

TEMA 5

ANÁLISIS FACTORIAL

Curso 2017/18

1 Introducción

Objetivo:

- El objetivo del Análisis Factorial (AF) es describir, si es posible, la estructura de covarianzas entre diversas variables en términos de un número reducido de variables latentes, no observables, denominadas **factores**.
- Supuesto que las variables están correladas, el objetivo es reducir la redundancia usando un menor número de factores
- El AF se sustenta en el siguiente argumento: Si las variables observables pueden agruparse a través de sus correlaciones (altas correlaciones dentro de los grupos y bajas entre variables de distintos grupos) parece lógico pensar que cada grupo de variables representa una variable simple subyacente (o factor) que es responsable de las correlaciones observadas, aunque dicho factor no sea observable.

Nota



Originalmente desarrollado por los psicólogos (principios de 1940).

Psicología: No es posible medir exactamente los conceptos en los que se está interesado (inteligencia, capacidad de liderazgo, ...)

La aplicación práctica del AF no siempre produce resultados satisfactorios y algunos lo consideran una técnica polémica (subjetividad, interpretación de los factores,...)

Nota

Similitud entre AF y ACP: tratan de explicar un conjunto de datos en un menor número de dimensiones.

Diferencias entre AF y ACP: se verán al final del tema

2 Modelo factorial ortogonal (Problema poblacional)

Sea \underline{X} un vector aleatorio (observable) p -dimensional, $\underline{X} \sim (\underline{\mu}, \Sigma)$.

El modelo factorial postula que el vector aleatorio \underline{X} depende linealmente de "un número reducido" de variables aleatorias no observables F_1, \dots, F_m , denominadas **factores comunes** y p fuentes de variación adicionales $\varepsilon_1, \dots, \varepsilon_p$ denominados **errores aleatorios** o **factores específicos**:

$$X_i - \mu_i = \sum_{j=1}^m l_{ij} \cdot F_j + \varepsilon_i \quad i = 1, \dots, p \quad \Longleftrightarrow \quad \begin{matrix} \underline{X} \\ (p \times 1) \end{matrix} - \begin{matrix} \underline{\mu} \\ (p \times 1) \end{matrix} = \begin{matrix} \mathbf{L} \\ (p \times m) \end{matrix} \begin{matrix} \underline{F} \\ (m \times 1) \end{matrix} + \begin{matrix} \underline{\varepsilon} \\ (p \times 1) \end{matrix}$$

La matriz $\mathbf{L} = (l_{ij})$ se denomina **matriz de cargas factoriales**. l_{ij} es la **carga factorial** de la i -ésima variable sobre el j -ésimo factor.

Los factores comunes pueden estar relacionados con todas las variables originales, mientras que cada factor específico está relacionado con una única variable original. **Tanto los factores comunes como los factores específicos son no observables.**

En el modelo

$$\begin{matrix} \underline{X} \\ (p \times 1) \end{matrix} - \begin{matrix} \underline{\mu} \\ (p \times 1) \end{matrix} = \begin{matrix} \underline{\mathbf{L}} \\ (p \times m) \end{matrix} \begin{matrix} \underline{F} \\ (m \times 1) \end{matrix} + \begin{matrix} \underline{\varepsilon} \\ (p \times 1) \end{matrix}$$

se asumen las siguientes hipótesis:

$$E[\underline{F}] = \underline{0} \quad Cov[\underline{F}] = E[\underline{F} \underline{F}^t] = \mathbf{I}_m$$

(factores comunes incorrelados y con varianza unidad)

$$E[\underline{\varepsilon}] = \underline{0} \quad Cov(\underline{\varepsilon}) = E[\underline{\varepsilon} \underline{\varepsilon}^t] = \mathbf{\Psi} = diag(\psi_1, \dots, \psi_p)$$

(factores específicos incorrelados y varianza ψ_i)

$$Cov(\underline{F}, \underline{\varepsilon}) = E[\underline{F} \underline{\varepsilon}^t] = \mathbf{0}$$

(factores comunes y específicos incorrelados)

Proposición 1 *Bajo las hipótesis del modelo:*

1.

$$\Sigma = \mathbf{L}\mathbf{L}^t + \Psi \Leftrightarrow \begin{cases} \sigma_{ii} = l_{i1}^2 + \dots + l_{im}^2 + \psi_i \\ \sigma_{ik} = \sum_{j=1}^m l_{ij} \cdot l_{kj} \end{cases} \quad (1)$$

2.

$$Cov(\underline{X}, \underline{F}) = \mathbf{L} \Leftrightarrow l_{ij} = Cov(X_i, F_j) \quad (2)$$

Nota 1 El modelo supone que los $p(p+1)/2$ elementos de Σ pueden ser "reproducidos" a través de las pm cargas factoriales y las p varianzas específicas.

Comunalidad: parte de variabilidad de X_i explicada por los m factores comunes

$$\sigma_{ii} = \underbrace{l_{i1}^2 + \dots + l_{im}^2}_{\text{Comunalidad}} + \underbrace{\psi_i}_{\text{varianza específica}}$$

$$\sigma_{ii} = h_i^2 + \psi_i$$

Comunalidades: **elementos diagonales de \mathbf{LL}^t** .

Nota 2 Interpretación de las cargas factoriales.

Dado que

$$\text{Cov}(\underline{X}, \underline{F}) = \mathbf{L} \implies \text{Cov}(X_i, F_j) = l_{ij}$$

el coeficiente de correlación lineal entre X_i y F_j es:

$$\rho_{X_i, F_j} = l_{ij} / \sqrt{\sigma_{ii}}.$$

En consecuencia, el j -th factor común podrá interpretarse en términos de las variables originales con las que posee una mayor correlación.

Nota 3 Esfericidad en las variables originales. *Si la matriz Σ es diagonal, o sus elementos no diagonales son muy pequeños, entonces las variables originales no están muy relacionadas. En esta situación el AF no proporcionará resultados muy útiles. En estas circunstancias, los factores específicos juegan un papel dominante, dejando un papel casi irrelevante a los factores comunes.*

Si la matriz Σ se desvía significativamente de una matriz diagonal, el problema será determinar las cargas factoriales y las varianzas específicas.

Nota 4 No existencia de solución.

Cuando $m = p$ puede encontrarse una representación exacta.

Al ser Σ simétrica, semidefinida positiva, por su descomposición espectral

$$\Sigma = \sum_{j=1}^p \lambda_j \underline{e}_j \underline{e}_j^t = \begin{bmatrix} \sqrt{\lambda_1} \underline{e}_1 & \sqrt{\lambda_2} \underline{e}_2 & \dots & \sqrt{\lambda_p} \underline{e}_p \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \underline{e}_1^t \\ \sqrt{\lambda_2} \underline{e}_2^t \\ \vdots \\ \sqrt{\lambda_p} \underline{e}_p^t \end{bmatrix} = \mathbf{L} \mathbf{L}^t$$

donde $\lambda_1, \dots, \lambda_p$ son sus autovalores y $\underline{e}_1, \dots, \underline{e}_p$ los autovectores unitarios asociados.

