Obviamente, la construcción de test de hipótesis en dicho contexto se basa, en un enfoque clásico, en la “metodología del cociente de verosimilitudes” y su comportamiento asintótico en distribución bajo conocidas condiciones de regularidad. Esta metodología es viable bajo la hipótesis de normalidad del modelo Factorial. De esta manera se obtiene a continuación el test asintótico de Bartlett (o Bartlett-Lawley) sobre la existencia de un Análisis Factorial con un número dado  $m$  de factores comunes ortogonales. También se considera la obtención de un contraste en Análisis Factorial, basado en la metodología de criterios de información (test de Akaike) que marca una dirección más reciente en la cuestión. En el Anexo ?? de este tema, se hace un estudio pormenorizado de los contrastes clásicos en Análisis Factorial ortogonal (ver Anexo ??) en el Modelo Lawley-Maxwell, obteniéndose los contrastes de Bartlett-Lawley y Lawley-Maxwell sobre el número  $m$  de factores adecuados.

### 1.7.1. Test asintótico (Bartlett) sobre la adecuación de un modelo factorial ortogonal con $m$ factores comunes

Consideramos aquí el contraste asintótico de Bartlett-Lawley en un Análisis Factorial ortogonal de Lawley-Maxwell. Lo hacemos de manera resumida. Su estudio detallado se realiza en el citado Anexo ??, junto con su expresión alternativa (test de Lawley-Maxwell).

La aceptación de la hipótesis de una población normal nos lleva directamente a los test de hipótesis sobre la adecuación del modelo. Supongamos que el modelo tiene  $m$  factores comunes; en estas circunstancias,  $\Sigma = LL' + \Psi$  y el test de adecuación del modelo con  $m$  factores comunes es equivalente a contrastar que:

$$H_0 : \Sigma_{p \times p} = L_{p \times m} L'_{m \times p} + \Psi_{p \times p} \quad (1.1)$$

frente a  $H_1 : \Sigma$  es otra cualquiera matriz definida positiva. Cuando  $\Sigma$  no tiene ninguna forma especial, el máximo de la función de verosimilitud (con  $\hat{\Sigma} = \frac{N-1}{N} S = S_N = \frac{A}{N}$ ) es proporcional a

$$|\hat{\Sigma}|^{-N/2} \exp\left(-\frac{Np}{2}\right).$$

Concretamente, la verosimilitud maximizada, bajo la hipótesis  $H_1$  (alternativa) tiene la forma:

$$\mathbb{L}\left(\bar{X}; \frac{A}{N}\right) = \exp\left(-\frac{pN}{2} \ln(2\pi)\right) \left|\frac{A}{N}\right|^{-N/2} \exp\left(-\frac{pN}{2}\right)$$

(ver Anexo ??).

Bajo la hipótesis nula,  $H_0$ ,  $\Sigma$  está restringida a tener la forma de la ecuación 1.1. En este caso, el máximo de la verosimilitud (con  $\hat{\mu} = \bar{x}$  y  $\hat{\Sigma} = \hat{L}\hat{L}' + \hat{\Psi}$ , donde  $\hat{L}$  y  $\hat{\Psi}$  son los estimadores de máxima verosimilitud de  $L$  y  $\Psi$ ) es proporcional a

$$\begin{aligned} |\hat{\Sigma}| \exp\left(-\frac{1}{2} \text{tr}(\hat{\Sigma}^{-1} A)\right) &= |\hat{L}\hat{L}' + \hat{\Psi}|^{-N/2} \exp\left(-\frac{1}{2} \text{tr}(N\hat{\Sigma}^{-1} S_N)\right) = \\ &= |\hat{L}\hat{L}' + \hat{\Psi}|^{-N/2} \exp\left(-\frac{N}{2} \text{tr}(\hat{\Sigma}^{-1} S_N)\right). \end{aligned}$$

Concretamente (véase Anexo ??) la verosimilitud maximizada es

$$\underset{A.\text{Factorial}}{\text{Max}} \mathbb{L} = \exp\left(-\frac{pN}{2} \ln(2\pi)\right) |\hat{\Psi} + \hat{L}\hat{L}'|^{-N/2} \exp\left(-\frac{pN}{2}\right).$$

En consecuencia, siguiendo la metodología del “cociente de verosimilitudes (multivariante de Wilks)” y calculados los dos máximos del cociente de verosimilitudes, éste se expresa finalmente así:

$$\frac{\underset{A, \text{Factorial}}{\text{Max}} \mathbb{L}}{\underset{\mu; \Sigma}{\text{Max}} \mathbb{L}} = \frac{\left| \frac{A}{N} \right|^{N/2}}{\left| \hat{\Psi} + \hat{L}\hat{L}' \right|^{N/2}} = \left( \frac{\left| \frac{A}{N} \right|}{\left| \hat{\Psi} + \hat{L}\hat{L}' \right|} \right)^{N/2} = W.$$

Por tanto, según el método de construcción de test de hipótesis basado en el cociente de verosimilitudes (procedimiento habitual en hipótesis nulas y alternativas compuestas con estimadores obtenidos bajo máxima verosimilitud), “se ha de tender a rechazar  $H_0$  (en este caso la existencia de un Análisis Factorial con  $m$  factores) en la medida en que  $W$  sea pequeño”. La cuestión es como definir “valores pequeño” de  $W$ , en términos estadísticos. Habría que conocer para ello la distribución exacta de  $W$ , que no es posible. Como es sabido, existe un resultado asintótico para la distribución de  $W$  en el contexto aquí supuesto, de Normalidad multivariante (test de la  $\chi^2$ ). Para ello consideraremos  $-2 \ln W$  es decir:

$$-2 \ln W = -2 \frac{N}{2} \ln \left( \frac{\left| \frac{A}{N} \right|}{\left| \hat{\Psi} + \hat{L}\hat{L}' \right|} \right) = -N \ln \left( \frac{\left| \frac{A}{N} \right|}{\left| \hat{\Psi} + \hat{L}\hat{L}' \right|} \right) = N \ln \left( \frac{\left| \hat{\Psi} + \hat{L}\hat{L}' \right|}{\left| \frac{A}{N} \right|} \right)$$

que es tal que se comporta (test de Bartlett), bajo la  $H_0$  de existencia de Análisis Factorial, así:

$$\left[ N - 1 - \frac{2p + 4m + 5}{6} \right] \ln \left( \frac{\left| \hat{\Psi} + \hat{L}\hat{L}' \right|}{\left| \frac{A}{N} \right|} \right) \xrightarrow[N \rightarrow \infty]{} \chi_{v-v_0}^2$$

en donde se ha sustituido  $N$  por la expresión indicada (aproximación sugerida por Bartlett (1947)) y en donde  $v - v_0 = \frac{1}{2}p(p+1) - [p(m+1) - m(m-1)/2] = \frac{1}{2}[(p-m)^2 - p - m] = d$  es decir: El número de parámetros (elementos) de  $\Sigma$ ,  $\frac{p(p+1)}{2}$ , más el número de restricciones identificadas,  $\frac{m(m-1)}{2}$ , menos el número de parámetros en  $L$  y  $\Psi$ ,  $pm + p = p^2(m+1)$ .

Obsérvese que hemos sustituido: “rechazar  $H_0$  para valores pequeños de  $W$ , por rechazar  $H_0$  para valores grandes de  $-2 \ln W$ ”. Por tanto el test final obtenido puede formularse en los siguientes términos:

Se rechaza  $H_0$  (existencia de un Análisis Factorial con  $m$  factores comunes), al nivel de significación  $\alpha$ , si:

$$\left[ N - 1 - \frac{2p + 4m + 5}{6} \right] \ln \left( \frac{\left| \hat{\Psi} + \hat{L}\hat{L}' \right|}{\left| \frac{A}{N} \right|} \right) > \chi_{\frac{(p-m)^2-p-m}{2}}^2(\alpha).$$

En este test aproximado, como los grados de libertad han de ser positivos en una  $\chi^2$ , es claro que el número de factores comunes  $m$ , ha de verificar:

$$m < \frac{1}{2} [2p + 1 - \sqrt{8p + 1}]$$

es decir,  $(p-m)^2 - p - m > 0$ .

**Nota 17.** al implementar el test, estamos contrastando la adecuación de los  $m$  factores del modelo comparando las varianzas generalizadas,  $|\hat{L}\hat{L}' + \hat{\Psi}|$  y  $|S_N|$ . Si  $N$  es grande y  $m$  es pequeño en relación a  $p$ , la hipótesis  $H_0$  será normalmente rechazada, indicándonos ello la necesidad de tomar más factores. Sin embargo, si  $\hat{\Sigma} = \hat{L}\hat{L}' + \hat{\Psi}$  puede ser aproximarse lo suficiente a  $S_N$  de tal manera que aumentar el número de factores no produce ninguna mejora, aunque estos con los que se aumenta, fueran significativos. •

Veamos a continuación la forma que toma el test de Bartlett-Lawley, cuando los datos están “tipificados”. Ya se vió en la sección 1.3.4 la forma que adoptan los E.M.V.,  $\hat{L}$  y  $\hat{\Psi}$  cuando se efectúa un tal cambio de tipificación. En efecto el estadístico del test basado en el cociente de varianzas generalizadas será:

$$\frac{|\hat{L}\hat{L}' + \hat{\Psi}|}{|S_N|}.$$

Sea  $\hat{V}^{-1/2}$  la matriz diagonal tal que  $\hat{V}^{-1/2}S_N\hat{V}^{-1/2} = R$ . Por las propiedades de los determinantes:

$$|\hat{V}^{-1/2}| |\hat{L}\hat{L}' + \hat{\Psi}| |\hat{V}^{-1/2}| = |\hat{V}^{-1/2}\hat{L}\hat{L}'\hat{V}^{-1/2} + \hat{V}^{-1/2}\hat{\Psi}\hat{V}^{-1/2}|$$

y

$$|\hat{V}^{-1/2}| |\hat{L}\hat{L}' + \hat{\Psi}| |\hat{V}^{-1/2}| = |\hat{V}^{-1/2}S_N\hat{V}^{-1/2}|$$

por tanto

$$\frac{|\hat{L}\hat{L}' + \hat{\Psi}|}{|S_N|} = \frac{|\hat{V}^{-1/2}|}{|\hat{V}^{-1/2}|} \frac{|\hat{L}\hat{L}' + \hat{\Psi}|}{|S_N|} \frac{|\hat{V}^{-1/2}|}{|\hat{V}^{-1/2}|} = \frac{|\hat{V}^{-1/2}\hat{L}\hat{L}'\hat{V}^{-1/2} + \hat{V}^{-1/2}\hat{\Psi}\hat{V}^{-1/2}|}{|\hat{V}^{-1/2}S_N\hat{V}^{-1/2}|} = \frac{|\hat{L}_z\hat{L}'_z + \hat{\Psi}_z|}{|R|}.$$

Para muestras grandes, las varianzas y covarianzas para los estimadores máximo verosímiles de  $\hat{l}_{ij}, \hat{\psi}_{ij}$  han sido obtenidos cuando estos estimadores han sido determinados por la matriz de covarianzas  $S$ . Estas expresiones son en general bastante complicadas. Un análisis detallado de la Inferencia Estadística sobre  $L$  y  $\Psi$  en un modelo de Análisis Factorial y en otros modelos más generales puede verse en el Anexo ??.

