

# **Análise Fatorial**

Francisco Ricardo Duarte Carvalho



# Análise Fatorial

Francisco Ricardo Duarte Carvalho

Dissertação para a obtenção do Grau de **Mestre em Matemática**  
Área de Especialização em **Estatística, Otimização e Matemática Financeira**

## Júri

**Presidente:** Maria Emília Mesquita Nogueira

**Orientador:** Cristina Maria Tavares Martins

**Vogais:** Maria Emília Mesquita Nogueira

Maria da Graça Santos Temido Neves Mendes

**Data: Setembro 2013**



# Resumo

A análise fatorial é um método estatístico multivariado cujo objetivo é agrupar  $p$  variáveis aleatórias,  $X_1, \dots, X_p$ , em grupos formados por variáveis fortemente correlacionadas. Tais grupos constituem os chamados fatores ou variáveis latentes. Os fatores são variáveis aleatórias não observáveis, preferencialmente em número inferior ao das variáveis originais. Neste trabalho, considera-se o modelo fatorial ortogonal, no qual os fatores são ortogonais dois a dois. No modelo fatorial, cada variável original é escrita como combinação linear dos fatores comuns adicionada de um fator específico. Para estimar os coeficientes dos fatores comuns, denominados pesos fatoriais, são abordados dois métodos, nomeadamente, o método das componentes principais e o método da máxima verossimilhança, sendo o primeiro desenvolvido com mais pormenor. Neste sentido, começa-se por definir e obter as componentes principais de uma população. O procedimento correspondente envolve os valores próprios e os vetores próprios da matriz de correlações ou da matriz de variâncias-covariâncias das variáveis  $X_1, \dots, X_p$ . Seguidamente, obtêm-se as componentes principais amostrais e apresentam-se estimadores para os parâmetros envolvidos, em particular, estimadores de máxima verossimilhança no caso em que o vetor aleatório  $[X_1 \dots X_p]^T$  tem distribuição normal multivariada.

Referem-se vários critérios para escolher o número  $m$  de fatores,  $m < p$ , e, considerando os fatores como eixos ortogonais, aborda-se a rotação ortogonal dos mesmos, com vista a facilitar a sua interpretação.

Apesar dos fatores comuns serem variáveis não observáveis, é possível estimar o valor de cada fator (*score*) para cada indivíduo da amostra. Neste trabalho referem-se dois métodos para atingir esse objetivo: o método dos mínimos quadrados ponderados e o método da regressão.

Finalmente, apresenta-se um exemplo de aplicação da análise fatorial, desenvolvido com recurso ao software SPSS.

## Palavras Chave:

Análise fatorial, componentes principais, matriz de variâncias-covariâncias, valores e vetores próprios, lei normal multivariada, estimadores de máxima verossimilhança.

# Abstract

Factor analysis is a multivariate statistical method with the objective of grouping  $p$  random variables  $X_1, \dots, X_p$  in groups formed by strongly correlated variables. These groups are called factors or latent variables. The factors are unobservable random variables, preferably in smallest number that the original variables. In this work, is considered the orthogonal factorial model, in which the factors are orthogonal two

by two. In factor analysis model, each original variable is written as a linear combination of the common factors and added to a specific factor. To estimate the coefficients of common factors, called loadings, we will see two methods, namely, the method of principal components and the method of maximum likelihood, the first being developed in more detail. We starts by define and obtain principal components of a population. The corresponding procedure involves the eigenvalues and eigenvectors of the correlation matrix or the variance-covariance matrix of the variables  $X_1, \dots, X_p$ . Then, we obtain the principal components of sample and we present the estimators for involved parameters, in particular, maximum likelihood estimators in case that the random vector  $[X_1 \dots X_p]^T$  has multivariate normal distribution.

We refer several criteria to choose the number  $m$  of factors,  $m < p$ , and, considering the factors as orthogonal axes, we study the orthogonal rotation, to facilitate their interpretation.

Although the common factors are unobservable variables, we can estimate the value of each factor (score) for each element of sample. In this work we refer two methods to achieve this objective: the method of weighted least squares and the regression method.

Finally, we present an example of application of factor analysis, developed using the SPSS software.

**Keywords:** Factor analysis, principal components, variance-covariance matrix, eigenvalues and eigenvectors, multivariate normal distribution, maximum likelihood estimators.





# Agradecimentos

*É com muita satisfação que agradeço a todos aqueles que, direta ou indiretamente, tornaram a realização desta dissertação de mestrado possível.*

*À minha orientadora, a Doutora Cristina Maria Tavares Martins, pela sua disponibilidade, acompanhamento no estudo e paciência.*

*À minha namorada, Ana Gonçalves, por todo o apoio, pela sua compreensão e encorajamento.*

*À minha mãe, pelo enorme apoio que tem oferecido ao longo destes anos.*

*Ao meu pai, que embora já não esteja comigo, me incutiu princípios que tento seguir, entre os quais, nunca desistir.*



# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Componentes principais</b>	<b>5</b>
2.1	Dedução das componentes principais . . . . .	6
2.2	Estimação em componentes principais . . . . .	13
2.2.1	Componentes principais amostrais . . . . .	13
2.2.2	Estimação de parâmetros . . . . .	15
<b>3</b>	<b>Análise fatorial</b>	<b>23</b>
3.1	Modelo ortogonal . . . . .	24
3.2	Adequação da aplicação do modelo fatorial . . . . .	26
3.3	Extração dos fatores . . . . .	27
3.3.1	Método das componentes principais . . . . .	27
3.3.2	Método da máxima verossimilhança . . . . .	28
3.4	Escolha do número de fatores . . . . .	29
3.5	Rotação dos fatores . . . . .	32
3.6	Scores fatoriais . . . . .	35
3.6.1	Método dos mínimos quadrados ponderados . . . . .	35
3.6.2	Método da regressão . . . . .	36
3.6.3	Relação entre as estimativas obtidas pelos dois métodos . . . . .	37
<b>4</b>	<b>Exemplo de aplicação da Análise Fatorial</b>	<b>39</b>
4.1	Análise preliminar dos dados . . . . .	40
4.2	Aplicação da análise fatorial . . . . .	41



# Capítulo 1

## Introdução

As diversas ciências têm a pretensão de conhecer a realidade e de interpretar os acontecimentos, baseadas em variáveis intervenientes consideradas importantes nesses eventos. Frequentemente, tais variáveis são em número muito elevado e é natural que muitas delas sejam fortemente correlacionadas. Neste contexto, a análise fatorial surge como um procedimento estatístico que permite reduzir a complexidade do problema original, agrupando  $p$  variáveis aleatórias,  $X_1, \dots, X_p$ , em grupos formados por variáveis fortemente correlacionadas. Estes grupos dão origem aos fatores ou variáveis latentes que são variáveis aleatórias não observáveis, preferencialmente em número inferior ao das variáveis originais. Além disso, variáveis correspondentes a fatores distintos devem apresentar correlações pequenas.

Para ilustrar o contexto de aplicação da análise fatorial, considere-se o caso em que o diretor de uma fábrica de automóveis pretende entender o que leva o consumidor a escolher determinado modelo específico de automóvel, de forma a direcionar a produção nesse sentido. Para isso, foram consideradas as opiniões dos consumidores relativamente à importância de 14 variáveis na sua decisão (custos de reparação baixos, espaço interior amplo, facilidade de manuseamento, bom motor, ...). O elevado número de variáveis dificulta a sua avaliação e conseqüentemente o desenvolvimento de planos de ação. Aplicando a análise fatorial, as 14 variáveis passam a constituir 4 fatores relacionados com o conforto, a relação custo/eficiência, o estilo e a facilidade de manipulação, simplificando significativamente a tarefa do diretor.

Historicamente, a ideia subjacente à análise fatorial é atribuída aos psicólogos, que sempre tentaram uma descrição clara e organizada das habilidades intelectuais e comportamentos humanos, surgindo a noção primitiva de fator (ou constructo, na área da psicologia). No entanto, foi Karl Pearson (1901) o primeiro a introduzir a análise fatorial. Mais tarde, Charles Spearman (1904) publicou um artigo no qual menciona as bases da análise fatorial.

O modelo fatorial constituído pelos fatores  $F_1, \dots, F_m$ ,  $m \leq p$ , é definido por

$$X_1 - \mu_1 = \alpha_{11}F_1 + \alpha_{12}F_2 + \dots + \alpha_{1m}F_m + \epsilon_1$$

$$X_2 - \mu_2 = \alpha_{21}F_1 + \alpha_{22}F_2 + \dots + \alpha_{2m}F_m + \epsilon_2$$

...

$$X_p - \mu_p = \alpha_{p1}F_1 + \alpha_{p2}F_2 + \dots + \alpha_{pm}F_m + \epsilon_p,$$

onde  $\alpha_{ij}$  são coeficientes reais, chamados pesos fatoriais (*loadings*),  $\mu_i$  é o valor médio da variável  $X_i$  e  $\epsilon_i$  é uma variável aleatória que constitui o chamado fator específico da variável  $X_i$ ,  $i = 1, \dots, p$ . Note-se que cada fator específico está associado apenas a uma variável, enquanto os fatores  $F_1, \dots, F_m$  estão associados a todas as variáveis. Chamam-se, por isso, fatores comuns. O modelo fatorial mais usado é o modelo fatorial ortogonal, no qual os fatores são não correlacionados e têm variância 1.

Um dos problemas associados à análise fatorial é a estimação dos pesos fatoriais. Os métodos mais comuns para efetuar esta estimação são o método das componentes principais e o método da máxima verosimilhança. Este último parte do pressuposto de que o vetor aleatório  $\mathbf{X} = [X_1 \dots X_p]^T$  é normal multivariado. O método das componentes principais não faz exigências sobre a distribuição de  $\mathbf{X}$  sendo, por isso, o mais usado na prática. Tal como o nome indica, é baseado nas componentes principais,  $Y_1, \dots, Y_p$ , associadas às variáveis  $X_1, \dots, X_p$ . As componentes principais são combinações lineares normalizadas destas variáveis iniciais, não correlacionadas entre si e com variâncias decrescentes, ou seja,  $V(Y_1) \leq \dots \leq V(Y_p)$ . Frequentemente as variáveis  $X_1, X_2, \dots, X_p$  são de natureza diversa e algumas delas têm dispersões muito diferentes, pelo que é habitual considerarem-se as componentes principais obtidas das correspondentes variáveis standardizadas,  $Z_1, \dots, Z_p$ .