En este caso Ψ es la matriz nula.

Esta solución no es válida ya que el objetivo es encontrar **un número pequeño de factores**.

Desafortunadamente, la gran mayoría de las matrices Σ no admiten factorizaciones de esta forma cuando se considera un número m de factores muy pequeño*

*Véase Ejemplo 9.2 (pp. 382-383) del libro:
Johnson, R.A. y Wichern, D.W. (1988). *Applied multivariate statistical analysis* (2^o edition). Prentice-Hall.

Nota 5 Existencia de infinitas soluciones (rotaciones)

Si existe solución para la representación de la matriz de covarianzas: $\Sigma = \mathbf{L}\mathbf{L}^t + \Psi$ entonces hay una ambigüedad inherente.

Para cualquier matriz $m \times m$ ortogonal \mathbf{T} , $(\mathbf{T}\mathbf{T}^t = \mathbf{T}^t\mathbf{T} = \mathbf{I}_m)$, se tiene:

$$\Sigma = \mathbf{L}\mathbf{L}^t + \Psi = \mathbf{L}\mathbf{T}\mathbf{T}^t\mathbf{L}^t + \Psi = \mathbf{L}^*\mathbf{L}^{*t} + \Psi$$

y los modelos

$$\underline{X} - \underline{\mu} = \mathbf{L} \underline{F} + \underline{\varepsilon}$$

$$\underline{X} - \underline{\mu} = \mathbf{L}\mathbf{T}\mathbf{T}^t \underline{F} + \underline{\varepsilon} = \mathbf{L}^* \underline{F}^* + \underline{\varepsilon}$$

verifican las mismas propiedades.

Esta ambigüedad permite utilizar las transformaciones ortogonales, **rotaciones**, para obtener una mejor interpretación de los factores comunes.

- La rotación modifica la matriz de cargas factoriales (\mathbf{L} y \mathbf{L}^* son distintas).
- La rotación no altera las comunales h_i^2 (los elementos diagonales de $\mathbf{L}\mathbf{L}^t$ coinciden con los de $\mathbf{L}^*\mathbf{L}^{*t}$) ni afecta a los factores específicos ($\underline{\varepsilon}$) ni a las varianzas específicas ψ_i .

Nota 6 AF sobre las variables estandarizadas

Dada $\mathbf{\Lambda} = \text{diag}\{\sigma_{11}, \dots, \sigma_{pp}\}$,

$$\underline{\mathbf{Z}} = \mathbf{\Lambda}^{-1/2}(\underline{\mathbf{X}} - \underline{\boldsymbol{\mu}}) \implies \text{Cov}(\underline{\mathbf{Z}}) = \mathbf{\Lambda}^{-1/2}\mathbf{\Sigma}\mathbf{\Lambda}^{-1/2} = \mathbf{R}$$

Para una factorización

$$\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^t + \mathbf{\Psi}$$

se tiene una factorización

$$\mathbf{R} = \mathbf{\Lambda}^{-1/2}\mathbf{L}\mathbf{L}^t\mathbf{\Lambda}^{-1/2} + \mathbf{\Lambda}^{-1/2}\mathbf{\Psi}\mathbf{\Lambda}^{-1/2} = \mathbf{G}_z\mathbf{G}_z^t + \mathbf{\Psi}_z$$

con matriz de cargas $\mathbf{G}_z = \mathbf{\Lambda}^{-1/2}\mathbf{L}$ y varianzas específicas $\mathbf{\Psi}_z = \mathbf{\Lambda}^{-1/2}\mathbf{\Psi}\mathbf{\Lambda}^{-1/2}$.

3 Métodos de estimación (problema muestral)

Problema teórico (poblacional), consiste en:

$$\text{Dado } (\Sigma, m) \longrightarrow \text{determinar } \longrightarrow (\mathbf{L}, \Psi) : \Sigma = \mathbf{L}\mathbf{L}^t + \Psi$$

Problema muestral

Sea $\underline{x}_1, \dots, \underline{x}_n$ una muestra aleatoria de $\underline{X} \sim (\underline{\mu}, \Sigma)$ y sea

$$\hat{\Sigma} = \frac{1}{n-1} \mathbf{S} = \frac{1}{n-1} \sum_{k=1}^n (\underline{x}_k - \bar{\underline{x}}) (\underline{x}_k - \bar{\underline{x}})^t.$$

El problema muestral consiste en

$$\text{Dado } (\hat{\Sigma}, m) \longrightarrow \text{determinar } \longrightarrow (\hat{\mathbf{L}}, \hat{\Psi}) : \hat{\Sigma} \approx \hat{\mathbf{L}}\hat{\mathbf{L}}^t + \hat{\Psi}$$

Vamos a construir el modelo factorial a partir de $\hat{\Sigma}$ (análogamente sobre $\hat{\mathbf{R}}$) usando los métodos de estimación más usuales: **método de las componentes principales**, el **método del factor principal** y el **método de máxima verosimilitud**.

3.1 Método de componentes principales

La descomposición espectral proporciona una factorización de la matriz Σ con $m = p$ factores y varianzas específicas nulas.

$$\Sigma = \sum_{j=1}^p \lambda_j \underline{e}_j \underline{e}_j^t = \mathbf{L} \mathbf{L}^t$$

$$\text{siendo } \mathbf{L} = \begin{bmatrix} \sqrt{\lambda_1} \underline{e}_1 & \sqrt{\lambda_2} \underline{e}_2 & \dots & \sqrt{\lambda_p} \underline{e}_p \end{bmatrix}.$$

Las columnas de \mathbf{L} son proporcionales a los autovectores de Σ .

Aunque la factorización sea exacta, no es útil dado que se obtienen p factores comunes (tantos como variables iniciales), y el objetivo es obtener un número de factores comunes más pequeño.

Si los últimos $p - m$ autovalores son pequeños y despreciamos

$$\sum_{j=m+1}^p \lambda_j \underline{e}_j \underline{e}_j^t$$

se obtiene la aproximación

$$\Sigma \approx \begin{bmatrix} \sqrt{\lambda_1} \underline{e}_1 & \sqrt{\lambda_2} \underline{e}_2 & \dots & \sqrt{\lambda_m} \underline{e}_m \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \underline{e}_1^t \\ \sqrt{\lambda_2} \underline{e}_2^t \\ \vdots \\ \sqrt{\lambda_m} \underline{e}_m^t \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{bmatrix} \quad (3)$$

$$\Sigma \approx \mathbf{L}_m \mathbf{L}_m^t + \Psi \quad (4)$$

donde

$$\psi_i = \sigma_{ii} - \sum_{j=1}^m l_{ij}^2 \quad \text{para } i = 1, \dots, p$$

La descomposición (4) aplicada a la matriz $\hat{\Sigma}$ proporciona la denominada **solución de componentes principales**.