### 1.7.2. Ejemplo Stock-prices

Aplicamos el test de Bartlett-Lawley deducido. Teniendo en cuenta que los datos están tipificados aplicamos dicho test en términos de  $R$  (matriz de correlaciones muestrales). Por tanto, utilizaremos los estimadores de  $\hat{L}_z, \hat{\Psi}_z$  obtenidos antes (véase sección 1.3.5), se tiene:

$$\frac{|\hat{L}_z\hat{L}'_z + \hat{\Psi}_z|}{|R|} = \frac{\begin{vmatrix} 1 & & & & \\ 0,572 & 1 & & & \\ 0,513 & 0,602 & 1 & & \\ 0,411 & 0,393 & 0,405 & 1 & \\ 0,458 & 0,322 & 0,430 & 0,523 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & & & & \\ 0,577 & 1 & & & \\ 0,509 & 0,599 & 1 & & \\ 0,387 & 0,389 & 0,436 & 1 & \\ 0,462 & 0,322 & 0,426 & 0,523 & 1 \end{vmatrix}} = \frac{0,194414}{0,193163} = 1,0065.$$

Usando el test corregido de Bartlett-Lawley, con  $S_N = R$ , se tiene:

$$(N - 1 - (2p + 4m + 5)/6) \ln \frac{|\hat{L}\hat{L}' + \hat{\Psi}|}{|S_N|} = \left[ 100 - 1 - \frac{10 + 8 + 5}{6} \right] \ln 1,0065 = 0,62$$

y como  $\frac{1}{2}[(p - m)^2 - p - m] = \frac{1}{2}[(5 - 2)^2 - 5 - 2] = 1$ , entonces, el valor de la  $\chi^2_{1;(0,05)} = 3,84$ , por lo que aceptamos la hipótesis nula. Por lo que concluimos que el modelo con dos factores es aceptable. De hecho, vemos que  $P[\chi^2_{1;(0,05)} > 62] = 0,43$ , lo que implica que no será rechazada la hipótesis para niveles razonables.

### 1.7.3. Criterios basados en la información estadística. El criterio de Akaike (AIC) para la elección de un modelo factorial con r-factores comunes

Recientemente, se ha investigado la utilización de criterios basados en estadísticos de información basados en la Entropía para estimar el número de factores explicatorios que pueden extraerse de una matriz de covarianzas. La razón principal que motiva la utilización de estos criterios es que la estimación clásica de factores por máxima verosimilitud tiende frecuentemente a elegir un número mayor de factores (por ejemplo, en Análisis Factorial), es decir en general, a elegir un modelo con un número excesivo de dimensiones que pueden ser interpretadas. Es decir el criterio de Máxima Verosimilitud, tiende a no ser apropiado para obtener un modelo (factorial) correcto (en el sentido del número de factores). El problema se presenta también en el contexto de las series temporales. También, por otra parte, ya hemos comentado antes otros inconvenientes de los test

basados en máxima verosimilitud en el Análisis Factorial, relacionados con su carácter asintótico y el tamaño muestral adecuado (ver sección 1.7.1 y Anexo ??).

El criterio básico de esta línea de test basados en la Información, es el criterio de Akaike (AIC) que fue inicialmente desarrollado en Series Temporales y en Análisis Factorial máximo verosímil (Akaike (1974) y (1987)) y su idea clave es la de penalizar un exceso de parámetros ajustados, cosa que no hace el test de la  $\chi^2$  asintótico. Siguiendo a Basilewsky (1994) (sección 6.2.2) que a su vez sigue a Bozdogan (1987), los criterios basados en Información para seleccionar modelos se basan en el planteamiento general siguiente:

Sea un vector aleatorio  $X$  con densidad  $f(X; \theta)$ , con  $\theta$  un vector paramétrico. Si existe un verdadero valor  $\theta^*$  del parámetro  $\theta$ , la medida de bondad de ajuste viene dada por la Información de Kullback-Leibler,  $-I$  dada por

$$E[\ln f(X; \theta) - \ln f(X; \theta^*)] = -I. \quad (1.2)$$

Bajo el principio de “maximización de Entropía” se estima  $f(X; \theta^*)$  por medio de  $f(X; \theta)$  tal que se maximice la entropía media

$$E_x[B] = E_x\left[E[\ln f(X; \theta)] - E[\ln f(X; \theta^*)]\right] = E_x\left[E[\ln f(X; \theta)]\right], \text{ entropía media,}$$

puesto que grandes valores de  $E[B]$  implican que el modelo  $f(X; \theta)$  es un buen ajuste de  $f(X; \theta^*)$ . Obsérvese que en la medida que esto último es cierto,  $\frac{f(X/\theta)}{f(X/\theta^*)}$  tiende a 1, con lo que  $\ln \frac{f(X/\theta)}{f(X/\theta^*)}$  tiende a 0. Obsérvese también que maximizar  $E[B]$  equivale a minimizar  $(-I)$ .

En consecuencia,  $E[B] = E[-I]$  puede ser considerado como una función riesgo que mide el error medio de estimación (average estimation error) del modelo ajustado. En tales circunstancias se define el criterio de Akaike.

### Criterio de Akaike

El criterio de Akaike es un estimador muestral de  $E[\ln f(X; \theta)]$ , esperanza de la log-verosimilitud (o negantropía) que es dada por la expresión general:

$$AIC(r) = -2 \ln L(r) + 2m \quad (1.3)$$

en donde:  $L(r)$  es la verosimilitud del modelo ajustado y  $m$  es el número de parámetros “libres” después de que un modelo ha sido ajustado.

La ecuación 1.3 obtenida por Akaike (1974) para series temporales, puede ser adaptada al caso de la **metodología máximo verosímil del Análisis Factorial exploratorio**. Supongamos que hemos ajustado un modelo con  $r$  factores comunes. Entonces sabemos que la verosimilitud  $L(r)$  es:

$$L(r) = \frac{N}{2} \sum_{i=r+1}^p \ln \hat{\theta}_i$$

y el valor de  $m$  será el número de parámetros “libres” que será:

$$m = p(r+1) - \frac{1}{2}r(r-1)$$

en consecuencia el criterio de Akaike (AIC) toma la forma:

$$AIC(r) = -2 \frac{N}{2} \sum_{i=r+1}^p \ln \hat{\theta}_i + [2p(r+1) - r(r-1)] \quad (1.4)$$

En lo anterior, como ya se vio,  $\hat{\theta}_i$ ;  $i = r+1, \dots, p$ , son las  $(p-r)$  últimas (más pequeñas) raíces características. Para más detalles, puede verse el Anexo ??, sección ??.

En base al  $AIC(r)$ , se establece la siguiente metodología para la selección de un modelo factorial de un mínimo de factores,  $r$  adecuado. En efecto, se utiliza la ecuación 1.4 para  $r = 1$  y se va variando el número de factores posibles, seleccionándose el modelo factorial con el  $r$  tal que para dicho modelo se minimice  $AIC(r)$ .

Esta forma de selección de un modelo factorial adecuado no es un test realmente, aunque depende de los valores de  $L(r)$ , es decir, las  $(p - r)$  raíces características estimadas  $\hat{\theta}_i$ , como, en definitiva, le ocurre al criterio de Lawley-Maxwell antes visto:

$$\chi^2 = - \left[ N - 1 - \frac{2p + 4r + 5}{6} \right] \sum_{i=r+1}^p \ln(\eta_i + 1).$$

Sin embargo, el  $m$  del criterio de Akaike no es el número de grados de libertad de una  $\chi^2$  como ocurre en el test de hipótesis del criterio de la  $\chi^2$ , sino que responde al número de parámetros libres del sistema (Basilewsky (1994), p.388).

### Corrección de Schwarz al criterio de AIC

Es curioso observar que el término de penalización ( $2m$ ) no depende del tamaño muestral  $N$ . Esto conduce al hecho de que el mismo número de parámetros comunes es seleccionado por el AIC(r) para muestras pequeñas o para muestras grandes, es decir, una misma estructura común. Por tanto, el AIC(r) no es un estimador consistente del número adecuado (o correcto) de factores comunes. Para evitar en lo posible estos inconvenientes del AIC(r), Schwarz (1987) introdujo un nuevo criterio que se expresa de la forma:

$$SIC(r) = -\frac{N}{2} \sum_{i=r+1}^p \ln \hat{\theta}_i + \frac{m}{2} \ln N \quad (1.5)$$

El valor de  $r$  es calculado minimizando SIC(r) en  $r$ . Puede comprobarse que el criterio SIC(r) selecciona modelos con un número de factores  $r$  más pequeño que los seleccionados por el AIC(r), cuando  $N > 8$ . Hay otros criterios alternativos al AIC(r) y al SIC(r), como los propuestos por Sclove (1987) y Bozdogan (1987).

### Ejemplos metodológicos

Un interesante ejemplo metodológico (sobre 32 modelo de automóvil), resuelto mediante los criterios de Akaike y de Schwarz, con discusión, puede verse en Basilewsky (1994) pp. 388-392, que compara además con los resultados obtenidos vía test de Lawley-Bartlett.

## Tema 2

# Aplicación mediante R

Para realizar un análisis factorial en R, existen varias funciones que podemos usar, las más usuales son las que comentamos a continuación.

### 2.1. Paquete Factanal

Este paquete está incluido dentro de la librería **stats**, por lo que tendremos que cargarlo. Este método realiza un análisis factorial sobre una matriz de datos o de covarianzas por el método de máxima verosimilitud. Su sintaxis es:

```
factanal(x, factors, data = NULL, covmat = NULL, n.obs = NA, subset, na.action,
start = NULL, scores = c("none", "regression", "Bartlett"),
rotation = "varimax", control = NULL, ...)
```

donde:

- **x**: fórmula o matriz de datos u objeto que contenga una matriz numérica.
- **factors**: número de factores que se extraen.
- **data**: fichero de datos (solo si **x** es formula).
- **covmat**: matriz de covarianzas (o correlaciones).
- **n.obs**: número de observaciones.
- **subset**: vector opcional que selecciona filas de la matriz de datos.
- **na.action**: actitud ante los datos faltantes.
- **stat**: valor nulo o matriz con los valores de las unicidades iniciales.
- **scores**: criterio para el cálculo de las puntuaciones factoriales (no calcularlas, por regresión o por el método de Bartlett).
- **rotation**: para elegir entre no rotar o una función (por defecto se realiza la varimax).