Dada a sua importância, o capítulo 2 deste trabalho é dedicado à obtenção das componentes principais de uma população e à estimação dos parâmetros subjacentes. O procedimento correspondente à determinação das componentes principais envolve os valores próprios e os vetores próprios da matriz de correlações ou da matriz de variâncias-covariâncias do vetor aleatório  $\mathbf{X}$  bem como o método dos multiplicadores de Lagrange.

No capítulo 3 estuda-se o modelo ortogonal de análise fatorial. Depois da definição do modelo, deduzem-se algumas propriedades decorrentes dos pressupostos envolvidos. Em particular, mostra-se que o peso fatorial  $\alpha_{ij}$  corresponde à covari-

ância entre a variável  $X_i$  e o fator  $F_j$ , pelo que aquele coeficiente é uma medida da relação entre  $X_i$  e  $F_j$ . Assim, os pesos fatoriais elevados (em valor absoluto) identificam o fator a que cada variável se associa. Outra propriedade importante prende-se com o facto da matriz dos pesos fatoriais não ser única, podendo ser rotacionada de modo a produzir novos pesos fatoriais. No entanto, esta rotação não altera a variância de  $X_i$ ,  $i = 1, \dots, p$ .

Na prática, só faz sentido aplicar a análise fatorial se as variáveis  $X_1, \dots, X_n$  forem correlacionadas. Para verificar se os dados corroboram esta hipótese, apresentam-se dois procedimentos, nomeadamente, o teste de esfericidade de Bartlett, desenvolvido sob a hipótese de normalidade do vetor aleatório  $\mathbf{X}$ , e a estatística KMO que não envolve condições sobre a distribuição de probabilidade de  $\mathbf{X}$ . Apresentam-se os métodos de estimação dos pesos fatoriais já mencionados, destacando-se o método das componentes principais, e referem-se processos para escolher do número de fatores a considerar na análise fatorial que, como já foi indicado, devem ser em menor número do que as variáveis iniciais.

Os  $m$  fatores determinados podem ser vistos como um sistema de eixos no qual se representam os  $p$  pontos  $(\alpha_{i1}, \dots, \alpha_{im})$ ,  $i = 1, \dots, p$ . Tendo em conta o que foi referido sobre a rotação da matriz dos pesos fatoriais, tais eixos podem ser rotacionados com o objetivo de conseguir que os pesos fatoriais elevados (resp., baixos) fiquem ainda mais elevados (resp., baixos) de modo a facilitar a interpretação dos fatores. Existem vários processos para escolher a matriz de rotação correspondente. Neste trabalho damos especial atenção ao método Varimax.

Apesar dos fatores comuns serem variáveis não observáveis, é possível estimar o valor de cada fator (*score*) para cada indivíduo da amostra. O capítulo 3 termina com a apresentação de dois métodos para fazer essa estimação: o método dos mínimos quadrados ponderados e o método da regressão.

No último capítulo, apresenta-se um exemplo de aplicação da análise fatorial, desenvolvido com recurso ao software SPSS.



# Capítulo 2

## Componentes principais

O objetivo essencial da análise de componentes principais é descrever a estrutura de covariâncias entre  $p$  variáveis aleatórias reais,  $X_1, X_2, \dots, X_p$ , efetivamente correlacionadas entre si, num outro conjunto de  $p$  variáveis não correlacionadas que expliquem a estrutura de variâncias-covariâncias das variáveis iniciais. Estas novas  $p$  variáveis resultam de combinações lineares normalizadas das variáveis originais e constituem as componentes principais.

O processo subjacente à obtenção das componentes principais é construtivo na medida que, a cada passo, se impõe que a nova componente principal seja não correlacionada com as anteriores e que as componentes principais surjam com variâncias decrescentes.

Frequentemente as variáveis  $X_1, X_2, \dots, X_p$  são de natureza diversa e algumas delas têm dispersões muito diferentes, pelo que é habitual considerarem-se as correspondentes variáveis standardizadas,  $Z_1, Z_2, \dots, Z_p$ . Neste sentido, deduzem-se, na secção 2.1, as componentes principais,  $Y_1, Y_2, \dots, Y_p$ , associadas às variáveis aleatórias standardizadas e apresentam-se resultados que relacionam a estrutura de correlações destes dois conjuntos de variáveis. Apresentam-se ainda as componentes principais associadas às variáveis iniciais,  $X_1, X_2, \dots, X_p$ , e deduzem-se os resultados correspondentes.

Na prática, não conhecemos a matriz de correlações nem a matriz de variâncias-covariâncias, pelo que as componentes principais são estimadas a partir da matriz de correlações ou da matriz de variâncias-covariâncias de uma amostra aleatória do vetor aleatório real  $\mathbf{X}$ . Tais componentes principais estimadas designam-se por componentes principais amostrais e são introduzidas na secção 2.2. Nesta secção faz-se também a estimação da matriz de variâncias-covariâncias, bem como dos seus valores próprios e vetores próprios associados. Em particular, determinam-se estimadores de máxima verosimilhança no caso em que  $\mathbf{X}$  segue uma lei normal multivariada. Analisa-se ainda a centricidade dos estimadores obtidos.

## 2.1. Dedução das componentes principais

Seja  $\mathbf{X} = [X_1 X_2 \dots X_p]^T$  um vetor aleatório real de dimensão  $p$  com vetor média  $\boldsymbol{\mu} = [\mu_1 \mu_2 \dots \mu_p]^T$  e matriz de variâncias-covariâncias  $\Sigma$ . Seja  $\mathbf{Z} = [Z_1 Z_2 \dots Z_p]^T$  o vetor aleatório das correspondentes variáveis standardizadas, isto é,

$$Z_j = \frac{X_j - \mu_j}{\sigma_j},$$

com  $\sigma_j^2 = V(X_j)$ ,  $j = 1, 2, \dots, p$ . Tendo em conta que

$$\text{Cov}(Z_i, Z_j) = \frac{1}{\sigma_i \sigma_j} \text{Cov}(X_i, X_j), \quad i, j = 1, \dots, p,$$

a matriz de variâncias-covariâncias de  $\mathbf{Z}$  coincide com a matriz de correlação de  $\mathbf{X}$ , que aqui designamos por  $C$ .

O vetor aleatório real  $\mathbf{Z}$  tem associadas  $p$  componentes principais,  $Y_1, Y_2, \dots, Y_p$ , tais que

$$Y_j = a_{1j}Z_1 + a_{2j}Z_2 + \dots + a_{pj}Z_p = \sum_{i=1}^p a_{ij}Z_i, \quad j = 1, 2, \dots, p,$$

onde os coeficientes  $a_{ij}$ ,  $i, j = 1, 2, \dots, p$ , são constantes reais satisfazendo determinadas condições. As  $p$  componentes principais podem ser representadas na forma matricial  $\mathbf{Y} = A^T \mathbf{Z}$ , com  $A = [a_{ij}]_{i,j=1,2,\dots,p}$ .

Prova-se facilmente que o vetor média e a matriz de variâncias-covariâncias de  $\mathbf{Y}$  são, respetivamente,

$$\boldsymbol{\mu}_{\mathbf{Y}} = \mathbf{0} \quad \text{e} \quad \Sigma_{\mathbf{Y}} = A^T C A.$$

De facto,  $\boldsymbol{\mu}_{\mathbf{Y}} = E(A^T \mathbf{Z}) = A^T E(\mathbf{Z}) = \mathbf{0}$ . No que diz respeito à matriz de variâncias-covariâncias  $\Sigma_{\mathbf{Y}}$ , considere-se o seu elemento genérico  $\text{Cov}(Y_r, Y_s)$ . Como  $\boldsymbol{\mu}_{\mathbf{Y}} = \mathbf{0}$ , tem-se

$$\begin{aligned} \text{Cov}(Y_r, Y_s) &= E(Y_r Y_s) = E\left(\sum_{i=1}^p a_{ir} Z_i \sum_{i=1}^p a_{is} Z_i\right) \\ &= \sum_{i=1}^p \sum_{j=1}^p a_{ir} a_{js} E(Z_i Z_j) \\ &= \sum_{i=1}^p \sum_{j=1}^p a_{ir} a_{js} \text{Cov}(Z_i, Z_j) \\ &= \mathbf{a}_r^T C \mathbf{a}_s, \quad r, s = 1, 2, \dots, p, \end{aligned} \tag{2.1}$$

onde  $\mathbf{a}_j$  representa a  $j$ -ésima coluna de  $A$ ,  $j = 1, 2, \dots, p$ . Ora  $\mathbf{a}_r^T C \mathbf{a}_s$  é precisamente o elemento da linha  $r$  e da coluna  $s$  da matriz  $A^T C A$ .

As condições a que devem obedecer os coeficientes  $a_{ij}$ ,  $i, j = 1, 2, \dots, p$ , são as seguintes:

- $V(Y_1) \geq V(Y_2) \geq \dots \geq V(Y_p)$ ;
- $\mathbf{a}_j^T \mathbf{a}_j = 1$ ,  $j = 1, 2, \dots, p$ ;
- $Cov(Y_j, Y_k) = 0$ ,  $j \neq k$ ,  $j, k = 1, 2, \dots, p$ .

Assim, o vetor  $\mathbf{a}_1$  é determinado de forma a maximizar a variância de  $Y_1$  e tal que  $\mathbf{a}_1^T \mathbf{a}_1 = 1$  e, para  $j = 2, \dots, p$ , o vetor  $\mathbf{a}_j$  é determinado de forma a maximizar a variância de  $Y_j$ , satisfazendo  $\mathbf{a}_j^T \mathbf{a}_j = 1$  e  $Cov(Y_k, Y_j) = 0$ ,  $k < j$ .

No que se segue, admitimos que a matriz  $C$  é definida positiva. O primeiro resultado desta secção estabelece que as componentes principais associadas às variáveis aleatórias  $Z_1, \dots, Z_p$  são determinadas à custa dos valores próprios e dos vetores próprios da matriz  $C$ . Como esta matriz é definida positiva, os seus valores próprios são todos positivos.