Solución de componentes principales:

Sean $(\hat{\lambda}_j, \hat{e}_j)$, $j = 1, \dots, p$ los autovalores $(\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_p \geq 0)$ y autovectores unitarios de $\hat{\Sigma}$.

Fijado el número de factores comunes $m < p$, las matrices de cargas factoriales y varianzas específicas estimadas son

$$\hat{\mathbf{L}}_m = \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{e}_1 & \dots & \sqrt{\hat{\lambda}_m} \hat{e}_m \end{bmatrix}$$

$$\hat{\Psi} = \text{diag}\{\hat{\psi}_1, \dots, \hat{\psi}_p\}$$

con $\hat{\psi}_i = \hat{\sigma}_{ii} - \sum_{j=1}^m \hat{l}_{ij}^2$.

Las estimaciones de las cargas factoriales y de las comunales son

$$\hat{l}_{ij} = \sqrt{\hat{\lambda}_j} \hat{e}_{ji} \quad i = 1, \dots, p; \quad j = 1, \dots, m$$

$$\hat{h}_i^2 = \sum_{j=1}^m \hat{l}_{ij}^2, \quad i = 1, \dots, p$$

Nota 7 *En la solución de componentes principales, las cargas factoriales estimadas no cambian cuando se incrementa el número m de factores elegidos.*

- *Los elementos diagonales de $\hat{\Sigma}$ quedan perfectamente representados por la factorización. No ocurre igual con los elementos no diagonales.*
- *La elección de m puede basarse en las magnitudes de los autovalores de forma semejante al ACP.*
- *Es recomendable usar las variables estandarizadas si las unidades de las variables originales son muy dispares (equivale a sustituir $\hat{\Sigma}$ por $\hat{\mathbf{R}}$).*

Nota 8 Matriz residual: $\hat{\Phi}$

$$\hat{\Phi} = \hat{\Sigma} - (\hat{\mathbf{L}}_m \hat{\mathbf{L}}_m^t + \hat{\Psi})$$

Los elementos diagonales $\hat{\Phi}$ son nulos. Si los no diagonales son "suficientemente pequeños", entonces se puede considerar que la factorización obtenida con los m factores es "suficientemente adecuada".

Proposición 2. Sea $\hat{\Phi} = (\hat{\varphi}_{ij})$. Se verifica que

$$\|\hat{\Phi}\|_2^2 = \sum_{i=1}^p \sum_{k=1}^p \hat{\varphi}_{ik}^2 = \sum_{i=1}^p \sum_{\substack{k=1 \\ k \neq i}}^p \hat{\varphi}_{ik}^2 \leq \sum_{j=m+1}^p \hat{\lambda}_j^2 \quad (5)$$

Nota 9 *En consecuencia, un valor pequeño de la suma de cuadrados de los autovalores implicará un valor pequeño en la suma de cuadrados de los elementos de la matriz de residuos, por lo que se podría concluir que la factorización es adecuada.*

Demostración de (5)

$$\widehat{\Phi} = \widehat{\Sigma} - (\widehat{\mathbf{L}}_m \widehat{\mathbf{L}}_m^t + \widehat{\Psi}) = \sum_{j=m+1}^p \lambda_j \widehat{\underline{e}}_j \widehat{\underline{e}}_j^t - \widehat{\Psi} = \mathbf{G} - \widehat{\Psi}$$

Los elementos diagonales de $\widehat{\Phi}$ son nulos y por tanto $\|\widehat{\Phi}\|_2^2 \leq \|\mathbf{G}\|_2^2$.

Por otro lado

$$\lambda_j \widehat{\underline{e}}_j \widehat{\underline{e}}_j^t = \begin{pmatrix} \lambda_j e_{j1} \widehat{\underline{e}}_j^t \\ \vdots \\ \lambda_j e_{jp} \widehat{\underline{e}}_j^t \end{pmatrix} \quad y \quad \mathbf{G} = \sum_{j=m+1}^p \lambda_j \widehat{\underline{e}}_j \widehat{\underline{e}}_j^t = \begin{pmatrix} \sum_{j=m+1}^p \lambda_j e_{j1} \widehat{\underline{e}}_j^t \\ \vdots \\ \sum_{j=m+1}^p \lambda_j e_{jp} \widehat{\underline{e}}_j^t \end{pmatrix}$$

Por tanto

$$\begin{aligned} \|\mathbf{G}\|_2^2 &= \sum_{i=1}^p \left(\sum_{j=m+1}^p \lambda_j e_{ji} \widehat{\underline{e}}_j^t \right) \left(\sum_{j=m+1}^p \lambda_j e_{ji} \widehat{\underline{e}}_j^t \right)^t \\ &= \sum_{i=1}^p \sum_{j=m+1}^p \lambda_j^2 e_{ji}^2 = \sum_{j=m+1}^p \lambda_j^2 \sum_{i=1}^p e_{ji}^2 = \sum_{j=m+1}^p \lambda_j^2. \end{aligned}$$

Nota 10 Varianza muestral de la i -ésima variable $\hat{\sigma}_{ii}$

$$\hat{\sigma}_{ii} = \sum_{j=1}^m \hat{l}_{ij}^2 + \hat{\psi}_i$$

Contribución de F_1 a $\hat{\sigma}_{ii}$ es el cuadrado de la correspondiente carga factorial

$$\hat{l}_{i1}^2 = \hat{\lambda}_1 \hat{e}_{1i}^2.$$

Contribución de F_1 a la variabilidad total $\sum_{i=1}^p \hat{\sigma}_{ii} = \text{tr}(\hat{\Sigma})$ es

$$\sum_{i=1}^p \hat{l}_{i1}^2 = \left(\sqrt{\hat{\lambda}_1} \hat{e}_1 \right)^t \left(\sqrt{\hat{\lambda}_1} \hat{e}_1 \right) = \hat{\lambda}_1$$

En general

$$\begin{array}{l} \text{Proporción de} \\ \text{variabilidad total} \\ \text{debida al } j\text{-ésimo factor} \end{array} = \begin{cases} \hat{\lambda}_j / \text{tr}(\hat{\Sigma}) & \text{si el estudio se hace sobre } \hat{\Sigma} \\ \hat{\lambda}_j / p & \text{si el estudio se hace sobre } \hat{\mathbf{R}} \end{cases}$$

3.2 La solución del factor principal

Es una modificación del método de las componentes principales. Haremos el desarrollo partiendo de la matriz de correlación muestral $\widehat{\mathbf{R}} = (r_{ij})$, como estimador de la matriz de correlación $\mathbf{R} = (\rho_{ij})$.