Los objetos de esta clase son:

- **loadings**: matriz con las cargas factoriales. Los factores están ordenados en forma decreciente de las sumas de los cuadrados de las cargas.
- **uniquenesses**: vector con las unicidades.
- **correlation**: matriz de correlaciones.

- **criteria:** resultados del proceso de optimización: menos el logaritmo de la verosimilitud
- **factors:** factores introducidos.
- **dof:** grados de libertad.
- **method:** método utilizado (siempre mle, estimación máximo verosímil).
- **scores:** matriz de puntuaciones factoriales.
- **n.obs:** número de observaciones.
- **STATISTIC, PVAL:** significación de la adecuación del modelo.

También son interesantes los paquetes:

```
varimax(x, normalize = TRUE, eps = 1e-5)
promax(x, m = 4)
```

donde:

- **varimax:** realiza la rotación varimax.
- **x:** es una matriz de factor loadings.
- **normalize:** es el procedimiento de normalización de Kaiser.
- **eps:** es la tolerancia de la rotación.
- **m:** la potencia de la rotación promax.

y el paquete:

```
loadings(x)

## S3 method for class 'loadings':
print(x, digits = 3, cutoff = 0.1, sort = FALSE, ...)

## S3 method for class 'factanal':
print(x, digits = 3, ...)
```

con el que podemos imprimir las cargas factoriales, decidiendo el número de dígitos (digits = 3), el punto de corte a partir de que valor de las cargas no se imprimen (cutoff = 0.1) y si se ordenan los factores de menor a mayor.

## 2.2. Paquete prcomp

El paquete *prcomp* esta incluido en la librería *stats* y lo utilizaremos para realizar análisis de componentes principales:

```
## S3 method for class 'formula':
prcomp(formula, data = NULL, subset, na.action, ...)

## Default S3 method:
prcomp(x, retx = TRUE, center = TRUE, scale. = FALSE,
tol = NULL, ...)

## S3 method for class 'prcomp':
predict(object, newdata, ...)
```

donde:

- **formula**: formula que incluye solo las variables numéricas con ninguna dependiente.
- **data**: objeto que contiene las variables de la formula anterior.
- **subset**: vector opcional para seleccionar filas de la matriz de datos.
- **na.action**: indica que hacer con los datos faltantes.
- **x**: fórmula o matriz de datos (o un objeto que contenga una matriz numérica).
- **retx**: indica si se devuelve la rotación.
- **center**: valor lógico que indica si las variables serán centradas por las medias.
- **scale**: valor lógico que indica si las variables deben ser escaladas para tener varianza uno.
- **tol**: valor a partir del cuál las componentes no se calculan. Una componente será omitida si su desviación típica es menor o igual que tol veces la correspondiente a la primera componente.

los objetos de esta clase son:

- **sdev**: desviaciones típicas de cada componente
- **rotation**: matriz de las cargas factoriales.
- **x**: si retx es verdadero, devuelve el valor de los datos rotados.

También podemos usar las órdenes *screeplot* o *biplot*.

## 2.3. Ejemplo de aplicación

En primer lugar, vamos a realizar el análisis mediante la extracción de factores por componentes principales. Para ello, abriremos el fichero “factorial.txt”. y aplicaremos el paquete *prcomp*:

```
datos<-read.table("factorial.txt",header=TRUE,row.names=1)
attach(datos)
```

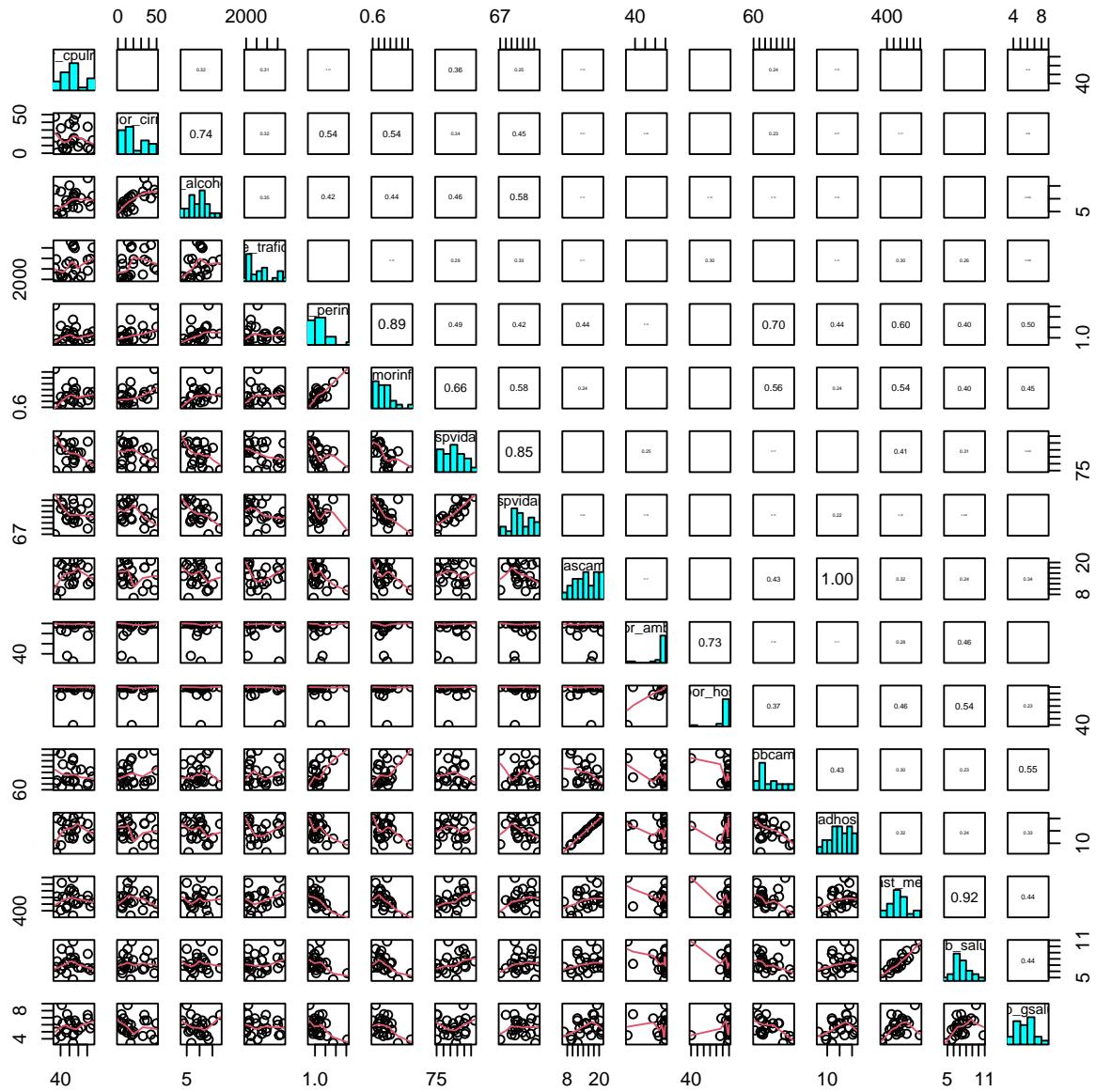
Antes de empezar a trabajar debemos contrastar una serie de hipótesis previas. Estas son la normalidad, la linealidad y la correlación.

```
panel.hist <- function(x, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(usr[1:2], 0, 1.5) )
  h <- hist(x, plot = FALSE)
  breaks <- h$breaks; nB <- length(breaks)
  y <- h$counts; y <- y/max(y)
  rect(breaks[-nB], 0, breaks[-1], y, col = "cyan", ...)
}
#linealidad
panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, ...)
{
  8
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
  r <- abs(cor(x, y))
  txt <- format(c(r, 0.123456789), digits = digits)[1]
```

```

txt <- paste(prefix, txt, sep = "")
if(missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)
text(0.5, 0.5, txt, cex = cex.cor * r)
}
pairs(datos, diag.panel = panel.hist, lower.panel = panel.smooth, upper.panel = panel.cor)

```



Con estas órdenes podemos analizar gráficamente las opciones de normalidad y linealidad

Evidentemente al tener muchas variables (y este no es un ejemplo demasiado extenso) analizar estas salidas gráficas es complicado, por ello contrastaremos la normalidad de la siguiente forma:

```

library(mvnormtest)
mshapiro.test(t(datos))

##

```

```
## Shapiro-Wilk normality test
##
## data: Z
## W = 0.38221, p-value = 7.654e-09
```

El contraste de normalidad tiene como hipótesis nula que los datos sigan una distribución normal, por lo que rechazamos esta hipótesis. El rechazar esta hipótesis no es un problema para realizar el Análisis Factorial, solo que no podremos utilizar (o no serán fiables) la metodología máximo verosímil.

Finalmente estudiaremos la correlación de los datos. Esta hipótesis es complicada de contrastar, ya que bivariantemente es fácil pero multivariantemente es muy complicada. Para contrastarla se suelen utilizar tres métodos.

- Estudio de la matriz de correlaciones: en este caso se estudia si las correlaciones son significativas y si el determinante de la matriz es pequeño
- Medida de adecuación muestral de Kayser, Meyer y Olkin (KMO): esta medida está comprendida entre 0 y 1 y cuanto mayor sea mejor se adecuan los datos a un análisis factorial.
- ] Contraste de esfericidad de Bartlet: contrasta si la matriz de correlaciones es igual a la identidad (en cuyo caso no existiría correlación) frente a que la matriz de correlaciones sea significativa.

```
Redatos<-cor(datos,method = c("pearson"))
symnum(Redatos)

##          mr_cp mr_cr c h mr_p mr espvdh espvdm d pr_m pr_h pb a g
## mor_cpulmon    1
## mor_cirro      1
## c_alcohol     . ,   1
## he_trafico   . .   . 1
## mor_perinatal . .   . 1
## morinf        . .   +   1
## espvidah     . .   . ,   1
## espvidam     . .   . .   +   1
## diascama       .
## por_ambu         .
## por_hos          ,
## pobcama        , .
## adhos          .
## gast_medi     . .   .
## pib_salud      . .   .
## pib_gsalud     . .   .
##                  pb_s pb_g
## mor_cpulmon
## mor_cirro
## c_alcohol
## he_trafico
## mor_perinatal
## morinf
## espvidah
## espvidam
## diascama
## por_ambu
## por_hos
## pobcama
```

```

## adhos
## gast_medi
## pib_salud      1
## pib_gsalud     .    1
## attr(,"legend")
## [1] 0 ' ' 0.3 '.' 0.6 ',' 0.8 '+' 0.9 '*' 0.95 'B' 1

det(Redatos)

## [1] 1.353152e-10

```

Vemos que pares de variables son significativos y que el determinante es pequeño, por lo que asumiremos que la matriz va a ser significativa.