**Teorema 1.** *Sejam  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$  os valores próprios da matriz  $C$  e  $\mathbf{e}_j = [e_{1j} \ e_{2j} \ \dots \ e_{pj}]^T$  um vetor próprio normado associado a  $\lambda_j$ ,  $j = 1, 2, \dots, p$ . A  $j$ -ésima componente principal do vetor aleatório  $\mathbf{Z}$  é dada por*

$$Y_j = \mathbf{e}_j^T \mathbf{Z} = e_{1j}Z_1 + e_{2j}Z_2 + \dots + e_{pj}Z_p, \quad j = 1, \dots, p.$$

Tem-se ainda que

$$V(Y_j) = \lambda_j \quad e \quad Cov(Y_j, Y_k) = 0, \quad j \neq k, \quad j, k = 1, \dots, p.$$

*Demonstração.* Usamos o método multiplicadores de Lagrange para determinar  $\mathbf{a}_1 = [a_{11} \ a_{21} \ \dots \ a_{p1}]^T$  de modo a maximizar  $V(Y_1)$  sujeito a  $\mathbf{a}_1^T \mathbf{a}_1 = 1$ .

De (2.1), tem-se  $V(Y_1) = \mathbf{a}_1^T C \mathbf{a}_1$ . Considerem-se as funções

$$f(\mathbf{a}_1) = f(a_{11}, a_{21}, \dots, a_{p1}) = V(Y_1) = \mathbf{a}_1^T C \mathbf{a}_1 = \sum_{i=1}^p \sum_{k=1}^p a_{i1} a_{k1} Cov(Z_i, Z_k)$$

e

$$g(\mathbf{a}_1) = g(a_{11}, a_{21}, \dots, a_{p1}) = \mathbf{a}_1^T \mathbf{a}_1 - 1 = \sum_{i=1}^p a_{i1}^2 - 1.$$

A função de Lagrange é dada por

$$\phi_1(\mathbf{a}_1) = \mathbf{a}_1^T C \mathbf{a}_1 - l(\mathbf{a}_1^T \mathbf{a}_1 - 1),$$

onde  $l$  é o multiplicador de Lagrange. O máximo da função será solução do sistema

$$\begin{cases} \nabla f = l \nabla g \\ g = 0 \end{cases} \Leftrightarrow \begin{cases} \nabla f = l \nabla g \\ \mathbf{a}_1^T \mathbf{a}_1 = 1 \end{cases} \quad (2.2)$$

Relativamente ao gradiente de  $g$ , tem-se

$$\frac{\partial g}{\partial a_{j1}} = 2a_{j1}, \quad j = 1, 2, \dots, p.$$

## Capítulo 2 Componentes principais

---

Quanto ao gradiente de  $f$ , verifica-se que

$$\frac{\partial f}{\partial a_{j1}} = 2 \sum_{k=1}^p a_{k1} \text{Cov}(Z_k, Z_j), \quad j = 1, \dots, p.$$

De facto,

$$\begin{aligned} \frac{\partial f}{\partial a_{j1}} &= \sum_{i=1}^p \sum_{k=1}^p \frac{\partial}{\partial a_{j1}} (a_{i1} a_{k1} \text{Cov}(Z_i, Z_k)) \\ &= \sum_{i=1}^p \frac{\partial}{\partial a_{j1}} (a_{i1}^2 V(Z_i)) + \sum_{i=1}^p \sum_{\substack{k=1 \\ k \neq i}}^p \frac{\partial}{\partial a_{j1}} (a_{i1} a_{k1} \text{Cov}(Z_i, Z_k)) \\ &= 2a_{j1} V(Z_j) + \sum_{\substack{k=1 \\ k \neq j}}^p \frac{\partial}{\partial a_{j1}} (a_{j1} a_{k1} \text{Cov}(Z_j, Z_k)) + \sum_{\substack{i=1 \\ i \neq j}}^p \frac{\partial}{\partial a_{j1}} (a_{i1} a_{j1} \text{Cov}(Z_i, Z_j)) \\ &= 2a_{j1} V(Z_j) + \sum_{\substack{k=1 \\ k \neq j}}^p a_{k1} \text{Cov}(Z_j, Z_k) + \sum_{\substack{i=1 \\ i \neq j}}^p a_{i1} \text{Cov}(Z_i, Z_j) \\ &= 2a_{j1} V(Z_j) + 2 \sum_{\substack{k=1 \\ k \neq j}}^p a_{k1} \text{Cov}(Z_k, Z_j) \\ &= 2 \sum_{k=1}^p a_{k1} \text{Cov}(Z_k, Z_j). \end{aligned}$$

Então

$$\begin{aligned} \nabla f = l \nabla g &\Leftrightarrow \sum_{k=1}^p a_{k1} \text{Cov}(Z_k, Z_j) = l a_{j1}, \quad j = 1, 2, \dots, p \\ &\Leftrightarrow \begin{bmatrix} \text{Cov}(Z_1, Z_j) & \dots & \text{Cov}(Z_p, Z_j) \end{bmatrix} \begin{bmatrix} a_{11} \\ \vdots \\ a_{p1} \end{bmatrix} = l a_{j1}, \quad j = 1, \dots, p \\ &\Leftrightarrow C \mathbf{a}_1 = l \mathbf{a}_1 \\ &\Leftrightarrow (C - l I_p) \mathbf{a}_1 = 0, \end{aligned}$$

onde  $I_p$  representa a matriz identidade de ordem  $p$ .

Para encontrar uma solução  $\mathbf{a}_1 \neq 0$  da equação  $(C - l I_p) \mathbf{a}_1 = 0$ , a matriz  $C - l I_p$  deve ser singular, isto é,  $\det(C - l I_p) = 0$ . Por outro lado, atendendo a (2.2)

$$V(Y_1) = \mathbf{a}_1^T C \mathbf{a}_1 = \mathbf{a}_1^T l \mathbf{a}_1 = l.$$

Desta forma, para que  $V(Y_1)$  seja máxima, devemos utilizar a maior das  $p$  raízes reais do polinómio  $\det(C - l I_p)$ , ou seja, o maior valor próprio de  $C$ ,  $\lambda_1$ . Consequentemente, obtém-se  $\mathbf{a}_1 = \mathbf{e}_1$ , atendendo à restrição  $\mathbf{a}_1^T \mathbf{a}_1 = 1$ . Assim, a primeira componente principal é dada por  $Y_1 = \mathbf{e}_1^T \mathbf{Z}$ .

O passo seguinte consiste em encontrar  $\mathbf{a}_2$  de forma a maximizar  $V(Y_2)$  de modo que  $Cov(Y_2, Y_1) = 0$  e  $\mathbf{a}_2^T \mathbf{a}_2 = 1$ . Tendo em conta que  $E(Y_j) = 0$ ,  $j = 1, 2, \dots, p$ , tem-se

$$Cov(Y_2, Y_1) = E(Y_2 Y_1) = E(\mathbf{a}_2^T \mathbf{Z} \mathbf{Z}^T \mathbf{e}_1) = \mathbf{a}_2^T C \mathbf{e}_1 = \lambda_1 \mathbf{a}_2^T \mathbf{e}_1.$$

Considerem-se as funções

$$f(\mathbf{a}_2) = V(Y_2) = \mathbf{a}_2^T C \mathbf{a}_2$$

$$g_1(\mathbf{a}_2) = \mathbf{a}_2^T \mathbf{a}_2 - 1$$

$$g_2(\mathbf{a}_2) = \lambda_1 \mathbf{a}_2^T \mathbf{e}_1$$

A função de Lagrange é então dada por

$$\phi_2(\mathbf{a}_2) = \mathbf{a}_2^T C \mathbf{a}_2 - l_1(\mathbf{a}_2^T \mathbf{a}_2 - 1) - l_2 \lambda_1 \mathbf{a}_2^T \mathbf{e}_1,$$

onde  $l_1$  e  $l_2$  são multiplicadores de Lagrange. Efetuando cálculos análogos aos do caso anterior, tem-se que o máximo da função é solução do sistema

$$\begin{cases} \nabla f = l_1 \nabla g_1 + l_2 \nabla g_2 \\ g_1 = 0 \\ g_2 = 0 \end{cases} \Leftrightarrow \begin{cases} 2C\mathbf{a}_2 = 2l_1 \mathbf{a}_2 + l_2 \lambda_1 \mathbf{e}_1 \\ \mathbf{a}_2^T \mathbf{a}_2 - 1 = 0 \\ \mathbf{a}_2^T \mathbf{e}_1 = 0 \end{cases} \quad (2.3)$$

Multiplicando ambos os membros da primeira equação de (2.3), à esquerda, por  $\mathbf{e}_1^T$  resulta, atendendo a que a matriz  $C$  é simétrica e  $\mathbf{e}_1^T \mathbf{a}_2 = (\mathbf{a}_2^T \mathbf{e}_1)^T = 0$ ,

$$\begin{aligned} 2C\mathbf{a}_2 = 2l_1 \mathbf{a}_2 + l_2 C \mathbf{e}_1 &\Leftrightarrow 2\mathbf{e}_1^T C \mathbf{a}_2 - 2l_1 \mathbf{e}_1^T \mathbf{a}_2 - l_2 \lambda_1 \mathbf{e}_1^T \mathbf{e}_1 = 0 \\ &\Leftrightarrow 2\mathbf{a}_2^T C \mathbf{e}_1 - l_2 \lambda_1 = 0 \\ &\Leftrightarrow 2\lambda_1 \mathbf{a}_2^T \mathbf{e}_1 - l_2 \lambda_1 = 0 \\ &\Leftrightarrow l_2 = 0, \end{aligned}$$

uma vez que  $\lambda_1 > 0$ . A primeira igualdade de (2.3) é então equivalente a  $C\mathbf{a}_2 = l_1 \mathbf{a}_2$ . Temos assim um caso análogo ao anterior, obtendo-se  $l_1 = \lambda_2$  e  $\mathbf{a}_2 = \mathbf{e}_2$ , vetor próprio de norma 1 associado ao valor próprio  $\lambda_2$ . A segunda componente principal é então dada por  $Y_2 = \mathbf{e}_2^T \mathbf{Z}$ .