Si el modelo factorial está correctamente especificado:

$$\mathbf{R} = \mathbf{L}\mathbf{L}^t + \Psi$$

y por tanto

$$\mathbf{R} - \Psi = \mathbf{L}\mathbf{L}^t.$$

Paso 0. Considerar unos estimadores iniciales de las varianzas específicas, $\Psi_0^* = \text{diag} \{ \psi_{1,0}^*, \dots, \psi_{p,0}^* \}$

Considerar la descomposición

$$\widehat{\mathbf{R}} = (\widehat{\mathbf{R}} - \Psi_0^*) + \Psi_0^* = \widehat{\mathbf{R}}_{r,0} + \Psi_0^*$$

Matriz de correlación muestral reducida $\widehat{\mathbf{R}}_{r,0} = \widehat{\mathbf{R}} - \Psi_0^*$

$$\widehat{\mathbf{R}}_{r,0} = \begin{bmatrix} 1 - \psi_{1,0}^* & r_{12} & \dots & r_{1p} \\ r_{12} & 1 - \psi_{2,0}^* & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \dots & 1 - \psi_{p,0}^* \end{bmatrix}$$

Estimación de las comunales: $h_{i,0}^{*2} = 1 - \psi_{i,0}^*$

Estimación de las varianzas específicas: $\psi_{i,0}^*$

Paso 1. Realizar la aproximación

$$\widehat{\mathbf{R}}_{r,0} \approx \mathbf{L}_{r,0}^* \left(\mathbf{L}_{r,0}^* \right)^t$$

con

$$\mathbf{L}_{r,0}^* = \begin{bmatrix} \sqrt{\widehat{\lambda}_1^*} \widehat{\underline{e}}_1^* & \sqrt{\widehat{\lambda}_2^*} \widehat{\underline{e}}_2^* & \dots & \sqrt{\widehat{\lambda}_m^*} \widehat{\underline{e}}_m^* \end{bmatrix}$$

donde $(\widehat{\underline{e}}_i^*; \widehat{\lambda}_i^*)$, $i = 1, \dots, m$, son los m mayores autovalores-autovectores de $\widehat{\mathbf{R}}_{r,0}$.

Considerar la descomposición

$$\widehat{\mathbf{R}} \approx \mathbf{L}_{r,0}^* \left(\mathbf{L}_{r,0}^* \right)^t + \Psi_1^*$$

con

$$\psi_{i,1}^* = 1 - \sum_{j=1}^m l_{ij,0}^{*2} \quad (\text{igualdades en la diagonal})$$

Estimaciones de las comunales: $\widehat{h}_{i,1}^{*2} = \sum_{j=1}^m l_{ij,0}^{*2}$

Estimaciones de las varianzas específicas: $\psi_{i,1}^*$

Paso 3: Iterar...

Matriz de correlación muestral reducida y aproximación

$$\widehat{\mathbf{R}}_{r,1} = \widehat{\mathbf{R}} - \Psi_1^* \approx \mathbf{L}_{r,1}^* \left(\mathbf{L}_{r,1}^* \right)^t$$

Considerar la descomposición

$$\widehat{\mathbf{R}} \approx \mathbf{L}_{r,1}^* \left(\mathbf{L}_{r,1}^* \right)^t + \Psi_2^* \quad (\text{igualdades en la diagonal})$$

Estimaciones de las comunales: $\widehat{h}_{i,2}^{*2} = \sum_{j=1}^m l_{ij,1}^{*2}$

Estimaciones de las varianzas específicas: $\psi_{i,2}^* = 1 - \sum_{j=1}^m l_{ij,1}^{*2}$

Elección de los estimadores iniciales de las varianzas específicas

La más utilizada, cuando se trabaja con $\widehat{\mathbf{R}}$, es

$$\psi_{i,0}^* = 1/r^{ii}$$

donde r^{ii} es el elemento i —ésimo diagonal de $\widehat{\mathbf{R}}^{-1}$.

En consecuencia, las comunalidades iniciales estimadas son

$$h_{i,0}^{*2} = 1 - \psi_{i,0}^* = 1 - 1/r^{ii}$$

que coincide con el cuadrado del coeficiente de correlación múltiple entre X_i y el resto de las $(p - 1)$ variables (estas comunalidades pueden estimarse incluso si la matriz $\widehat{\mathbf{R}}$ no es de rango total).

Si se parte de $\widehat{\Sigma}$, se consideran como estimadores iniciales de las varianzas específicas los elementos diagonales $\hat{\sigma}^{ii}$ de la matriz $\widehat{\Sigma}^{-1} = (\sigma^{ij})$.

Nota 11 *Al igual que en el método de componentes principales, los autovalores estimados $\hat{\lambda}_1^* \geq \hat{\lambda}_2^* \geq \dots \geq \hat{\lambda}_p^*$ pueden utilizarse para determinar el número m de factores comunes que debe considerarse. Sin embargo, en este procedimiento puede ocurrir que algunos **autovalores sean negativos** debido a la introducción de las comunalidades estimadas en la diagonal de la matriz. El número de factores comunes "ideal" sería igual al rango de la matriz reducida poblacional (no muestral). Sin embargo, generalmente no es posible determinarlo.*

Nota 12 *El método de componentes principales sobre la matriz de correlación puede considerarse como un método de factor principal con las comunalidades iniciales iguales a la unidad (y las varianzas específicas iguales a cero). Aunque ambos métodos son computacional y geométricamente diferentes, en muchas ocasiones producen cargas factoriales semejantes si el número de variables es grande y el número de factores comunes pequeños.*

3.3 Método de máxima verosimilitud

Supone la normalidad del vector $\underline{X} \sim N_p(\underline{\mu}, \underline{\Sigma})$.

La función de verosimilitud asociada a una muestra aleatoria $\underline{X}_1, \dots, \underline{X}_n$

$$L(\underline{\mu}, \underline{\Sigma}) = cte. |\underline{\Sigma}|^{-n/2} \cdot \exp \left\{ -\frac{1}{2} \cdot tr [\underline{\Sigma}^{-1} \underline{S}] \right\} \cdot \exp \left\{ -\frac{n}{2} (\overline{\underline{X}} - \underline{\mu})^t \underline{\Sigma}^{-1} (\overline{\underline{X}} - \underline{\mu}) \right\}$$

Los estimadores de máxima verosimilitud (en general) vienen dados por

$$\arg \max_{\underline{\mu}, \underline{\Sigma}} L(\underline{\mu}, \underline{\Sigma}) = \left(\overline{\underline{X}}, \frac{1}{n} \underline{S} \right)$$

En el modelo factorial $\underline{\Sigma} = \underline{L}\underline{L}^t + \underline{\Psi}$ la función de verosimilitud es $L(\underline{\mu}, \underline{L}, \underline{\Psi})$ y los estimadores de máxima verosimilitud vienen dados por

$$\arg \max_{\underline{\mu}, \underline{L}, \underline{\Psi}} L(\underline{\mu}, \underline{L}, \underline{\Psi})$$

Debido a la falta de unicidad en la solución $\Sigma = \mathbf{L}\mathbf{L}^t + \Psi$ se debe imponer alguna condición para obtener solución única (condición de unicidad).

Una de las más utilizadas es:

$$\arg \max_{\underline{\mu}, \mathbf{L}, \Psi} L(\underline{\mu}, \mathbf{L}, \Psi)$$

$$\text{sujeto a } \mathbf{L}^t \Psi^{-1} \mathbf{L} = \Delta \text{ (diagonal)}$$

La solución de este problema requiere el uso de métodos numéricos.