Si realizamos la prueba KMO, en este caso el valor es pequeño, como podemos observar 0,48, pero por el contraste de esfericidad podemos asumir que la matriz de correlaciones es significativa.

```

library(psych)
KMO(datos)

## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = datos)
## Overall MSA = 0.48
## MSA for each item =
##   mor_cpulmon   mor_cirro   c_alcohol   he_trafico   mor_perinatal
##   0.19          0.40        0.43         0.45          0.64
##   morinf       espvidah   espvidam   diascama      por_ambu
##   0.65          0.45        0.42         0.45          0.27
##   por_hos       pobcama    adhos       gast_medi    pib_salud
##   0.53          0.60        0.45         0.58          0.46
##   pib_gsalud
##   0.65

```

```

cortest.bartlett(datos,n=NULL)

## R was not square, finding R from data

## $chisq
## [1] 359.7874
##
## $p.value
## [1] 8.871193e-26
##
## $df
## [1] 120

```

Una vez analizados estos pasos previos, pasamos al análisis factorial propiamente dicho. En primer lugar, es importante, imprimir el resumen del análisis y la gráfica de los autovalores (plot) para determinar el número de factores

```

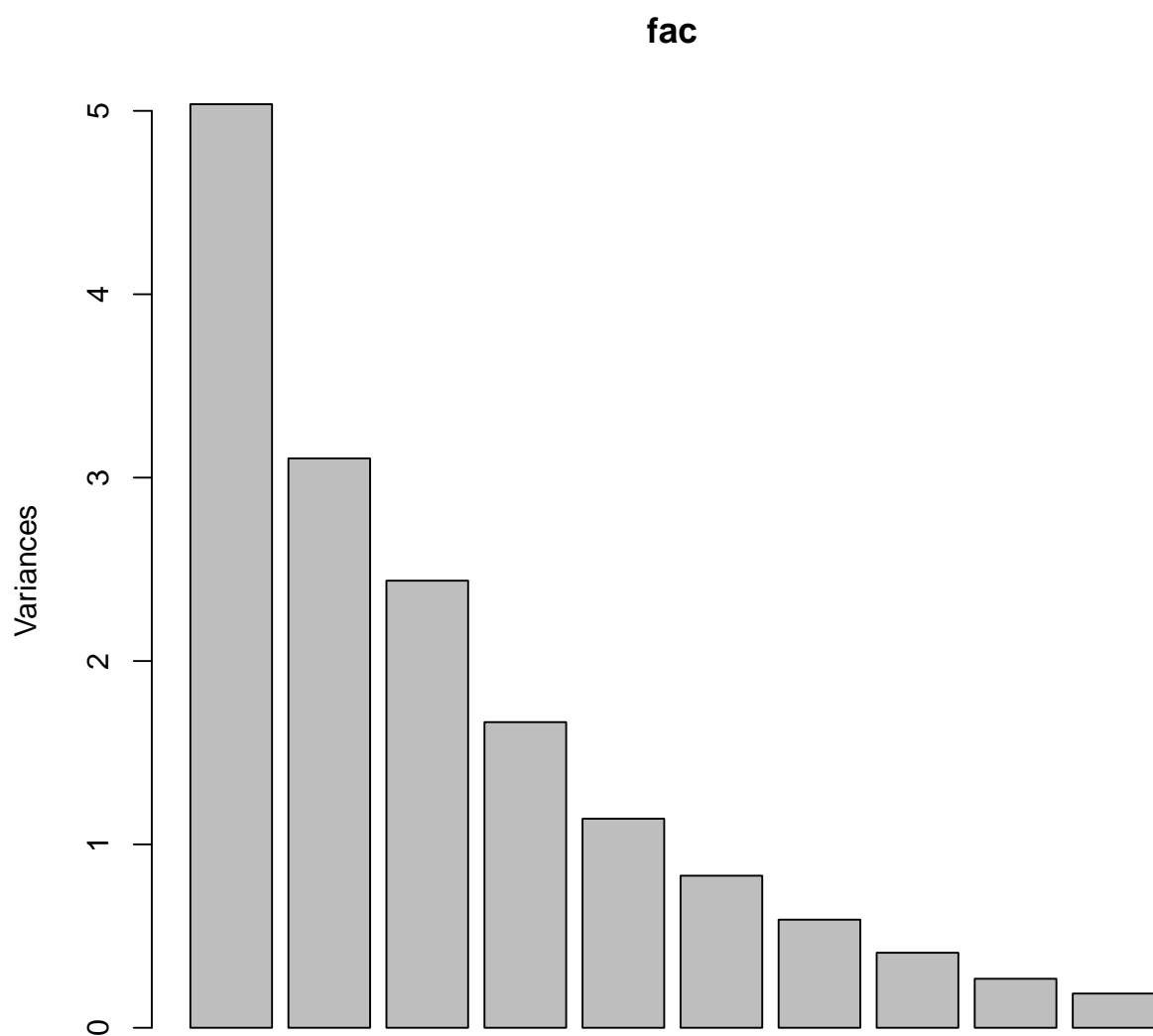
fac<-prcomp(datos, retx=,center=TRUE,scale.=TRUE,tol=NULL)
summary(fac)

## Importance of components:
##                 PC1    PC2    PC3    PC4    PC5    PC6
## Standard deviation 2.2442 1.7621 1.5614 1.2910 1.06768 0.91070

```

```
## Proportion of Variance 0.3148 0.1941 0.1524 0.1042 0.07125 0.05184
## Cumulative Proportion  0.3148 0.5088 0.6612 0.7654 0.83664 0.88848
##                               PC7      PC8      PC9      PC10     PC11     PC12
## Standard deviation      0.76781 0.63965 0.51701 0.43202 0.36799 0.31884
## Proportion of Variance 0.03685 0.02557 0.01671 0.01166 0.00846 0.00635
## Cumulative Proportion  0.92532 0.95089 0.96760 0.97926 0.98773 0.99408
##                               PC13     PC14     PC15     PC16
## Standard deviation      0.23911 0.15520 0.11249 0.02793
## Proportion of Variance 0.00357 0.00151 0.00079 0.00005
## Cumulative Proportion  0.99765 0.99916 0.99995 1.00000

plot(fac)
```



donde el gráfico de será (solo se muestran los 8 primeros valores).

Al igual que en el caso anterior, vamos a seleccionar una estructura factorial con 4 factores. Y las cargas factoriales serán:

```
fac
```

```
## Standard deviations (1, ..., p=16):
## [1] 2.24420782 1.76208722 1.56144644 1.29102974 1.06768372 0.91069652
## [7] 0.76781157 0.63964659 0.51701318 0.43201613 0.36798650 0.31883943
## [13] 0.23911088 0.15519530 0.11249484 0.02793139
##
## Rotation (n x k) = (16 x 16):
##          PC1       PC2       PC3       PC4
## mor_cpulmon 0.002168196 0.24391538 0.21620153 -0.15629688
## mor_cirro   -0.274470311 0.15946595 0.02306329 -0.35162771
## c_alcohol   -0.243529081 0.28720356 0.05367817 -0.35495787
## he_trafico  -0.083559197 0.30921647 -0.11425616 -0.34620682
## mor_perinatal -0.406846958 -0.02365215 -0.07845725 0.13711280
## morinf      -0.396082688 0.08733273 0.04789417 0.12148711
## espvidah    0.278959235 -0.28955421 -0.25057776 -0.17050323
## espvidam    0.240003427 -0.36167443 -0.24289345 -0.07703057
## diascama     0.223234628 0.24830160 0.32273752 0.31526741
## por_ambu     -0.022042121 -0.29450448 0.31056835 -0.41217839
## por_hos      0.006599818 -0.30750787 0.46182849 -0.21065155
## pobcama     -0.287841288 -0.06081815 -0.31438870 0.22459135
## adhos        0.221617224 0.25176743 0.32435086 0.31792081
## gast_medi   0.302546520 0.28236802 -0.25471943 -0.16802946
## pib_salud    0.240571634 0.32789209 -0.30715603 -0.13333042
## pib_gsalud   0.262312659 0.12984779 0.17142060 -0.16225424
##          PC5       PC6       PC7       PC8
## mor_cpulmon 0.64670436 -0.01548086 0.458070975 0.31312566
## mor_cirro   -0.44665337 0.08451930 0.135267471 -0.12330996
## c_alcohol   -0.12788598 -0.09828201 0.466513784 -0.13131549
## he_trafico  0.27314116 0.39955560 -0.572574003 0.05489151
## mor_perinatal -0.18437938 -0.15741798 0.054032604 0.17930899
## morinf      -0.08938349 0.01923148 -0.120216372 0.29340132
## espvidah   -0.18482621 0.19661434 0.246357878 0.11539037
## espvidam   0.06623038 0.05265113 0.200477691 0.16301672
## diascama    -0.21652480 0.28408459 0.083709229 0.11936571
## por_ambu    -0.12708072 0.36253492 0.003337878 0.30010666
## por_hos     -0.11744582 -0.11600423 -0.197398309 0.08861104
## pobcama    -0.05037179 0.13319804 0.045125808 0.60210887
## adhos       -0.21288421 0.27827114 0.103986998 0.09620288
## gast_medi  -0.13588538 0.09063991 -0.030579053 0.18715637
## pib_salud   -0.23077746 -0.11546293 -0.004176529 0.07600669
## pib_gsalud  -0.11989525 -0.64375210 -0.219899890 0.42849270
##          PC9       PC10      PC11      PC12
## mor_cpulmon -0.18315512 0.01121982 -0.147097893 0.27544210
## mor_cirro   -0.23473904 0.51882605 -0.326557593 0.17596029
## c_alcohol   0.27719338 -0.08251489 0.503701363 -0.30121624
## he_trafico -0.09039804 0.19308880 0.298824130 -0.01452360
## mor_perinatal -0.25623341 -0.16560177 0.182072608 0.19195731
## morinf      -0.48520347 -0.36955297 -0.008758639 -0.29630419
## espvidah   -0.06573952 -0.14307077 0.147706827 0.13427616
## espvidam   -0.47748574 0.35132897 0.273751048 -0.27318343
## diascama    -0.06200301 0.12826199 0.138815809 -0.06835936
```

```

## por_ambu      0.16713467 -0.24701443 -0.379579680 -0.23123832
## por_hos      -0.01785333 -0.06411059  0.438204906  0.52608321
## pobcama      0.47291246  0.28176445  0.052118793  0.16564159
## adhos       -0.05981854  0.08341361  0.098698710  0.02009988
## gast_medi    0.08306029 -0.31939293  0.073442595 -0.03317821
## pib_salud   -0.15689247 -0.23163853 -0.152169053  0.38517378
## pib_gsalud   0.04599678  0.23476008 -0.048437507 -0.26965890
##                  PC13        PC14        PC15        PC16
## mor_cpulmon -0.0125490166 -0.03438996  0.09824789 -0.028552850
## mor_cirro     0.0964441318  0.01118108  0.25219226 -0.002263703
## c_alcohol     -0.0448466115  0.12059012 -0.15226902  0.024657921
## he_trafico   -0.2227244252 -0.09969925 -0.06958286  0.005898783
## mor_perinatal -0.0905213011 -0.71522997 -0.17462231 -0.042364411
## morinf       -0.0209752650  0.45019830  0.21423540  0.025748558
## espvidah    -0.6266094839  0.02005375  0.37183765 -0.060920541
## espvidam     0.3107593764  0.02207572 -0.26481106  0.067509566
## diascama     -0.0004524565 -0.01290453 -0.06887355 -0.697831268
## por_ambu      0.0303680119 -0.15125158 -0.31377594  0.004968104
## por_hos       0.2310465485  0.20720898  0.06391198  0.006957487
## pobcama      0.0750781697  0.19875576 -0.02145276  0.022909413
## adhos        -0.1085776821 -0.06764019 -0.04606519  0.706562952
## gast_medi    0.5598272655 -0.24845997  0.42639435  0.024869279
## pib_salud   -0.0486000902  0.28331619 -0.56335858 -0.017804547
## pib_gsalud   -0.2314723558 -0.08384447  0.06992847  0.008781721