De uma forma geral, para gerar a  $j$ -ésima componente principal,  $j = 2, \dots, p$ , determina-se  $\mathbf{a}_j$  de forma a maximizar a variância de  $Y_j = \mathbf{a}_j^T \mathbf{Z}$ , sujeito a  $\mathbf{a}_j^T \mathbf{a}_j = 1$  e à não correlação de  $Y_j$  com  $Y_1, \dots, Y_{j-1}$ . Sejam

$$f(\mathbf{a}_j) = V(Y_j) = \mathbf{a}_j^T C \mathbf{a}_j$$

$$g_1(\mathbf{a}_j) = \mathbf{a}_j^T \mathbf{a}_j - 1$$

$$g_i(\mathbf{a}_j) = \lambda_{i-1} \mathbf{a}_j^T \mathbf{e}_{i-1}, \quad i = 2, \dots, j$$

A função de Lagrange é agora

$$\phi_j(\mathbf{a}_j) = \mathbf{a}_j^T C \mathbf{a}_j - l_1 (\mathbf{a}_j^T \mathbf{a}_j - 1) - \sum_{i=2}^j l_i \lambda_{i-1} \mathbf{a}_j^T \mathbf{e}_{i-1},$$

onde  $l_1, l_2, \dots, l_j$  são multiplicadores de Lagrange. O máximo de  $\phi_j$  é solução do sistema

$$\begin{cases} \nabla f = l_1 \nabla g_1 + \sum_{i=2}^j l_i \nabla g_i \\ g_1 = 0 \\ g_i = 0, \quad i = 2, \dots, j \end{cases} \Leftrightarrow \begin{cases} 2 C \mathbf{a}_j = 2 l_1 \mathbf{a}_j + \sum_{i=2}^j l_i \lambda_{i-1} \mathbf{e}_{i-1} \\ \mathbf{a}_j^T \mathbf{a}_j - 1 = 0 \\ \mathbf{a}_j^T \mathbf{e}_{i-1} = 0, \quad i = 2, \dots, j \end{cases} \quad (2.4)$$

Após multiplicarmos ambos os membros da primeira igualdade de (2.4) por  $\mathbf{e}_{j-1}^T$ , à esquerda, obtém-se  $l_j = 0$ , de modo análogo ao caso anterior. Aquela igualdade é então equivalente a

$$2 C \mathbf{a}_j = 2 l_1 \mathbf{a}_j + \sum_{i=2}^{j-1} l_i \lambda_{i-1} \mathbf{e}_{i-1}.$$

Seguidamente, a multiplicação de ambos os membros desta igualdade por  $\mathbf{e}_{j-2}^T$ , à esquerda, conduz a  $l_{j-1} = 0$ . Este procedimento é, pois, repetido  $j - 1$  vezes, concluindo-se que  $l_2 = \dots = l_j = 0$ . Assim, a primeira igualdade de (2.4) é equivalente a  $(C - l_1 I_p) \mathbf{a}_j = 0$ . Obtém-se então, analogamente aos casos anteriores,  $l_1 = \lambda_j$  e  $\mathbf{a}_j = \mathbf{e}_j$ . A  $j$ -ésima componente principal é então dada por  $Y_j = \mathbf{e}_j^T \mathbf{Z}$ .  $\square$

Note-se que a componente principal  $Y_j$ ,  $j = 1, \dots, p$ , não é única por não ser único o vetor  $\mathbf{e}_j$ . No entanto, quando o valor próprio  $\lambda_j$  tem multiplicidade 1, caso em que há duas possibilidades (simétricas) para a escolha de  $\mathbf{e}_j$ , podemos ter unicidade se fixarmos  $\mathbf{e}_j$  como o vetor cuja primeira componente não nula é positiva.

A proposição seguinte estabelece que, no seu conjunto, as componentes principais mantêm a variância total das variáveis iniciais.

**Proposição 1.** *Sejam  $Y_1, Y_2, \dots, Y_p$  as  $p$  componentes principais correspondentes às variáveis aleatórias reais  $Z_1, Z_2, \dots, Z_p$ . Tem-se*

$$\sum_{j=1}^p V(Y_j) = \sum_{j=1}^p V(Z_j) = p.$$

*Demonstração.* Seja  $\Gamma$  a matriz cujas colunas são os vetores próprios  $\mathbf{e}_1, \dots, \mathbf{e}_p$  associados aos valores próprios  $\lambda_1, \dots, \lambda_p$ , respetivamente, da matriz  $C$ . Como  $\mathbf{e}_1, \dots, \mathbf{e}_p$

são normados e mutuamente ortogonais, a matriz  $\Gamma$  é ortogonal e tem-se

$$\Sigma_{\mathbf{Y}} = \Gamma^T C \Gamma, \text{ com } \Sigma_{\mathbf{Y}} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_p \end{bmatrix}$$

Então

$$\sum_{j=1}^p V(Y_j) = \text{tr}(\Sigma_{\mathbf{Y}}) = \text{tr}(\Gamma^T C \Gamma) = \text{tr}(\Gamma \Gamma^T C) = \text{tr}(C) = \sum_{j=1}^p V(Z_j) = p.$$

□

Este resultado permite afirmar que a proporção da variância total das variáveis standardizadas explicada pela  $j$ -ésima componente principal é dada por

$$\frac{\lambda_j}{p}, \quad j = 1, 2, \dots, p.$$

Assim, a percentagem de variância total explicada pelas  $k$  primeiras componentes principais,  $k \leq p$ , é

$$\frac{\lambda_1 + \lambda_2 + \dots + \lambda_k}{p} \times 100\%.$$

Se grande parte da percentagem da variância total for atribuída às primeiras  $k$  componentes principais ( $k < p$ ), então estas podem ser usadas em vez das  $p$  variáveis  $Z_1, Z_2, \dots, Z_p$  sem perda significativa de informação.

No resultado seguinte, verifica-se que o coeficiente de correlação entre a componente principal  $Y_j$  e a variável  $Z_k$ , é proporcional à componente  $e_{kj}$  do vetor próprio  $\mathbf{e}_j$ ,  $j, k = 1, 2, \dots, p$ .

**Proposição 2.** *Sejam  $Y_1, Y_2, \dots, Y_p$  as componentes principais das variáveis aleatórias reais  $Z_1, Z_2, \dots, Z_p$ . O coeficiente de correlação entre  $Y_j$  e  $Z_k$  é dado por*

$$\rho_{(Y_j, Z_k)} = e_{kj} \sqrt{\lambda_j}, \quad j, k = 1, 2, \dots, p.$$

*Demonstração.* Para cada  $k = 1, \dots, p$ , tem-se  $Z_k = \mathbf{u}_k^T \mathbf{Z}$ , onde  $\mathbf{u}_k^T = [0 \dots 0 \ 1 \ 0 \dots 0]$ , com o valor 1 na posição  $k$ . Como  $Y_j = \mathbf{e}_j^T \mathbf{Z}$ , obtemos

$$\begin{aligned} \text{Cov}(Z_k, Y_j) &= \text{Cov}(\mathbf{u}_k^T \mathbf{Z}, \mathbf{e}_j^T \mathbf{Z}) = \mathbf{u}_k^T C \mathbf{e}_j \\ &= \mathbf{u}_k^T \lambda_j \mathbf{e}_j = \lambda_j e_{kj}, \quad j, k = 1, \dots, p. \end{aligned}$$

Assim

$$\rho_{(Y_j, Z_k)} = \frac{Cov(Y_j, Z_k)}{\sqrt{V(Y_j)}\sqrt{V(Z_k)}} = \frac{\lambda_j e_{kj}}{\sqrt{\lambda_j}} = e_{kj} \sqrt{\lambda_j}, \quad j, k = 1, 2, \dots, p.$$

□

Como já foi referido, as componentes principais também podem ser obtidas através das variáveis originais,  $X_1, \dots, X_p$ . Neste caso, o processo de obtenção das componentes principais é análogo ao que foi descrito para as variáveis standardizadas, usando agora a matriz de variâncias-covariâncias,  $\Sigma$ , do vetor aleatório  $\mathbf{X} = [X_1 \ X_2 \ \dots \ X_p]^T$ . No entanto, se o vetor média,  $\boldsymbol{\mu}$ , de  $\mathbf{X}$  for diferente de zero, o referido procedimento envolve cálculos mais elaborados.

Continuaremos a usar as notações  $Y_j$ ,  $\lambda_j$  e  $\mathbf{e}_j$ ,  $j = 1, \dots, p$ , com o mesmo significado, mas agora relativamente às variáveis  $X_1, \dots, X_p$  e à matriz  $\Sigma$ .

Analogamente ao resultado obtido na secção anterior para as variáveis standardizadas, a componente principal  $Y_j$  associada às variáveis  $X_1, X_2, \dots, X_p$  é dada por

$$Y_j = \mathbf{e}_j^T \mathbf{X}, \quad j = 1, \dots, p.$$

Além disso, usando  $\Sigma$  em vez de  $C$  na demonstração da proposição 1, obtém-se

$$\sum_{j=1}^p V(Y_j) = \sum_{j=1}^p V(X_j),$$

ou seja, a variância total das variáveis originais também é preservada pelo conjunto das correspondentes componentes principais. Neste caso, a proporção de variância total explicada pela  $j$ -ésima componente principal é  $\frac{\lambda_j}{\lambda_1 + \dots + \lambda_p}$ ,  $j = 1, \dots, p$ .

O resultado correspondente à proposição 2 para as variáveis  $X_1, \dots, X_p$  é, como facilmente se verifica,

$$\rho_{(Y_j, X_k)} = \frac{e_{kj} \sqrt{\lambda_j}}{\sqrt{V(X_k)}}, \quad j, k = 1, 2, \dots, p.$$

Finalmente, como se ilustra no exemplo seguinte, notamos que as componentes principais correspondentes às variáveis originais não são, em geral, iguais às componentes principais correspondentes às variáveis standardizadas.