4 Rotación de factores

Las cargas factoriales obtenidas a través de una transformación ortogonal de las cargas iniciales tienen la misma capacidad para reproducir la matriz de covarianza (o correlación). Dado que una transformación ortogonal realiza una rotación de los ejes coordenados, a las transformaciones ortogonales de las cargas factoriales, y por tanto de los factores, se denomina **rotación factorial** o **rotación de los factores**.

Sea $\hat{\mathbf{L}}$ la $(p \times m)$ -matriz estimada de cargas factoriales obtenida por alguno de los métodos anteriores, y \mathbf{T} una $(m \times m)$ -matriz ortogonal ($\mathbf{T}^t \mathbf{T} = \mathbf{T} \mathbf{T}^t = \mathbf{I}_m$).

La matriz $\hat{\mathbf{L}}^* = \hat{\mathbf{L}} \mathbf{T}$ es una $(p \times m)$ -matriz de cargas **rotada**. En particular, mantiene idéntica factorización de la matriz de covarianza (o correlación):

$$\hat{\Sigma} \approx \hat{\mathbf{L}} \hat{\mathbf{L}}^t + \hat{\Psi} = \hat{\mathbf{L}} \mathbf{T} \mathbf{T}^t \cdot \hat{\mathbf{L}}^t + \hat{\Psi} = \hat{\mathbf{L}}^* (\hat{\mathbf{L}}^*)^t + \hat{\Psi}$$

y por tanto, la matriz de residuos tampoco se ve modificada:

$$\hat{\Phi} = \hat{\Sigma} - (\hat{\mathbf{L}} \hat{\mathbf{L}}^t + \hat{\Psi}) = \hat{\Sigma} - \left[\hat{\mathbf{L}}^* (\hat{\mathbf{L}}^*)^t + \hat{\Psi} \right]$$

Además, las varianzas específicas estimadas $\hat{\psi}_i$ y las comunalidades \hat{h}_i^2 también se mantienen.

Dado que los factores comunes iniciales pueden no ser interpretables de forma adecuada, en la práctica se realizan rotaciones tratando de buscar una interpretación a los factores.

Para $m = 2$, las rotaciones adecuadas se pueden obtener a través de la representación gráfica de los pares de puntos $\{(\hat{l}_{i1}, \hat{l}_{i2}), i = 1, \dots, p\}$ y rotando los ejes un ángulo θ adecuado para buscar mayor representación (Véase ejemplo 9.8, pp 401). La matriz ortogonal viene dada por:

$$\mathbf{T} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

4.1 Criterio varimax

La interpretación de los factores se facilita si los que "afectan" (o correlacionan) a algunas variables no lo hacen a otras.

Dado que el modelo para cada variable viene dado por

$$X_i - \mu_i = \sum_{j=1}^m l_{ij} \cdot F_j + \varepsilon_i \quad i = 1 \dots p$$

se trata de buscar una representación tal que:

$$\text{si } \left(\hat{l}_{ij}^* \right)^2 \uparrow \text{ entonces } \left(\hat{l}_{kj}^* \right)^2 \downarrow \text{ para } k \neq i$$

(Se consideran los cuadrados para evitar el signo)

Una forma de alcanzar el objetivo anterior es buscar la representación con máxima variabilidad (**criterio varimax**) en los conjuntos de valores: $\left\{ \left(\hat{l}_{ij}^* \right)^2, i = 1, 2, \dots, p, j = 1, \dots, m \right\}$

Una medida de la variabilidad de los cuadrados de las cargas factoriales asociadas al factor j th viene dada por la varianza de $\left\{ \left(\hat{l}_{ij}^* \right)^2, i = 1, 2, \dots, p \right\}$:

$$V_j = \frac{1}{p} \sum_{i=1}^p \left[\left(\hat{l}_{ij}^* \right)^2 - \eta_j \right]^2$$

con

$$\eta_j = \frac{1}{p} \sum_{i=1}^p \left(\hat{l}_{ij}^* \right)^2 \quad (\text{aportación media de } F_j \text{ a } X_1, \dots, X_p)$$

La **rotación varimax** es la que maximiza la suma de las variabilidades de todos los factores:

$$V = \sum_{j=1}^m V_j$$

Nota 13 *Las rotaciones varimax obtenidas a partir de dos estimaciones distintas de las cargas factoriales resultan distintas. El aumento del número de factores comunes puede provocar grandes cambios en las rotación varimax.*

5 Puntuaciones factoriales

El desarrollo realizado hasta el momento ha tenido por objetivo estimar \mathbf{L} y Ψ .

Teniendo en cuenta que los **factores son variables no observable**, un problema de interés es **estimar** los valores de los factores (**puntuaciones factoriales**) en cada una de los n elementos muestrales, o bien, sobre un nuevo individuo.

Las puntuaciones factoriales pueden utilizarse como variables de entrada para análisis posteriores (análisis de conglomerados, MANOVA,...)

5.1 Método de mínimos cuadrados ponderados (Barlett)

Para cada observación muestral el modelo factorial teórico es

$$\underline{X}_k - \underline{\mu} = \mathbf{L}\underline{F}_k + \underline{\varepsilon}_k \quad , \quad k = 1, 2, \dots, n \quad (6)$$

$$\begin{aligned} E[\underline{\varepsilon}_k] &= 0 \\ Var[\underline{\varepsilon}_k] &= \mathbf{\Psi} = diag(\psi_i) \end{aligned}$$

- Si $\underline{\mu}$, \mathbf{L} y $\mathbf{\Psi}$ fuesen conocidos, (6) puede considerarse como un modelo de regresión múltiple con parámetros desconocidos \underline{F}_k .

- Barlett propone estimar \underline{F}_k mediante mínimos cuadrados ponderados (distintas varianzas):

$$\min_{\underline{F}_k} \sum_{i=1}^p \frac{1}{\psi_i} \varepsilon_{ki}^2 = \underline{\varepsilon}_k^t \Psi^{-1} \underline{\varepsilon}_k$$

$$\min_{\underline{F}_k} \left(\underline{X}_k - \underline{\mu} - \mathbf{L} \underline{F}_k \right)^t \Psi^{-1} \left(\underline{X}_k - \underline{\mu} - \mathbf{L} \underline{F}_k \right)$$

La solución del problema de minimización viene dado por:

$$\underline{f}_k = \left(\mathbf{L}^t \Psi^{-1} \mathbf{L} \right)^{-1} \mathbf{L}^t \Psi^{-1} \left(\underline{X}_k - \underline{\mu} \right)$$

Por tanto, un estimador de \underline{F}_k es

$$\hat{\underline{F}}_k = \left(\hat{\mathbf{L}}^t \hat{\Psi}^{-1} \hat{\mathbf{L}} \right)^{-1} \hat{\mathbf{L}}^t \hat{\Psi}^{-1} \left(\underline{X}_k - \overline{X} \right)$$