```

Estas cargas factoriales corresponden a los autovectores de la matriz de correlaciones. Para obtener las cargas factoriales debemos multiplicar cada columna por la raíz cuadrada del autovalor correspondiente. Con las siguientes ordenes obtenemos el procedimiento varimax con normalización de Kaiser:

```

cargas<-matrix(0,16,4)
for (i in 1:4) cargas[,i]<-fac$rotation[,i]*fac$sdev[i]
cargas2<-varimax(cargas,normalize=T)$loadings
print(cargas2,cutoff=0.3)

##
## Loadings:
##      [,1]   [,2]   [,3]   [,4]
## [1,] 0.451   0.338
## [2,] 0.731
## [3,] 0.857
## [4,] 0.679
## [5,] -0.860
## [6,] -0.822  0.406
## [7,] 0.617 -0.534   -0.434
## [8,] 0.466 -0.649   -0.454
## [9,]          0.891
## [10,]         0.849
## [11,]         0.933
## [12,] -0.676   -0.338 -0.426
## [13,]          0.897
## [14,] 0.764   -0.562
## [15,] 0.637   -0.670
## [16,] 0.613   0.359
## 
```

```

##          [,1]  [,2]  [,3]  [,4]
## SS loadings   4.093 2.941 2.579 2.634
## Proportion Var 0.256 0.184 0.161 0.165
## Cumulative Var 0.256 0.440 0.601 0.765

```

En relación a las communalidades, R no las obtiene de forma directa, pero se pueden hacer de modo inmediato sabiendo que el modelo verifica que  $h_i = \sum_{j=1}^m l_{ij}^2$ , donde  $h_i$  es la communalidad de la variable  $i$  y  $l_{ij}$  la carga de la variable  $i$  en el factor  $j$ ;  $y m$  el número de factores. También podemos calcular las unicidades a partir de la relación  $\sigma_i^2 = h_i + \psi_i^2$ , donde  $\sigma_i^2$  es la varianza de la  $i$ -ésima variable y  $\psi_i^2$  es la unicidad.

```

comunalidad<-matrix(0,16,2)
for (i in 1:16)
  {for (j in 1:4)
    {comunalidad[i,1]=comunalidad[i,1]+cargas[i,j]^2
     comunalidad[i,2]=1-comunalidad[i,1]}}
comunalidad

##          [,1]  [,2]
## [1,] 0.3394337 0.66056632
## [2,] 0.6657523 0.33424767
## [3,] 0.7718379 0.22816210
## [4,] 0.5636492 0.43635079
## [5,] 0.8817385 0.11826148
## [6,] 0.8440028 0.15599722
## [7,] 0.8537956 0.14620436
## [8,] 0.8499947 0.15000532
## [9,] 0.8620354 0.13796462
## [10,] 0.7900783 0.20992166
## [11,] 0.8878025 0.11219749
## [12,] 0.7538266 0.24617340
## [13,] 0.8691386 0.13086140
## [14,] 0.9138220 0.08617801
## [15,] 0.8849610 0.11503900
## [16,] 0.5144237 0.48557627

```

Podemos también calcular la contribución de cada factor en la explicación de cada variable, tanto para el total de la varianza de la variable (cargas3) como para el total explicado por el modelo factorial (cargas4):

```

cargas3<-matrix(0,16,4)
cargas4<-matrix(0,16,4)
for (i in 1:16){
  cargas3[i,]<-cargas2[i,]^2
  cargas4[i,]<-cargas3[i,]/comunalidad[i,1]}
cargas3

##          [,1]      [,2]      [,3]      [,4]
## [1,] 0.017248489 0.202987894 5.282009e-03 0.113915290
## [2,] 0.066487146 0.533836607 8.538384e-03 0.056890195
## [3,] 0.031265522 0.734258618 4.933675e-05 0.006264427
## [4,] 0.020143626 0.460466884 6.735249e-02 0.015686209
## [5,] 0.739802051 0.061144580 1.567207e-03 0.079224681
## [6,] 0.675585924 0.165080940 9.948114e-06 0.003325968
## [7,] 0.380238305 0.284904671 6.916454e-04 0.187961020
## [8,] 0.217212319 0.421236276 5.223742e-03 0.206322338
## [9,] 0.050122100 0.010331747 6.845696e-03 0.794735837

```

```
## [10,] 0.022405662 0.005826094 7.212907e-01 0.040555924
## [11,] 0.002584568 0.007580848 8.695894e-01 0.008047701
## [12,] 0.456647415 0.001293899 1.140132e-01 0.181872113
## [13,] 0.048178017 0.009554947 7.275845e-03 0.804129793
## [14,] 0.583933947 0.009478536 3.156569e-01 0.004752650
## [15,] 0.405683008 0.029299635 4.487775e-01 0.001200829
## [16,] 0.375328196 0.003380011 6.716384e-03 0.128999137
```

cargas4

```
## [,1]      [,2]      [,3]      [,4]
## [1,] 0.050815490 0.598019303 1.556124e-02 0.335603967
## [2,] 0.099867688 0.801854656 1.282517e-02 0.085452491
## [3,] 0.040507886 0.951311945 6.392113e-05 0.008116248
## [4,] 0.035737877 0.816938758 1.194936e-01 0.027829736
## [5,] 0.839026577 0.069345479 1.777406e-03 0.089850539
## [6,] 0.800454619 0.195592886 1.178683e-05 0.003940708
## [7,] 0.445350487 0.333691878 8.100831e-04 0.220147551
## [8,] 0.255545506 0.495575194 6.145618e-03 0.242733683
## [9,] 0.058143901 0.011985293 7.941317e-03 0.921929489
## [10,] 0.028358785 0.007374071 9.129356e-01 0.051331522
## [11,] 0.002911197 0.008538890 9.794852e-01 0.009064742
## [12,] 0.605772491 0.001716442 1.512459e-01 0.241265184
## [13,] 0.055431915 0.010993582 8.371329e-03 0.925203173
## [14,] 0.639001855 0.010372410 3.454249e-01 0.005200849
## [15,] 0.458419079 0.033108392 5.071156e-01 0.001356928
## [16,] 0.729609029 0.006570479 1.305613e-02 0.250764360
```

También podemos calcular la matriz de covarianzas reproducida por el modelo teniendo en cuenta que en el modelo ortogonal  $\Sigma = LL' + \Psi$ , donde  $L$  es la matriz de cargas factoriales y  $\Psi$  la matriz de unicidades. Por ello:

```
ajuste<-cor(datos)-cargas2%*%t(cargas2)
ajuste

##          mor_cpulmon    mor_cirro     c_alcohol   he_trafico
## mor_cpulmon 0.66056632 -0.285385639 -0.016327730 0.047268598
## mor_cirro   -0.28538564  0.334247668  0.048306656 -0.150075943
## c_alcohol   -0.01632773  0.048306656  0.228162096 -0.221238893
## he_trafico   0.04726860 -0.150075943 -0.221238893  0.436350793
## mor_perinatal -0.08195424  0.072282084  0.033533607 -0.113454819
## morinf      -0.04625032  0.012905722 -0.058198119  0.023197570
## espvidah   -0.05283505  0.105591041  0.076716127 -0.063147397
## espvidam   0.13361131  0.025594037  0.018457621  0.005961687
## diascama   -0.12680774  0.139537512  0.030093670  0.013078282
## por_ambu   -0.06689097  0.052610669 -0.028764848  0.058475065
## por_hos    -0.12056362  0.017934096 -0.019142356  0.005192506
## pobcama   0.03127195  0.006872933  0.006157417  0.025922133
## adhos     -0.11834557  0.138397868  0.032367012  0.002992101
## gast_medi -0.09241361  0.028060648  0.008076320 -0.015152724
## pib_salud  -0.13992665  0.104285225  0.009624363 -0.117849889
## pib_gsalud -0.09263401 -0.007144761 -0.008586208 -0.157931781
##          mor_perinatal    morinf     espvidah     espvidam
## mor_cpulmon  -0.081954235 -0.046250321 -0.052835048  0.133611310
## mor_cirro     0.072282084  0.012905722  0.105591041  0.025594037
```