**Exemplo 1.** Considere-se um vetor aleatório  $\mathbf{X} = [X_1 \ X_2]^T$  com matriz de variâncias-covariâncias  $\Sigma$  tal que

$$\Sigma = \begin{bmatrix} 4 & -1 \\ -1 & 1 \end{bmatrix}$$

Os valores próprios de  $\Sigma$  são  $\lambda_2 = 4.3028$  e  $\lambda_1 = 0.6972$ , tendo como correspondentes vetores próprios normados  $\mathbf{e}_1 = [0.9571 \quad -0.2898]^T$  e  $\mathbf{e}_2 = [0.2898 \quad 0.9571]^T$ . As componentes principais correspondentes são então dadas por

$$Y_1 = 0.9571 X_1 - 0.2898 X_2$$

$$Y_2 = 0.2898 X_1 + 0.9571 X_2$$

A matriz de correlações correspondente a  $\Sigma$  é dada por

$$C = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$$

Tem-se agora  $\lambda_1 = 1.5$ ,  $\lambda_2 = 0.5$ ,  $\mathbf{e}_1 = [0.7071 \quad -0.7071]^T$  e  $\mathbf{e}_2 = [0.7071 \quad 0.7071]^T$ .

Neste caso, as componentes principais são dadas por

$$\begin{aligned} Y_1 &= 0.7071 Z_1 - 0.7071 Z_2 \\ &= 0.7071 \left( \frac{X_1 - \mu_1}{2} \right) - 0.7071 (X_2 - \mu_2) \\ &= 0.3536 (X_1 - \mu_1) - 0.7071 (X_2 - \mu_2) \\ Y_2 &= 0.7071 Z_1 + 0.7071 Z_2 \\ &= 0.3536 (X_1 - \mu_1) + 0.7071 (X_2 - \mu_2), \end{aligned}$$

onde  $[\mu_1 \quad \mu_2]^T$  representa o vetor média de  $\mathbf{X}$ .

Podemos ainda observar que, no primeiro caso, a percentagem da variância total explicada pela primeira componente principal é aproximadamente 86% e, segundo no caso, tal percentagem é 75%.

## 2.2. Estimação em componentes principais

### 2.2.1. Componentes principais amostrais

Seja  $(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$  uma amostra aleatória de dimensão  $n$  do vetor aleatório real  $\mathbf{X} = [X_1 \quad X_2 \quad \dots \quad X_p]^T$ , onde  $\mathbf{X}_j = [X_{1j} \quad X_{2j} \quad \dots \quad X_{pj}]^T$ ,  $j = 1, 2, \dots, n$ . O vetor média amostral e a matriz de variâncias-covariâncias amostral são dados, respetivamente, por

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{j=1}^n \mathbf{X}_j \quad \text{e} \quad S = \frac{1}{n-1} \sum_{j=1}^n (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^T.$$

Tem-se  $\bar{\mathbf{X}} = [\bar{X}_1 \quad \bar{X}_2 \quad \dots \quad \bar{X}_p]^T$  e  $S = [S_{ik}]_{i,k=1,2,\dots,p}$ , com  $\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{ij}$  e

$$S_{ik} = \frac{1}{n-1} \sum_{j=1}^n (X_{ij} - \bar{X}_i)(X_{kj} - \bar{X}_k), \quad i, k = 1, 2, \dots, p.$$

A amostra standardizada correspondente é dada por  $(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n)$ , com

$$\mathbf{Z}_j = \left[ \frac{X_{1j} - \bar{X}_1}{\sqrt{S_{11}}} \quad \frac{X_{2j} - \bar{X}_2}{\sqrt{S_{22}}} \quad \dots \quad \frac{X_{pj} - \bar{X}_p}{\sqrt{S_{pp}}} \right]^T, \quad j = 1, 2, \dots, n.$$

A matriz de variâncias-covariâncias amostral das variáveis standardizadas é a matriz de correlações amostral das variáveis não standardizadas, que denotamos por  $R$ . De facto, tem-se  $\bar{Z}_i = \frac{1}{n} \sum_{j=1}^n \frac{X_{ij} - \bar{X}_i}{\sqrt{S_{ii}}} = 0$ , resultando que o elemento na posição  $(i, k)$  da matriz de variâncias-covariâncias amostral das variáveis standardizadas é dado por

$$\frac{1}{n-1} \sum_{j=1}^n Z_{ij} Z_{kj} = \frac{1}{n-1} \sum_{j=1}^n \frac{(X_{ij} - \bar{X}_i)}{\sqrt{S_{ii}}} \frac{(X_{kj} - \bar{X}_k)}{\sqrt{S_{kk}}} = \frac{S_{ik}}{\sqrt{S_{ii}} \sqrt{S_{kk}}},$$

$i, k = 1, 2, \dots, p$ , que corresponde ao elemento na posição  $(i, k)$  da matriz  $R$ . Se  $n > p$ , as matrizes  $R$  e  $S$  são definidas positivas (teorema 3.14 de [6]). A restrição  $n > p$  que passaremos a admitir, é natural pois indica que a dimensão da amostra deve ser superior à dimensão do vetor aleatório  $\mathbf{X}$ .

Denotamos por  $l_1, l_2, \dots, l_p$  os valores próprios de  $R$ , todos positivos. Estes valores próprios são todos distintos com probabilidade 1 ([1], p. 280). Consideramos  $l_1 > l_2 > \dots > l_p$ . Sendo  $\mathbf{q}_i = [q_{1i} \ q_{2i} \ \dots \ q_{pi}]^T$  um vetor próprio normado de  $R$  associado a  $l_i$ ,  $i = 1, 2, \dots, p$ , a matriz  $Q = [\mathbf{q}_1 \ \mathbf{q}_2 \ \dots \ \mathbf{q}_p]$  é ortogonal e tem-se

$$R = QLQ^T = \sum_{i=1}^p l_i \mathbf{q}_i \mathbf{q}_i^T,$$

onde  $L = \text{diag}(l_1, l_2, \dots, l_p)$  é a matriz diagonal cujos elementos diagonais são os valores próprios de  $R$ . Na prática, a unicidade do vetor próprio  $\mathbf{q}_i$  fica garantida se, dos dois vetores próprios normados associados a  $l_i$ , for escolhido, por exemplo, aquele cuja primeira componente não nula tiver sinal positivo.

A partir da matriz  $R$  obtêm-se as componentes principais amostrais procedendo como no capítulo anterior. Desta forma, a  $j$ -ésima componente principal amostral é dada por,  $Y_j^* = \mathbf{q}_j^T \mathbf{Z}$ ,  $j = 1, 2, \dots, p$ . Assim o vetor das componentes principais amostrais é dado por

$$Y^* = Q^T \mathbf{Z}.$$

A variância amostral de  $Y_j^*$  é igual a  $l_j$  e a covariância amostral entre  $Y_j^*$  e  $Y_k^*$  é igual a zero,  $k \neq j$ ,  $j, k = 1, 2, \dots, p$ . Como os  $p$  elementos diagonais de  $R$  são todos iguais a 1, a variância amostral total das variáveis standardizadas é igual a  $p$  e  $l_1 + l_2 + \dots + l_p = p$ . Desta forma, a  $j$ -ésima componente principal amostral explica uma proporção da variância total amostral que é dada por  $\frac{l_j}{p}$ ,  $j = 1, 2, \dots, p$ .

No que diz respeito ao coeficiente de correlação amostral entre  $Y_j^*$  e  $Z_k$  tem-se  $r_{Y_j^*, Z_k} = q_{kj} \sqrt{l_j}$ ,  $j, k = 1, 2, \dots, p$ .

As componentes principais amostrais podem também ser determinadas a partir da matriz  $S$ , obtendo-se resultados análogos a partir dos valores e vetores próprios desta matriz.

Quanto ao coeficiente de correlação amostral entre  $Y_j^*$  e  $X_k$  tem-se agora  $r_{Y_j^*, X_k} = \frac{q_{kj} \sqrt{l_j}}{\sqrt{S_{kk}}}$ ,  $j, k = 1, 2, \dots, p$ .

Daqui em diante, na maior parte dos casos, as notações introduzidas neste capítulo serão usadas tanto para as estatísticas que representam como para as correspondentes concretizações.

### 2.2.2. Estimação de parâmetros

O vetor aleatório  $\bar{\mathbf{X}}$  e a matriz aleatória  $S$ , são estimadores de  $\boldsymbol{\mu}$  e  $\Sigma$ , respetivamente. Quanto aos valores próprios e vetores próprios de  $\Sigma$ , eles podem ser estimados pelos correspondentes valores próprios e vetores próprios de  $S$ .

Admitindo que  $\mathbf{X}$  segue uma lei normal multivariada,  $N_p(\boldsymbol{\mu}, \Sigma)$ , obtêm-se seguidamente os estimadores de máxima verosimilhança dos parâmetros acima mencionados.

Consideremos então um vetor aleatório  $\mathbf{X}$  com distribuição  $N_p(\boldsymbol{\mu}, \Sigma)$  e seja  $(\mathbf{X}_1, \dots, \mathbf{X}_n)$  uma amostra aleatória de dimensão  $n$  de  $\mathbf{X}$ . Os  $n$  vetores aleatórios  $\mathbf{X}_1, \dots, \mathbf{X}_n$  são então independentes e seguem todos a lei normal multivariada  $N_p(\boldsymbol{\mu}, \Sigma)$ . Nestas condições, e atendendo a que a matriz  $\Sigma$  é definida positiva,  $(\mathbf{X}_1, \dots, \mathbf{X}_n)$  tem função densidade conjunta dada por

$$\begin{aligned} f(\mathbf{x}_1, \dots, \mathbf{x}_n) &= \prod_{j=1}^n \frac{1}{(2\pi)^{\frac{p}{2}} (\det \Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x}_j - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}_j - \boldsymbol{\mu})\right) \\ &= \frac{1}{(2\pi)^{\frac{np}{2}}} \frac{1}{(\det \Sigma)^{\frac{n}{2}}} \exp\left(-\frac{1}{2} \sum_{j=1}^n (\mathbf{x}_j - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}_j - \boldsymbol{\mu})\right), \end{aligned} \quad (2.5)$$

para quaisquer  $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$ .

Começamos por obter os estimadores de máxima verosimilhança de  $\boldsymbol{\mu}$  e  $\Sigma$ .