Nota 14 Si se estima \mathbf{L} por el método de las componentes principales, se suele usar el método mínimos cuadrados ordinarios (equivale a suponer que las ψ_i son iguales)

En este caso $\underline{\hat{F}}_k = (\hat{\mathbf{L}}^t \hat{\mathbf{L}})^{-1} \hat{\mathbf{L}}^t (\underline{X}_k - \underline{\bar{X}})$

Dado que

$$\begin{aligned} \hat{\mathbf{L}} &= \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1 & \sqrt{\hat{\lambda}_2} \hat{\mathbf{e}}_2 & \dots & \sqrt{\hat{\lambda}_m} \hat{\mathbf{e}}_m \end{bmatrix} \Rightarrow \hat{\mathbf{L}}^t \hat{\mathbf{L}} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_m) \Rightarrow \\ \Rightarrow \underline{\hat{F}}_k &= \begin{bmatrix} \frac{1}{\sqrt{\hat{\lambda}_1}} \hat{\mathbf{e}}_1^t (\underline{X}_k - \underline{\bar{X}}) \\ \vdots \\ \frac{1}{\sqrt{\hat{\lambda}_m}} \hat{\mathbf{e}}_m^t (\underline{X}_k - \underline{\bar{X}}) \end{bmatrix} \end{aligned}$$

En este caso la media muestral y varianza muestral de las estimaciones de las puntuaciones factoriales son:

$$\frac{1}{n} \sum_{k=1}^n \underline{\hat{F}}_k = \underline{0} \quad y \quad \frac{1}{n-1} \sum_{k=1}^n \underline{\hat{F}}_k \underline{\hat{F}}_k^t = \mathbf{I}_m$$

Obsérvese que las puntuaciones factoriales coinciden con los valores de las CP, salvo la constante multiplicativa de $\hat{\lambda}_j^{-1/2}$.

5.2 Método de regresión

Supone hipótesis de normalidad de los factores comunes \underline{F} y los factores específicos $\underline{\varepsilon}$.

Del modelo

$$\begin{matrix} \underline{X} \\ (p \times 1) \end{matrix} - \begin{matrix} \underline{\mu} \\ (p \times 1) \end{matrix} = \begin{matrix} \mathbf{L} \\ (p \times m) \end{matrix} \begin{matrix} \underline{F} \\ (m \times 1) \end{matrix} + \begin{matrix} \underline{\varepsilon} \\ (p \times 1) \end{matrix}$$

se deduce que $\underline{X} - \underline{\mu} = \begin{pmatrix} \mathbf{L} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \underline{F} \\ \underline{\varepsilon} \end{pmatrix} \sim N_p(\underline{0}, \mathbf{L}\mathbf{L}^t + \mathbf{\Psi})$.

- La distribución conjunta de $\underline{X} - \underline{\mu}$ y \underline{F} es una distribución normal

$$\begin{pmatrix} \underline{X} - \underline{\mu} \\ \underline{F} \end{pmatrix} = \begin{pmatrix} \mathbf{L} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \underline{F} \\ \underline{\varepsilon} \end{pmatrix} \sim N_{p+m}(\underline{0}, \mathbf{\Sigma}^*)$$

con

$$\mathbf{\Sigma}^* = \begin{bmatrix} \mathbf{L}\mathbf{L}^t + \mathbf{\Psi} & \mathbf{L} \\ \mathbf{L}^t & \mathbf{I}_m \end{bmatrix} = \begin{bmatrix} \mathbf{\Sigma} & \mathbf{L} \\ \mathbf{L}^t & \mathbf{I}_m \end{bmatrix}$$

- Por las propiedades de la distribución normal, la línea de regresión de \underline{F} sobre $\underline{X} - \underline{\mu}$ es

$$E[\underline{F} \mid \underline{X} = \underline{x}] = \mathbf{L}^t \boldsymbol{\Sigma}^{-1} (\underline{x} - \underline{\mu}) = \mathbf{L}^t \left(\mathbf{L} \mathbf{L}^t + \boldsymbol{\Psi} \right)^{-1} (\underline{x} - \underline{\mu})$$

y verifica

$$Cov[\underline{F} \mid \underline{X} = \underline{x}] = \mathbf{I}_m - \mathbf{L}^t \boldsymbol{\Sigma}^{-1} \mathbf{L} = \mathbf{I}_m - \mathbf{L}^t \left(\mathbf{L} \mathbf{L}^t + \boldsymbol{\Psi} \right)^{-1} \mathbf{L}.$$

- Dada una observación \underline{X}_k , se proponen como estimadores de las puntuaciones factoriales \underline{F}_k :

$$\hat{\underline{F}}_k = \hat{\mathbf{L}}^t \left(\hat{\mathbf{L}} \hat{\mathbf{L}}^t + \hat{\boldsymbol{\Psi}} \right)^{-1} (\underline{X}_k - \overline{\underline{X}}), \quad k = 1, \dots, n \quad (7)$$

$$\hat{\underline{F}}_k = \hat{\mathbf{L}}^t \hat{\boldsymbol{\Sigma}}^{-1} (\underline{X}_k - \overline{\underline{X}}) \quad (8)$$

con $\hat{\mathbf{L}}$ y $\hat{\boldsymbol{\Psi}}$ los estimadores de máxima verosimilitud.

- La expresión (7) es más correcta que (8), pero depende del número de factores considerados.

6 Medidas de adecuación muestral

Uno de los requisitos que debe cumplirse para que el AF tenga sentido es que las variables originales estén altamente correlacionadas. Si las correlaciones entre las variables son bajas, el AF no es apropiado dado que resultará imposible reducir significativamente la dimensión original (p).

Se espera que las variables que tienen correlación muy alta entre sí la tengan con el mismo factor o factores.

Se ha propuesto diversas medidas de adecuación muestral para que tenga sentido la aplicación del AF.

6.1 Índice de Kaiser-Meyer-Olkin (KMO)

- El coeficiente de correlación parcial es un indicador de la fuerza de las relaciones entre dos variables eliminando la influencia del resto.
- Si las variables comparten factores comunes, la correlación parcial entre pares de variables será baja, puesto que se eliminan los efectos lineales de las otras variables.
- Las correlaciones parciales son aproximaciones de las correlaciones entre los factores específicos, que son nulas según el modelo factorial

$$\left. \begin{array}{l} X_1 - \mu_1 = \sum_{j=1}^m l_{1j} F_j + \varepsilon_1 \\ X_2 - \mu_2 = \sum_{j=1}^m l_{2j} F_j + \varepsilon_2 \end{array} \right\} \Rightarrow \rho_{par}^2(X_1, X_2) \approx \rho^2(\varepsilon_1, \varepsilon_2) = 0.$$

- Por tanto si existe un número elevado de estos coeficientes distintos de cero es señal de que la aplicación del AF no es adecuada.