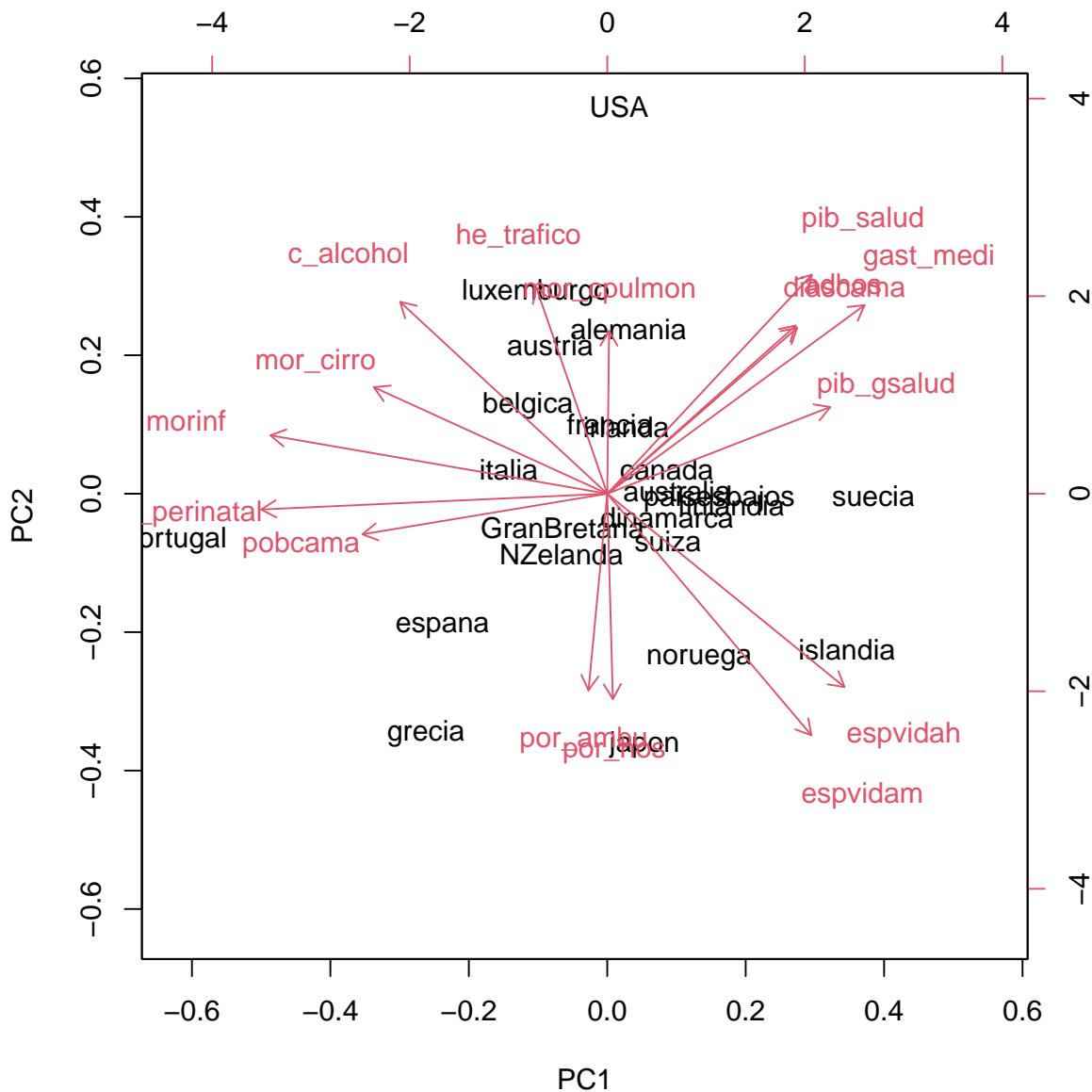
```

## c_alcohol      0.033533607 -0.058198119  0.076716127  0.018457621
## he_trafico   -0.113454819  0.023197570 -0.063147397  0.005961687
## mor_perinatal 0.118261481  0.064506999  0.046755301  0.019409038
## morinf        0.064506999  0.155997220  0.034510673  0.044202677
## espvidah     0.046755301  0.034510673  0.146204359  0.019828102
## espvidam     0.019409038  0.044202677  0.019828102  0.150005323
## diascama      0.022614985  0.035739319  0.108964749  0.037460040
## por_ambu       -0.013040248  0.054769544  0.090817772 -0.017019795
## por_hos        0.059410845  0.027258024 -0.008562621 -0.027642328
## pobcama       -0.001540014 -0.007474052  0.052087254  0.004467101
## adhos         0.024934739  0.030877174  0.114553326  0.029216712
## gast_medi     0.035269224  0.049885778  0.037719148 -0.016945844
## pib_salud     0.087309408  0.057599096  0.043161098 -0.028217303
## pib_gsalud    0.119302616  0.054552140 -0.094566345 -0.023747956
##               diascama   por_ambu   por_hos   pobcama
## mor_cpulmon  -0.1268077434 -0.06689097 -0.1205636189  0.031271952
## mor_cirro     0.1395375123  0.05261067  0.0179340964  0.006872933
## c_alcohol      0.0300936699 -0.02876485 -0.0191423557  0.006157417
## he_trafico    0.0130782819  0.05847507  0.0051925057  0.025922133
## mor_perinatal 0.0226149848 -0.01304025  0.0594108453 -0.001540014
## morinf        0.0357393186  0.05476954  0.0272580243 -0.007474052
## espvidah     0.1089647486  0.09081777 -0.0085626209  0.052087254
## espvidam     0.0374600402 -0.01701980 -0.0276423278  0.004467101
## diascama      0.1379646197  0.11771105 -0.0005443319  0.074127171
## por_ambu       0.1177110540  0.20992166 -0.0407139566  0.122423448
## por_hos        -0.0005443319 -0.04071396  0.1121974902  0.018796432
## pobcama       0.0741271712  0.12242345  0.0187964320  0.246173403
## adhos         0.1323224819  0.11470583 -0.0024708328  0.066526000
## gast_medi     0.0547935114  0.08548316  0.0323252487  0.057869950
## pib_salud     0.0251957357  0.01159676  0.0606538220 -0.006192487
## pib_gsalud    -0.1062193787 -0.12432385  0.0953688981  0.047342500
##               adhos   gast_medi   pib_salud   pib_gsalud
## mor_cpulmon  -0.118345565 -0.09241361 -0.139926650 -0.092634008
## mor_cirro     0.138397868  0.02806065  0.104285225 -0.007144761
## c_alcohol      0.032367012  0.00807632  0.009624363 -0.008586208
## he_trafico    0.002992101 -0.01515272 -0.117849889 -0.157931781
## mor_perinatal 0.024934739  0.03526922  0.087309408  0.119302616
## morinf        0.030877174  0.04988578  0.057599096  0.054552140
## espvidah     0.114553326  0.03771915  0.043161098 -0.094566345
## espvidam     0.029216712 -0.01694584 -0.028217303 -0.023747956
## diascama      0.132322482  0.05479351  0.025195736 -0.106219379
## por_ambu       0.114705826  0.08548316  0.011596760 -0.124323854
## por_hos        -0.002470833  0.03232525  0.060653822  0.095368898
## pobcama       0.066526000  0.05786995 -0.006192487  0.047342500
## adhos         0.130861397  0.05069499  0.029906942 -0.112831427
## gast_medi     0.050694988  0.08617801  0.034184690 -0.012120209
## pib_salud     0.029906942  0.03418469  0.115038998  0.084987961
## pib_gsalud    -0.112831427 -0.01212021  0.084987961  0.485576272

```

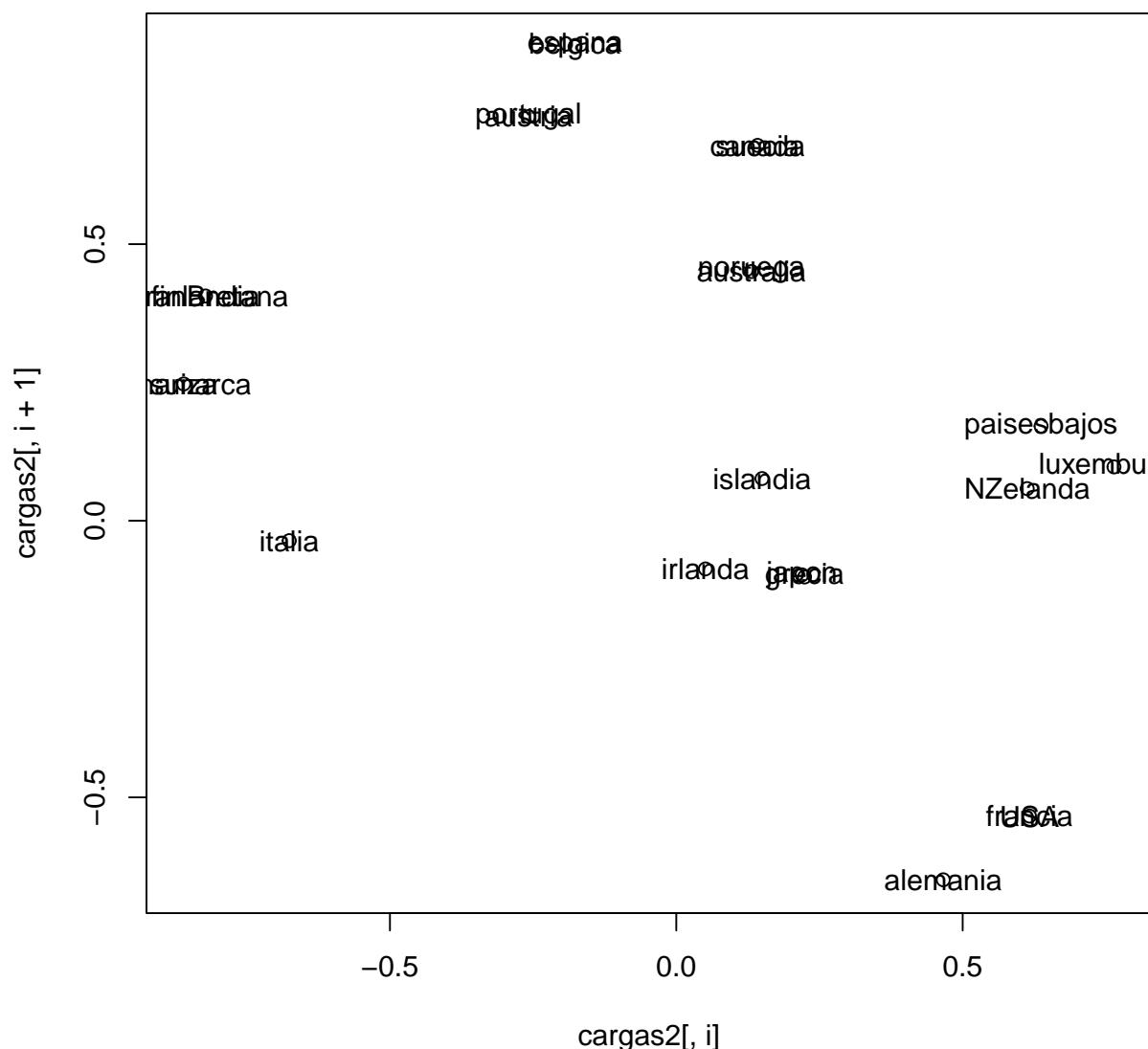
A continuación realizaremos las opciones gráficas. La primer de ellas, será representar el gráfico biplot