Note-se que a expressão (2.5) define também a função de verosimilhança associada a cada realização  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$  da amostra aleatória de  $\mathbf{X}$ , a qual pretendemos maximizar em termos de  $\boldsymbol{\mu}$  e  $\Sigma$ . No entanto, como o factor  $(2\pi)^{-\frac{np}{2}}$  não afeta tal problema de maximização, os estimadores de máxima verosimilhança pretendidos

corresponderão às soluções  $\hat{\boldsymbol{\mu}}$  e  $\hat{\Sigma}$  que maximizam a função

$$L(\boldsymbol{\mu}, \Sigma) = (\det \Sigma)^{-\frac{n}{2}} \exp \left( -\frac{1}{2} \sum_{j=1}^n (\mathbf{x}_j - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_j - \boldsymbol{\mu}) \right). \quad (2.6)$$

Vamos escrever de outra forma o argumento da exponencial em (2.6). Denotando por  $tr$  o traço de uma matriz quadrada e atendendo a que o produto  $(\mathbf{x}_j - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_j - \boldsymbol{\mu})$  é um escalar, tem-se

$$\begin{aligned} (\mathbf{x}_j - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_j - \boldsymbol{\mu}) &= tr \left( (\mathbf{x}_j - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_j - \boldsymbol{\mu}) \right) \\ &= tr \left( \Sigma^{-1} (\mathbf{x}_j - \boldsymbol{\mu}) (\mathbf{x}_j - \boldsymbol{\mu})^T \right). \end{aligned}$$

Então

$$\begin{aligned} \sum_{j=1}^n (\mathbf{x}_j - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_j - \boldsymbol{\mu}) &= \sum_{j=1}^n tr \left( \Sigma^{-1} (\mathbf{x}_j - \boldsymbol{\mu}) (\mathbf{x}_j - \boldsymbol{\mu})^T \right) \\ &= tr \left( \Sigma^{-1} \sum_{j=1}^n (\mathbf{x}_j - \boldsymbol{\mu}) (\mathbf{x}_j - \boldsymbol{\mu})^T \right). \end{aligned} \quad (2.7)$$

Subtraindo e adicionando  $\bar{\mathbf{x}} = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j$  a cada termo  $(\mathbf{x}_j - \boldsymbol{\mu})$  em (2.7), obtém-se

$$\begin{aligned} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}} + \bar{\mathbf{x}} - \boldsymbol{\mu}) (\mathbf{x}_j - \bar{\mathbf{x}} + \bar{\mathbf{x}} - \boldsymbol{\mu})^T &= \\ &= \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}}) (\mathbf{x}_j - \bar{\mathbf{x}})^T + \sum_{j=1}^n (\bar{\mathbf{x}} - \mathbf{x}_j) (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \\ &= \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}}) (\mathbf{x}_j - \bar{\mathbf{x}})^T + n (\bar{\mathbf{x}} - \boldsymbol{\mu}) (\bar{\mathbf{x}} - \boldsymbol{\mu})^T. \end{aligned} \quad (2.8)$$

Esta última igualdade é obtida da anterior pelo facto de se ter

$$\sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}}) (\bar{\mathbf{x}} - \boldsymbol{\mu})^T = \sum_{j=1}^n (\bar{\mathbf{x}} - \boldsymbol{\mu}) (\mathbf{x}_j - \bar{\mathbf{x}})^T = 0.$$

Substituindo (2.8) em (2.7), tem-se

$$\begin{aligned} tr \left( \Sigma^{-1} \sum_{j=1}^n (\mathbf{x}_j - \boldsymbol{\mu}) (\mathbf{x}_j - \boldsymbol{\mu})^T \right) &= \\ &= tr \left( \Sigma^{-1} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}}) (\mathbf{x}_j - \bar{\mathbf{x}})^T \right) + n tr \left( \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \right) \\ &= tr \left( \Sigma^{-1} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}}) (\mathbf{x}_j - \bar{\mathbf{x}})^T \right) + n (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}). \end{aligned} \quad (2.9)$$

Consequentemente, tendo em conta (2.6), (2.7) e (2.9), podemos escrever a função  $L(\boldsymbol{\mu}, \Sigma)$  na forma

$$\frac{1}{(\det \Sigma)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2} \left[ tr \left( \Sigma^{-1} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}}) (\mathbf{x}_j - \bar{\mathbf{x}})^T \right) + n (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \right] \right\} \quad (2.10)$$

O lema que se segue é a base da demonstração da proposição que estabelece os estimadores de máxima verosimilhança de  $\boldsymbol{\mu}$  e  $\Sigma$ . A sua demonstração pode ser encontrada em [3] (pgs. 146-147).

**Lema 1.** *Dada uma matriz  $B$ , quadrada, de ordem  $p$ , simétrica e definida positiva e um escalar  $b > 0$ , tem-se*

$$\frac{1}{(\det \Sigma)^b} e^{-\text{tr}(\Sigma^{-1}B)/2} \leq \frac{1}{(\det B)^b} (2b)^{bp} e^{-bp} \quad (2.11)$$

para qualquer matriz  $\Sigma$ , quadrada, de ordem  $p$ , e definida positiva, verificando-se a igualdade apenas para  $\Sigma = \frac{1}{2b}B$ .

**Proposição 3.** *Seja  $(\mathbf{X}_1, \dots, \mathbf{X}_n)$  uma amostra aleatória de dimensão  $n$  de um vetor aleatório com distribuição  $N_p(\boldsymbol{\mu}, \Sigma)$ . Então*

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{j=1}^n \mathbf{X}_j \quad \text{e} \quad \hat{\Sigma} = \frac{1}{n} \sum_{j=1}^n (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^T$$

são os estimadores da máxima verosimilhança de  $\boldsymbol{\mu}$  e  $\Sigma$ , respetivamente.

*Demonstração.* Como já foi referido, pretendemos maximizar a função  $L(\boldsymbol{\mu}, \Sigma)$  dada por (2.10), para cada realização  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ , arbitrariamente fixa.

Como  $\Sigma$  é definida positiva, o mesmo acontece com  $\Sigma^{-1}$  pelo que  $(\bar{\mathbf{x}} - \boldsymbol{\mu})^T \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \geq 0$ , tendo-se a igualdade se e só se  $\boldsymbol{\mu} = \bar{\mathbf{x}}$ . Então, relativamente a  $\boldsymbol{\mu}$ , a função  $L(\boldsymbol{\mu}, \Sigma)$  atinge o seu valor máximo em  $\bar{\mathbf{x}}$ .

Resta então maximizar a função

$$L(\bar{\mathbf{x}}, \Sigma) = \frac{1}{(\det \Sigma)^{\frac{n}{2}}} \exp \left[ -\frac{1}{2} \text{tr} \left( \Sigma^{-1} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T \right) \right].$$

Aplicando o lema anterior com  $b = \frac{n}{2}$  e  $B = \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T$ , verifica-se que o máximo, relativamente a  $\Sigma$ , ocorre em  $\frac{1}{n} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T$ . Obtêm-

se assim, para  $\boldsymbol{\mu}$  e  $\Sigma$ , os estimadores de máxima verosimilhança  $\bar{\mathbf{X}} = \frac{1}{n} \sum_{j=1}^n \mathbf{X}_j$  e

$$\hat{\Sigma} = \frac{1}{n} \sum_{j=1}^n (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^T, \text{ respetivamente.} \quad \square$$

Além de  $\boldsymbol{\mu}$  e  $\Sigma$ , pretende-se estimar  $\lambda_i$  e  $\mathbf{e}_i$ ,  $i = 1, 2, \dots, p$ . O próximo teorema estabelece o resultado correspondente. Na sua demonstração são usados os dois lemas que se seguem. O primeiro estabelece a invariância dos estimadores de máxima verosimilhança quando são aplicadas funções bijetivas (cf. [1], pg. 48).

**Lema 2.** Se  $\hat{\theta}_1, \dots, \hat{\theta}_m$  são estimadores de máxima verosimilhança dos parâmetros reais  $\theta_1, \dots, \theta_m$  de uma determinada distribuição de probabilidade e se  $\varphi_1, \dots, \varphi_m$  são funções reais bijetivas, então  $\varphi_1(\hat{\theta}_1, \dots, \hat{\theta}_m), \dots, \varphi_m(\hat{\theta}_1, \dots, \hat{\theta}_m)$  são estimadores de máxima verosimilhança de  $\varphi_1(\theta_1, \dots, \theta_m), \dots, \varphi_m(\theta_1, \dots, \theta_m)$ . Além disso, se os estimadores  $\hat{\theta}_1, \dots, \hat{\theta}_m$  são únicos, o mesmo acontece com  $\varphi_1(\hat{\theta}_1, \dots, \hat{\theta}_m), \dots, \varphi_m(\hat{\theta}_1, \dots, \hat{\theta}_m)$ .

O segundo lema é um resultado que envolve matrizes sem componentes aleatórias (cf. [6], pg. 386).

**Lema 3.** Se  $U = \text{diag}(u_1, \dots, u_k)$ , com  $u_1 > \dots > u_k > 0$  e  $V = \text{diag}(v_1, \dots, v_p)$  com  $v_1 > \dots > v_p > 0$ ,  $k \leq p$ , então para toda a matriz  $P$  de tipo  $p \times k$  tal que  $P^T P = I$ , tem-se

$$\text{tr}(UP^TVP) \leq \sum_{i=1}^k u_i v_i.$$

A igualdade verifica-se apenas quando  $P$  assume a forma

$$P = \begin{bmatrix} \pm 1 & & 0 \\ & \ddots & \\ 0 & & \pm 1 \\ \hline & & \mathbf{0} \end{bmatrix} \quad (2.12)$$

**Teorema 2.** Suponhamos que a matriz  $\Sigma$  tem valores próprios  $\delta_1 > \dots > \delta_r$  com multiplicidades  $m_1, \dots, m_r$ , respetivamente,  $r \leq p$ . Particionem-se as matrizes  $\Gamma$  e  $Q$  na forma

$$\Gamma = [\Gamma_1 | \Gamma_2 | \dots | \Gamma_r] \quad e \quad Q = [Q_1 | Q_2 | \dots | Q_r],$$

onde  $\Gamma_i$  e  $Q_i$  são matrizes  $p \times m_i$ . Um estimador de máxima verosimilhança de  $\delta_i$  é

$$\hat{\delta}_i = \frac{1}{m_i} \frac{n-1}{n} \sum_{j \in D_i} l_j, \quad i = 1, 2, \dots, r,$$

onde  $D_1 = \{1, \dots, m_1\}$  e  $D_i = \{m_1 + \dots + m_{i-1} + 1, \dots, m_1 + \dots + m_i\}$ ,  $i = 2, \dots, r$ . Além disso, um estimador de máxima verosimilhança de  $\Gamma_i$  é dado por  $\hat{\Gamma}_i = Q_i P_{ii}$ , onde  $P_{ii}$  é uma matriz ortogonal qualquer de ordem  $m_i$  tal que o primeiro elemento em cada coluna de  $\hat{\Gamma}_i$  é não-negativo,  $i = 1, 2, \dots, r$ .