Índice de Kaiser-Meyer-Olkin (KMO)

Sea r_{ih} el coeficiente de correlación muestral entre las variables X_i y X_h , y $r_{ih \cdot \{1, \dots, p\}}$ al coeficiente de correlación parcial muestral entre X_i y X_h , en el conjunto de variables originales $\{X_1, \dots, X_p\}$.

Índice KMO

$$KMO = \frac{\sum_{i=1}^p \sum_{h>i} r_{ih}^2}{\sum_{i=1}^p \sum_{h>i} r_{ih}^2 + \sum_{i=1}^p \sum_{h>i} r_{ih \cdot \{1, \dots, p\}}^2}$$

Baremos propuesto por Kaiser

Valor del índice	Interpretación
$0.9 \leq KMO \leq 1$	Muy bueno
$0.8 < KMO \leq 0.9$	Bueno
$0.7 < KMO \leq 0.8$	Mediano
$0.6 < KMO \leq 0.7$	Mediocre
$0.5 < KMO \leq 0.6$	Bajo
$KMO \leq 0.5$	Inaceptable

Kaiser (1974). An index of factorial simplicity. Psychometrika, Vol. 39 (1), 31-36.

6.2 Medida de Adecuación Muestral (MSA) para cada variable

$$MSA_i = \frac{\sum_{h \neq i} r_{ih}^2}{\sum_{h \neq i} r_{ih}^2 + \sum_{h \neq i} r_{ih \cdot \{1, \dots, p\}}^2}$$

Un bajo valor de MSA_i indica que el modelo del AF es poco adecuado para la variable X_i .

De esta forma si el KMO es bajo es posible localizar las variables responsables y eliminar dichas variables del análisis.

7 Relación entre AF y ACP

- Similitud: Ambos tratan de explicar un conjunto de datos en un número menor de dimensiones
- Diferencias:
 - ACP es una mera transformación de los datos: las variables originales se transforman en las CP
 - ACP no hace hipótesis sobre la matriz Σ asociada al vector \underline{X}
 - AF supone que los datos proceden de un modelo bien definido, donde los factores verifican una serie de hipótesis. Si las hipótesis no se verifican los resultados del AF pueden ser engañosos.
 - En AF se centra en la identificación de los factores subyacentes a las variables originales

- Si el modelo del AF se verifica y las varianzas específicas son pequeñas, ACP y AF proporcionan resultados similares

Si las varianzas específicas son grandes, el ACP las absorbe en las CP, las que mantiene y las que rechaza, mientras que el AF las tiene en cuenta de forma especial.

	Principal component					Maximum likelihood				
Variable	Estimated rotated factor loadings, $\tilde{\ell}_{ij}^*$				Specific variances $\tilde{\psi}_i = 1 - \tilde{h}_i^2$	Estimated rotated factor loadings, $\hat{\ell}_{ij}^*$				Specific variances $\hat{\psi}_i = 1 - \hat{h}_i^2$
	F_1^*	F_2^*	F_3^*	F_4^*		F_1^*	F_2^*	F_3^*	F_4^*	
100-m run	.884	.136	.156	-.113	.16	.167	.857	.246	-.138	.16
Long jump	.631	.194	.515	-.006	.30	.240	.477	.580	.011	.38
Shot put	.245	.825	.223	-.148	.19	.966	.154	.200	-.058	.00
High jump	.239	.150	.750	.076	.35	.242	.173	.632	.113	.50
400-m run	.797	.075	.102	.468	.13	.055	.709	.236	.330	.33
110-m hurdles	.404	.153	.635	-.170	.38	.205	.261	.589	-.071	.54
Discus	.186	.814	.147	-.079	.28	.697	.133	.180	-.009	.46
Pole vault	-.036	.176	.762	.217	.34	.137	.078	.513	.116	.70
Javelin	-.048	.735	.110	.141	.43	.416	.019	.175	.002	.80
1500-m run	.045	-.041	.112	.934	.11	-.055	.056	.113	.990	.00
Cumulative proportion of total sample variance explained	.21	.42	.61	.73		.18	.34	.50	.61	

Example 13.6. The speaking rate of four voices was artificially manipulated by means of a rate changer without altering the pitch (Brown, Strong, and Rencher 1973). There were five rates for each voice:

FF = 45% faster,

F = 25% faster,

N = normal rate,

S = 22% slower,

SS = 42% slower.

The resulting 20 voices were played to 30 judges, who rated them on 15 paired-opposite adjectives (variables) with a 14-point scale between poles. The following adjectives were used: intelligent, ambitious, polite, active, confident, happy, just, likeable, kind, sincere, dependable, religious, good-looking, sociable, and strong. The results were averaged over the 30 judges to produce 20 observation vectors of 15 variables each. The averaging produced very reliable data, so that even though there were only 20 observations on 15 variables, the factor analysis model fit very well. The correlation matrix is as follows:

$$\mathbf{R} = \begin{pmatrix} 1.00 & .90 & -.17 & .88 & .92 & .88 & .15 & .39 & -.02 & -.16 & .52 & -.15 & -.79 & -.78 & .73 \\ .90 & 1.00 & -.46 & .93 & .87 & .79 & -.16 & .10 & -.35 & -.42 & .25 & -.40 & .68 & -.60 & .62 \\ -.17 & -.46 & 1.00 & -.56 & -.13 & .07 & .85 & .75 & .88 & .91 & .68 & .88 & .21 & .31 & .25 \\ .88 & .93 & -.56 & 1.00 & .85 & .73 & -.25 & -.02 & -.45 & -.57 & .10 & -.53 & .58 & .84 & .50 \\ .92 & .87 & -.13 & .85 & 1.00 & .91 & .20 & .39 & -.09 & -.16 & .49 & -.10 & .85 & .80 & .81 \\ .88 & .79 & .07 & .73 & .91 & 1.00 & .27 & .53 & .12 & .06 & .66 & .08 & .90 & .85 & .78 \\ .15 & -.16 & .85 & -.25 & .20 & .27 & 1.00 & .85 & .81 & .79 & .79 & .81 & .43 & .54 & .53 \\ .39 & .10 & .75 & -.02 & .39 & .53 & .85 & 1.00 & .84 & .79 & .93 & .77 & .71 & .69 & .76 \\ -.02 & -.35 & .88 & -.45 & -.09 & .12 & .81 & .84 & 1.00 & .91 & .76 & .85 & .28 & .36 & .35 \\ -.16 & -.42 & .91 & -.57 & -.16 & .06 & .79 & .79 & .91 & 1.00 & .72 & .96 & .26 & .28 & .29 \\ .52 & .25 & .67 & .10 & .49 & .66 & .79 & .93 & .76 & .72 & 1.00 & .72 & .75 & .77 & .78 \\ -.15 & -.40 & .88 & -.53 & -.10 & .08 & .81 & .77 & .85 & .96 & .72 & 1.00 & .33 & .32 & .34 \\ .79 & .68 & .21 & .58 & .85 & .90 & .43 & .71 & .28 & .26 & .75 & .33 & 1.00 & .86 & .92 \\ .78 & .60 & .31 & .54 & .80 & .85 & .54 & .69 & .36 & .28 & .77 & .32 & .86 & 1.00 & .82 \\ .73 & .62 & .25 & .50 & .81 & .78 & .53 & .76 & .35 & .29 & .78 & .34 & .92 & .82 & 1.00 \end{pmatrix}$$

The eigenvalues of \mathbf{R} are 7.91, 5.85, .31, .26, \dots , .002, with the scree plot in Figure 13.5. Clearly, by any criterion for choosing m , there are two factors.