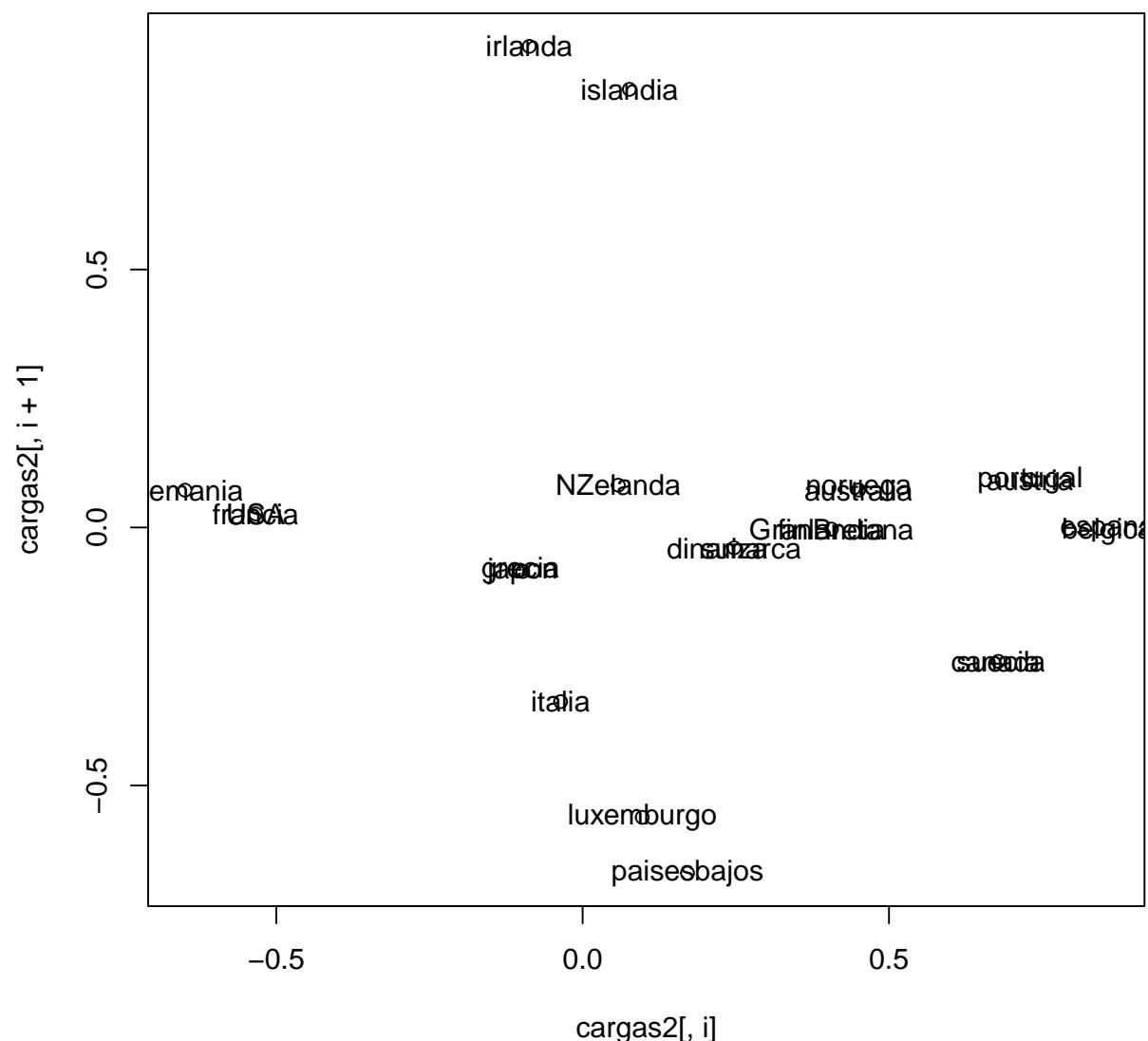
`biplot(fac)`

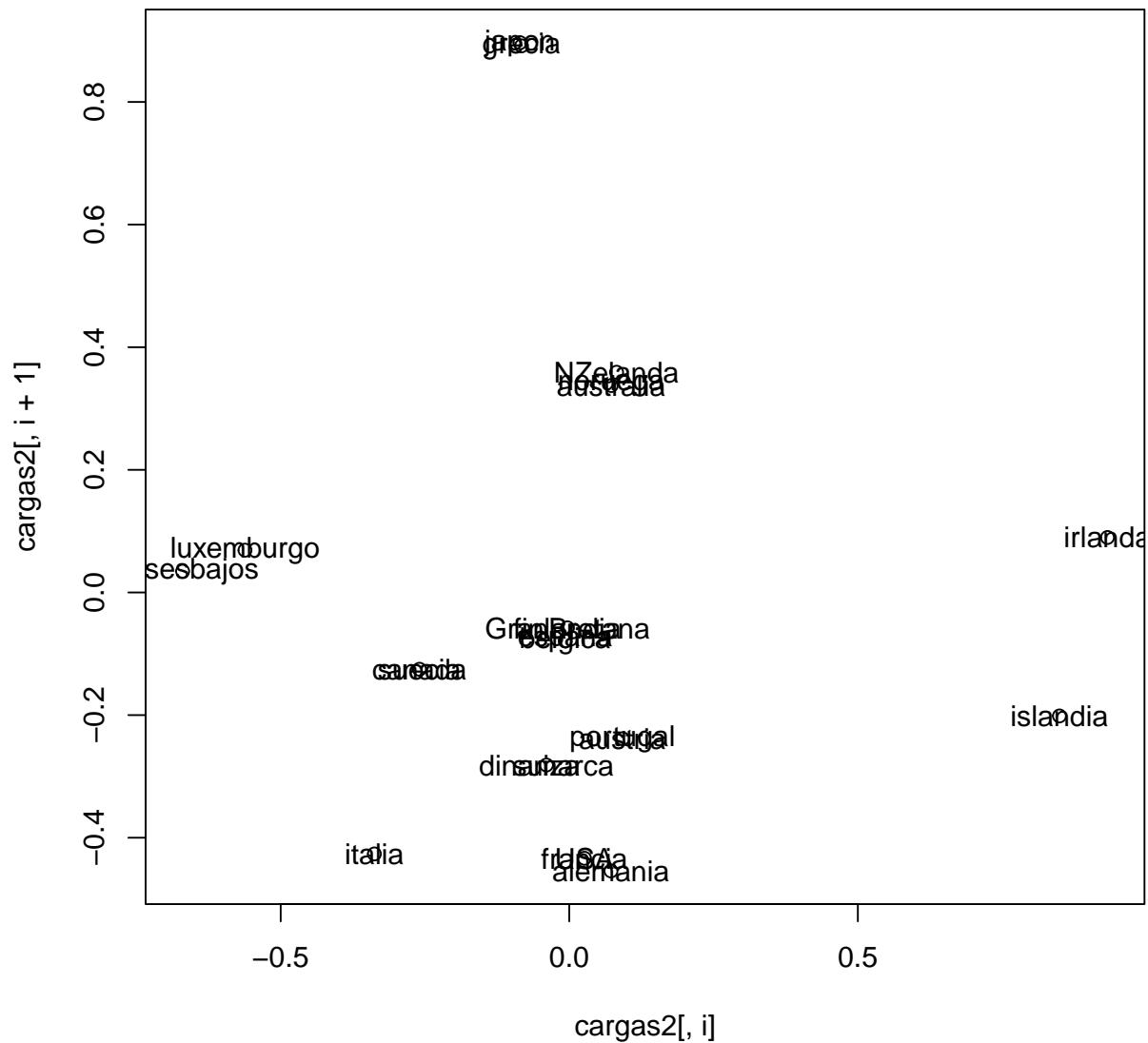


y las gráficas bidimensionales de las cargas factoriales con la orden

```
for (i in 1:3){  
  plot(cargas2[,i],cargas2[,i+1])  
  text(cargas2[,i],cargas2[,i+1],labels=row.names(datos))}
```