*Demonstração.* Na demonstração vamos considerar o caso em que apenas um valor próprio tem multiplicidade diferente de 1, sendo semelhante a prova do caso mais geral. Suponhamos então que os valores próprios de  $\Sigma$  são  $\lambda_1 > \dots > \lambda_k > \lambda_{k+1} =$

... =  $\lambda_p$ . Por simplificação de notação consideram-se os  $(p - k)$  menores valores próprios iguais a  $\lambda$ .

Já vimos, em (2.10), que a função de verosimilhança  $L(\boldsymbol{\mu}, \Sigma)$  associada a cada realização  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$  da amostra aleatória  $\mathbf{X}$ , pode ser escrita na forma

$$\frac{1}{(\det \Sigma)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2} \left[ \text{tr} \left( \Sigma^{-1} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T \right) + n(\bar{\mathbf{x}} - \boldsymbol{\mu})^T \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \right] \right\}.$$

Além disso, vimos que  $L(\boldsymbol{\mu}, \Sigma)$  é maximizada relativamente a  $\boldsymbol{\mu}$  para  $\boldsymbol{\mu} = \bar{\mathbf{x}}$ . Nesse caso, resta maximizar, relativamente a  $\lambda_1, \dots, \lambda_k, \lambda$ , a função

$$g(\Sigma) = \ln L(\bar{\mathbf{x}}, \Sigma) = -\frac{n}{2} \ln \det \Sigma - \frac{1}{2} \text{tr} \left( \Sigma^{-1} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T \right).$$

Podemos escrever  $\Sigma = \Gamma \Sigma_Y \Gamma^T$ , onde  $\Sigma_Y = \text{diag}(\lambda_1, \dots, \lambda_k, \lambda, \dots, \lambda)$  e  $\Gamma$  é uma matriz cujas colunas são os vetores próprios associados aos valores próprios que constituem a diagonal de  $\Sigma_Y$ . Analogamente, tem-se  $S = QLQ^T$ , com  $L = \text{diag}(l_1, \dots, l_p)$  e  $Q$  a matriz dos correspondentes vetores próprios. Tendo ainda em conta que  $\sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T = (n - 1)S$ , obtém-se

$$\begin{aligned} g(\Sigma) &= -\frac{n}{2} \log \det(\Gamma \Sigma_Y \Gamma^T) - \frac{1}{2} \text{tr} \left( (\Gamma \Sigma_Y \Gamma^T)^{-1} (n - 1) QLQ^T \right) \\ &= -\frac{n}{2} \log \left( \prod_{i=1}^p \lambda_i \right) - \frac{n - 1}{2} \text{tr} \left( \Gamma \Sigma_Y^{-1} \Gamma^T QLQ^T \right) \\ &= -\frac{n}{2} \log \left( \prod_{i=1}^k \lambda_i \right) - \frac{n}{2} \log \left( \prod_{j=k+1}^p \lambda_j \right) - \frac{n - 1}{2} \text{tr} \left( \Sigma_Y^{-1} P^T LP \right) \\ &= -\frac{n}{2} \sum_{i=1}^k \log \lambda_i - \frac{n(p - k)}{2} \log \lambda - \frac{n - 1}{2} \text{tr} \left( \Sigma_Y^{-1} P^T LP \right), \end{aligned}$$

onde  $P = Q^T \Gamma$  é uma matriz ortogonal de ordem  $p$ . Particionando  $P$  na forma  $P = [P_1 | P_2]$ , com  $P_1 \in \mathbb{R}^{p \times k}$  e  $P_2 \in \mathbb{R}^{p \times (p - k)}$  e escrevendo  $\Sigma_Y$  na forma

$$\Sigma_Y = \begin{bmatrix} \Sigma_{Y_1} & 0 \\ 0 & \lambda I_{p-k} \end{bmatrix}$$

onde  $\Sigma_{Y_1} = \text{diag}(\lambda_1, \dots, \lambda_k)$ , tem-se

$$\begin{aligned} \text{tr} \left( \Sigma_Y^{-1} P^T LP \right) &= \text{tr} \left( \begin{bmatrix} \Sigma_{Y_1}^{-1} & 0 \\ 0 & \lambda^{-1} I_{p-k} \end{bmatrix} \begin{bmatrix} P_1^T \\ P_2^T \end{bmatrix} L \begin{bmatrix} P_1^T & P_2^T \end{bmatrix} \right) \\ &= \text{tr} \left( \Sigma_{Y_1}^{-1} P_1^T L P_1 \right) + \lambda^{-1} \text{tr} \left( L P_2 P_2^T \right) \\ &= \frac{1}{\lambda} \sum_{i=1}^p l_i - \text{tr} \left( (\lambda^{-1} I_k - \Sigma_{Y_1}^{-1}) P_1^T L P_1 \right) \end{aligned}$$

No último passo usámos o facto de  $P_2 P_2^T = I - P_1 P_1^T$ . Desta forma,

$$g(\Sigma) = -\frac{n}{2} \sum_{i=1}^k \ln \lambda_i - \frac{n(p-k)}{2} \ln \lambda - \frac{n-1}{2\lambda} \sum_{i=1}^p l_i + \frac{n-1}{2} \text{tr} \left( (\lambda^{-1} I_k - \Sigma_{Y_1}^{-1}) P_1^T L P_1 \right) \quad (2.13)$$

Aplicando o lema anterior, com  $U = \lambda^{-1} I_k - \Sigma_{Y_1}^{-1}$  e  $V = L$ , o termo em (2.13) onde surge o traço é maximizado relativamente a  $P_1$  quando esta matriz tem a forma (2.12). Tal máximo é dado por

$$\frac{n-1}{2\lambda} \sum_{i=1}^k l_i - \frac{n-1}{2} \sum_{i=1}^k \frac{l_i}{\lambda_i}. \quad (2.14)$$

Atendendo a que  $P$  é ortogonal e  $P_1$  tem a forma indicada, a função  $g(\Sigma)$  é maximizada relativamente a  $P$  para

$$\widehat{P} = \left[ \begin{array}{cc|c} \pm 1 & 0 & \\ & \ddots & \mathbf{0} \\ 0 & \pm 1 & \\ \hline & \mathbf{0} & P_{22} \end{array} \right]$$

A matriz  $P_{22}$  é quadrada de ordem  $p - k$ . Desta forma, pelo lema 2,  $\widehat{\Gamma} = Q \widehat{P}$  é um estimador de máxima verosimilhança de  $\Gamma$ . Para concluir a demonstração, resta encontrar os estimadores de máxima verosimilhança de  $\lambda_i$  e  $\lambda$ . Para tal, por substituição de (2.14) em (2.13), maximiza-se relativamente a esses parâmetros a função

$$g(\Sigma) = -\frac{n}{2} \sum_{i=1}^k \ln \lambda_i - \frac{n(p-k)}{2} \ln \lambda - \frac{n-1}{2\lambda} \sum_{i=k+1}^p l_i - \frac{n-1}{2} \sum_{i=1}^k \frac{l_i}{\lambda_i},$$

obtendo-se para  $\lambda_i$  e  $\lambda$ , respetivamente, os estimadores de máxima verosimilhança

$$\widehat{\lambda}_i = \frac{n-1}{n} l_i, \quad i = 1, \dots, k$$

e

$$\widehat{\lambda} = \frac{n-1}{(p-k)n} \sum_{j=k+1}^p l_j.$$

□

**Exemplo 2.** No teorema anterior, suponhamos que  $p = 8$ ,  $r = 4$ ,  $m_1 = 1$ ,  $m_2 = 2$ ,  $m_3 = 3$  e  $m_4 = 2$ . Então  $D_1 = \{1\}$ ,  $D_2 = \{2, 3\}$ ,  $D_3 = \{4, 5, 6\}$  e  $D_4 = \{7, 8\}$ . De

acordo com o teorema, os estimadores de máxima verosimilhança são

$$\begin{aligned}\hat{\delta}_1 &= \frac{n-1}{n}l_1 \\ \hat{\delta}_2 &= \frac{n-1}{2n}(l_2 + l_3) \\ \hat{\delta}_3 &= \frac{n-1}{3n}(l_4 + l_5 + l_6) \\ \hat{\delta}_4 &= \frac{n-1}{2n}(l_7 + l_8)\end{aligned}$$

No que diz respeito à centricidade dos estimadores dos valores próprios de  $\Sigma$ , os resultados de Lawley ([4]), obtidos para o caso em que o vetor aleatório  $\mathbf{X}$  tem distribuição normal multivariada, permitem concluir que se  $\lambda_i$  é um valor próprio de  $\Sigma$  de multiplicidade 1, então o seu estimador  $l_i$  e, conseqüentemente, o seu estimador de máxima verosimilhança, são assintoticamente cêntricos.

Por outro lado, como se prova a seguir, os estimadores  $\bar{\mathbf{X}}$  e  $S$  são cêntricos de  $\boldsymbol{\mu}$  e  $\Sigma$ , respetivamente. Conseqüentemente, o estimador de máxima verosimilhança de  $\Sigma$  é apenas assintoticamente cêntrico.