All four major methods of factor extraction discussed in Section 13.3 produced nearly identical results (after rotation). We give the initial and rotated loadings obtained from the principal component method in Table 13.10.

The two rotated factors were labeled *competence* and *benevolence*. The same two factors emerged consistently in similar studies with different voices and different judges.

The two groupings of variables can also be seen in the correlation matrix. For example, in the first row, the large correlations correspond to the boldface rotated

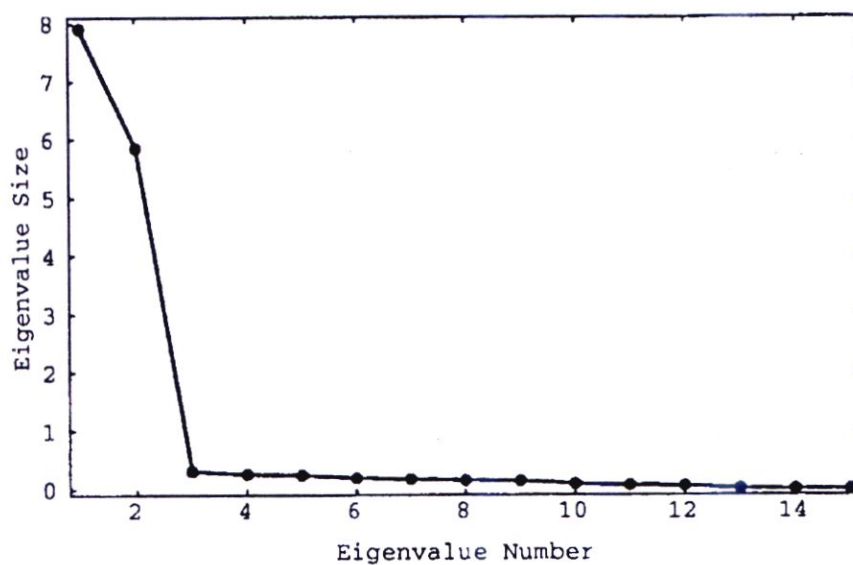


Figure 13.5. Scree graph for voice data.

Table 13.10. Initial and Varimax Rotated Loadings for the Voice Data

Variable	Initial Loadings		Rotated Loadings		Communalities
	\bar{f}_1	f_2	f_1	f_2	
Intelligent	.71	-.65	.96	-.06	.93
Ambitious	.48	-.84	.90	-.36	.94
Polite	.50	.81	-.12	.95	.92
Active	.37	-.91	.86	-.48	.97
Confident	.73	-.64	.97	-.04	.95
Happy	.83	-.47	.94	.15	.91
Just	.71	.58	.20	.89	.84
Likeable	.89	.39	.45	.87	.95
Kind	.58	.75	-.02	.95	.89
Sincere	.52	.82	-.11	.97	.95
Dependable	.93	.27	.56	.79	.94
Religious	.55	.79	-.07	.96	.92
Good looking	.91	-.29	.89	.35	.91
Sociable	.91	-.22	.84	.40	.87
Strong	.91	-.21	.84	.41	.86
Variance accounted for	7.91	5.85	7.11	6.65	13.76
Proportion of total variance	.53	.39	.47	.44	.92
Cumulative proportion	.53	.92	.47	.92	.92

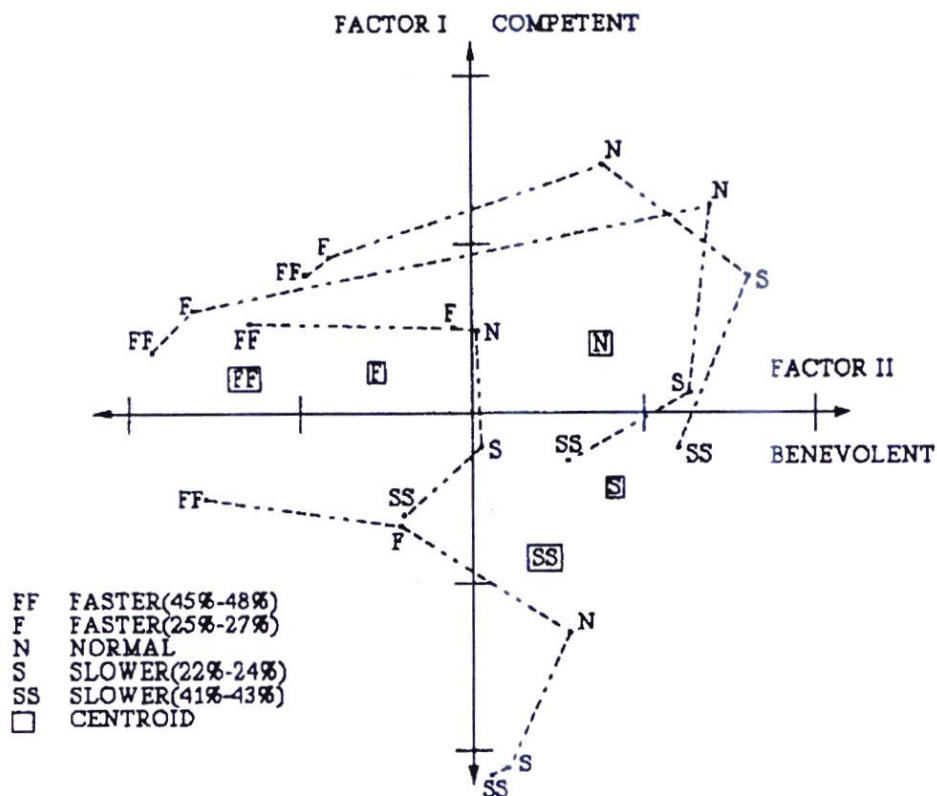


Figure 13.6. Factor scores of adjective rating of voices with five levels of manipulated rate.

loadings for f_1 , whereas in the third row, the large correlations correspond to the boldface rotated loadings for f_2 .

The factor scores were of primary interest in this study. The goal was to ascertain the effect of the rate manipulations on the two factors, that is, to determine the perceived change in competence and benevolence when the speaking rate is increased or decreased.

The two factor scores were obtained for each of the 20 voices; these are plotted in Figure 13.6, where a consistent effect of the manipulation of speaking rate on all four voices can clearly be seen. Decreasing the speaking rate causes the speaker to be rated less competent; increasing the rate causes the speaker to be rated less benevolent. The mean vectors (centroids) are also given in Figure 13.6 for the four speakers.