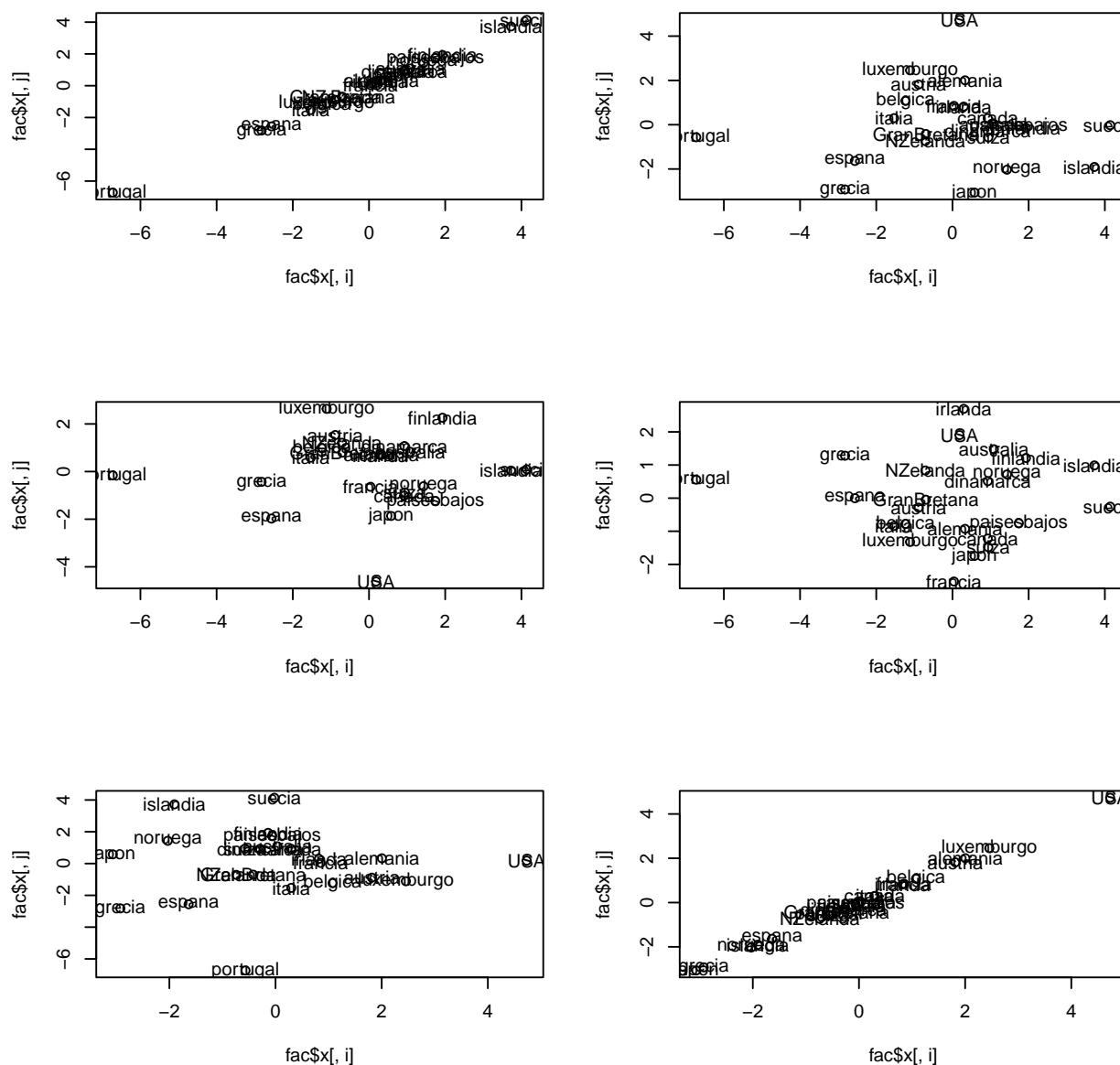


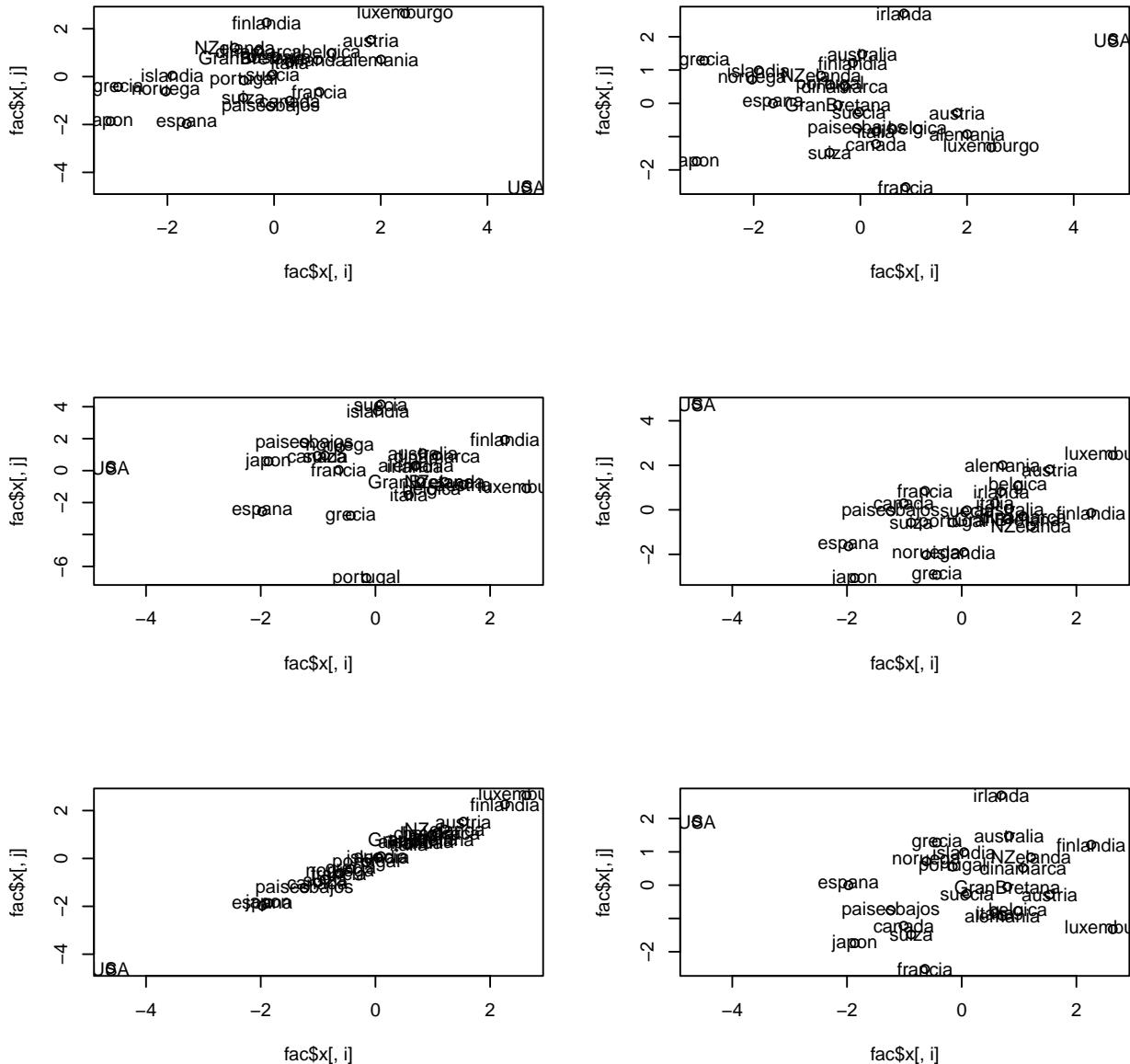




Por último representamos las puntuaciones factoriales de los individuos:

```
par(mfrow=c(3,2))
for (i in 1:3){
  for(j in 1:4){
    {plot(fac$x[,i],fac$x[,j])
    text(fac$x[,i],fac$x[,j],labels=row.names(fac$x))}}
```





Finalmente vamos a comprobar si la hipótesis del número de factores considerado es correcta. Para ello recurriremos al paquete factanal:

```
facmle<-vector("list",4)
for (i in 1:4) {facmle<-factanal(datos,i)}
facmle

## 
## Call:
## factanal(x = datos, factors = i)
## 
## Uniquenesses:
##   mor_cpulmon    mor_cirro    c_alcohol   he_trafico mor_perinatal
```

```

##          0.665      0.599      0.527      0.545      0.005
## morinf    espvidah    espvidam    diascama    por_ambu
## 0.129      0.113      0.149      0.005      0.684
## por_hos   pobcama     adhos     gast_medi   pib_salud
## 0.572      0.444      0.005      0.075      0.033
## pib_gsalud
##          0.670
##
## Loadings:
##           Factor1 Factor2 Factor3 Factor4
## mor_cpulmon -0.228      0.523
## mor_cirro    0.516      0.184      0.295 -0.117
## c_alcohol    0.373      0.224      0.519 -0.121
## he_trafico
##          0.324      0.547 -0.209
## mor_perinatal 0.970
##          -0.194
## morinf       0.868 -0.107      0.327
## espvidah    -0.455      0.152 -0.789 -0.184
## espvidam    -0.414
##          -0.773 -0.282
## diascama    -0.262
##          0.958
## por_ambu    -0.155 -0.517 -0.119 -0.101
## por_hos     -0.110 -0.639
## pobcama     0.689
##          -0.123 -0.257
## adhos        -0.256
##          0.959
## gast_medi   -0.509      0.801
## pib_salud   -0.313      0.927
## pib_gsalud  -0.461      0.271      0.195
##
##           Factor1 Factor2 Factor3 Factor4
## SS loadings  3.752      2.505      2.323      2.206
## Proportion Var 0.234      0.157      0.145      0.138
## Cumulative Var 0.234      0.391      0.536      0.674
##
## Test of the hypothesis that 4 factors are sufficient.
## The chi square statistic is 89.7 on 62 degrees of freedom.
## The p-value is 0.0123

```

Por lo que concluiríamos que 4 factores NO son suficientes, aunque como hemos recordado la hipótesis de normalidad no se cumple en este caso, por lo que sería conveniente subir a un 5 factor.

# Referencias Bibliográficas

- Afifi, A. A., and S. P. Azen. 1979. *Statistical analysis, A computer orieoriente approach*. Academic Press. Second Edition.
- Akaike, H. 1974. "A new look at the statistical model identification." *IEEE Trans Autom Control* AC-19:716–723.
- . 1987. "Factor analysis and AIC." *Psychometrika* 52:317–332.
- Anderson, T. W. 1984. *An introduction to multivariate statistical analysis*. John Wiley & Sons.
- Anderson, T. W., and H. Rubin. 1956. "Statistical inference in factor analysis." *Proceedings of the third Berkeley Symp.*, Volume Vol. V. University of California, Berkeley, 111–150.
- Bartlett, M. S. 1947. "Multivariate Analysis." *Journal of the Royal Statistical Society Suppl.* 9B:176–197.
- Basilewsky, A. 1994. *Statistical factor analysis and related methods: theory and applications*. New York.: John Wiley and Sons.
- Bozdogan, H. 1987. "Model selection and Akaike's information criterion (AIC): the general theory and its analytical extension." *Psychometrika* 52:345–370.
- Gutiérrez, R., and A. Gónzalez. 1991. *Introducción al análisis multivariante; Volumen 1*. Universidad de Granada.
- Harman, H. 1967. *Modern factor analysis*. U. University Chicago Press. (Traducción espesa Ansis factorial moderno,, Editorial Salt Madrid (1980)).
- Howe, W. G. 1955. "Some contributions in factor analysis." Technical Report, Atomic Energy Comm. Report Oak Ridge Laboratory, Tennessee, USA.
- Johnson, R. A., and D. W. Wichern. 1998. *Applied multivariate statistical analysis*. Prentice-Hall.
- Jreskog, K. G. 1969. "A general approach to confirmatory maximum likelihood factor analysis." *Psychometrika* 34:183–202.
- Kaiser, H. F. 1958. "The varimax criterion for analysis rotation in factor analysis." *Psychometrika* 23:187–200.
- Lawley, D. N., and E. Maxwell. 1971. *Factor analysis as a statistical method*. Ed. Butterworths, London.
- Reiersol, O. 1950. "On the identifiability of parameters in Thurstone's multiple factor analysis." *Psychometrika* 15:121–159.
- Schwarz, G. 1987. "Estimating the dimension of a model." *Annals of Statistics* 6:431–464.
- Sclove, S. 1987. "Applications of the model selection criteria to some problem in multivariate analysis." *Psychometrika* 52:333–343.
- Takeuchi, K., H. Yanai, and B. N. Mukherjee. 1982. *The foundations of multivariate analysis*. Wiley Eastern Ltd, New Delhi.
- Thurstone, L. L. 1945. *Multiple-Fsctor Analysis*. University Chicago Press.