**Proposição 4.** *Os estimadores  $\bar{\mathbf{X}}$  e  $S = \frac{n}{n-1}\hat{\Sigma}$  são cêntricos de  $\boldsymbol{\mu}$  e  $\Sigma$ , respetivamente.*

*Demonstração.* Tendo em conta a definição de esperança matemática de uma matriz aleatória real (matriz das esperanças matemáticas das variáveis aleatórias que formam a matriz aleatória, desde que existam), a linearidade da esperança matemática de variáveis aleatórias reais e ainda propriedades das operações com matrizes, tem-se

$$E(\bar{\mathbf{X}}) = \frac{1}{n} \sum_{j=1}^n E(\mathbf{X}_j) = \boldsymbol{\mu}$$

Para verificar a centricidade do estimador  $S$  de  $\Sigma$ , provemos primeiro que a matriz de variâncias-covariâncias de  $\bar{\mathbf{X}}$  é dada por

$$Cov(\bar{\mathbf{X}}) = \frac{1}{n}\Sigma.$$

Tem-se  $Cov(\bar{\mathbf{X}}) = E[(\bar{\mathbf{X}} - \boldsymbol{\mu})(\bar{\mathbf{X}} - \boldsymbol{\mu})^T]$  e, como a matriz  $(\bar{\mathbf{X}} - \boldsymbol{\mu})(\bar{\mathbf{X}} - \boldsymbol{\mu})^T$  tem elemento genérico  $(\bar{X}_i - \mu_i)(\bar{X}_j - \mu_j)$ ,  $i, j = 1, \dots, p$ , podemos escrever

$$\begin{aligned}Cov(\bar{\mathbf{X}}) &= E \left[ \left( \frac{1}{n} \sum_{j=1}^n (\mathbf{X}_j - \boldsymbol{\mu}) \right) \left( \frac{1}{n} \sum_{k=1}^n (\mathbf{X}_k - \boldsymbol{\mu}) \right)^T \right] \\ &= \frac{1}{n^2} E \left[ \sum_{j=1}^n \sum_{k=1}^n (\mathbf{X}_j - \boldsymbol{\mu})(\mathbf{X}_k - \boldsymbol{\mu})^T \right] \\ &= \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n E \left[ (\mathbf{X}_j - \boldsymbol{\mu})(\mathbf{X}_k - \boldsymbol{\mu})^T \right].\end{aligned}$$

## Capítulo 2 Componentes principais

---

Tendo em conta a independência dos vetores aleatórios  $\mathbf{X}_j$  e  $\mathbf{X}_k$ ,  $j \neq k$ , verifica-se  $E[(\mathbf{X}_j - \boldsymbol{\mu})(\mathbf{X}_k - \boldsymbol{\mu})^T] = 0$ ,  $j \neq k$ . Tem-se então

$$Cov(\bar{\mathbf{X}}) = \frac{1}{n^2} \sum_{j=1}^n E[(\mathbf{X}_j - \boldsymbol{\mu})(\mathbf{X}_j - \boldsymbol{\mu})^T] = \frac{1}{n} \Sigma,$$

uma vez que  $E[(\mathbf{X}_j - \boldsymbol{\mu})(\mathbf{X}_j - \boldsymbol{\mu})^T]$  corresponde à matriz de variâncias-covariâncias do vetor aleatório  $\mathbf{X}_j$  e este tem a mesma distribuição de  $\mathbf{X}$ ,  $j = 1, \dots, n$ .

Vamos então verificar que  $E(S) = \Sigma$ .

Tem-se

$$\begin{aligned} \sum_{j=1}^n (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^T &= \sum_{j=1}^n (\mathbf{X}_j - \bar{\mathbf{X}})\mathbf{X}_j^T + \left[ \sum_{j=1}^n (\mathbf{X}_j - \bar{\mathbf{X}}) \right] (-\bar{\mathbf{X}})^T \\ &= \sum_{j=1}^n \mathbf{X}_j \mathbf{X}_j^T - n\bar{\mathbf{X}}\bar{\mathbf{X}}^T, \end{aligned}$$

tendo em conta que  $\sum_{j=1}^n \mathbf{X}_j^T = n\bar{\mathbf{X}}^T$  e  $\sum_{j=1}^n (\mathbf{X}_j - \bar{\mathbf{X}}) = 0$ .

Então

$$\begin{aligned} E(S) &= \frac{1}{n-1} \sum_{j=1}^n E(\mathbf{X}_j \mathbf{X}_j^T) - \frac{n}{n-1} E(\bar{\mathbf{X}} \bar{\mathbf{X}}^T) \\ &= \frac{n}{n-1} [E(\mathbf{X} \mathbf{X}^T) - E(\bar{\mathbf{X}} \bar{\mathbf{X}}^T)], \end{aligned} \quad (2.15)$$

uma vez que os vetores aleatórios  $\mathbf{X}$  e  $\mathbf{X}_j$  seguem a mesma lei,  $j = 1, \dots, n$ .

Por outro lado, verifica-se facilmente que

$$\Sigma = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] \Leftrightarrow E(\mathbf{X} \mathbf{X}^T) = \Sigma + \boldsymbol{\mu} \boldsymbol{\mu}^T. \quad (2.16)$$

Analogamente, como  $E(\bar{\mathbf{X}}) = \boldsymbol{\mu}$  e  $Cov(\bar{\mathbf{X}}) = \frac{1}{n} \Sigma$ , tem-se

$$E(\bar{\mathbf{X}} \bar{\mathbf{X}}^T) = \frac{1}{n} \Sigma + \boldsymbol{\mu} \boldsymbol{\mu}^T. \quad (2.17)$$

O resultado  $E(S) = \Sigma$  obtém-se conjugando (2.15), (2.16) e (2.17).  $\square$

# Capítulo 3

## Análise fatorial

Em problemas que envolvam um grande número de variáveis observáveis são necessários muitos parâmetros para descrever toda a informação. Sempre que existam variáveis fortemente correlacionadas é concebível agrupá-las num grupo de modo que variáveis que estejam em grupos distintos apresentem fraca correlação. Esta é a base da análise fatorial, isto é, descrever a estrutura de covariâncias entre as variáveis iniciais através de um menor número de variáveis, os fatores ou variáveis latentes.

Este capítulo é dedicado ao modelo ortogonal de análise fatorial. Na secção 3.1, depois da definição do modelo, deduzem-se algumas propriedades decorrentes dos pressupostos envolvidos. Em particular, conclui-se que os pesos fatoriais com elevado valor absoluto identificam o fator a que cada variável se associa. Outra propriedade importante prende-se com o facto da matriz dos pesos fatoriais não ser única, podendo ser rotacionada de modo a produzir novos pesos fatoriais que facilitam a interpretação dos fatores. Na secção 3.2 apresentam-se dois procedimentos para, na prática, verificar a adequação dos dados à aplicação da análise fatorial, a qual só faz sentido se as variáveis iniciais forem correlacionadas. A secção 3.3 é dedicada à estimação dos pesos fatoriais, destacando-se o método das componentes principais. A redução da complexidade de um problema com muitas variáveis, passa pela escolha do número de fatores que podem substituir as variáveis originais sem perda significativa de informação. A escolha do número de fatores a considerar na análise fatorial é analisada na secção 3.4. Os fatores determinados podem ser vistos como um sistema de eixos no qual se representam os pontos cujas coordenadas são os pesos fatoriais correspondentes a cada uma das variáveis  $X_1, \dots, X_p$ . Tendo em conta o que foi referido sobre a rotação da matriz dos pesos fatoriais, tais eixos podem ser rotacionados com o objetivo de conseguir que os pesos fatoriais elevados (resp., baixos) fiquem ainda mais elevados (resp., baixos) de modo a facilitar a interpretação dos fatores. A rotação de fatores é estudada na secção 3.5. Apesar dos

fatores comuns serem variáveis não observáveis, é possível estimar o valor de cada fator (*score*) para cada indivíduo da amostra. Na secção 3.6 referem-se dois métodos para atingir esse objetivo: o método dos mínimos quadrados ponderados e o método da regressão.

### 3.1. Modelo ortogonal

Seja  $\mathbf{X} = [X_1 \ X_2 \ \dots \ X_p]^T$  um vetor aleatório real de dimensão  $p$ , com vetor média  $\boldsymbol{\mu} = [\mu_1 \ \mu_2 \ \dots \ \mu_p]^T$  e matriz de variâncias-covariâncias  $\Sigma$ , definida positiva. O modelo da análise fatorial expressa cada variável observável  $X_i$  como função linear de  $m$  variáveis aleatórias  $F_1, F_2, \dots, F_m$  ( $m < p$ ), chamadas fatores comuns, e de um fator único ou erro,  $\epsilon_i$ ,  $i = 1, 2, \dots, p$ . Este fator único é também uma variável aleatória cujo objetivo é explicar a parte da variância da respetiva variável não explicada pelos fatores comuns. Os  $m$  fatores comuns e os  $p$  fatores únicos não são observáveis. Mais precisamente, as variáveis observáveis são escritas na forma que representam a importância do  $j$ -ésimo factor comum na  $i$ -ésima variável observável,  $i = 1, 2, \dots, p$ ,  $j = 1, 2, \dots, m$ . O modelo pode escrever-se na forma matricial

$$\mathbf{X} - \boldsymbol{\mu} = \Lambda \mathbf{F} + \boldsymbol{\epsilon} \quad (3.1)$$

onde

$$\mathbf{F} = [F_1 \ F_2 \ \dots \ F_m]^T, \quad \boldsymbol{\epsilon} = [\epsilon_1 \ \epsilon_2 \ \dots \ \epsilon_p]^T \quad \text{e} \quad \Lambda = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1m} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2m} \\ \vdots & \vdots & & \vdots \\ \alpha_{p1} & \alpha_{p2} & \dots & \alpha_{pm} \end{bmatrix}$$

Assume-se que:

- $E(\mathbf{F}) = 0$ ;
- $E(\boldsymbol{\epsilon}) = 0$ ;
- $Cov(\mathbf{F}) = I$ ;

$$\bullet \quad Cov(\boldsymbol{\epsilon}) = \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \psi_p \end{bmatrix} = \Psi;$$

- Os vetores  $\mathbf{F}$  e  $\boldsymbol{\epsilon}$  são independentes.

Este conjunto de condições e a relação (3.1) constituem o chamado modelo ortogonal de análise fatorial. Se os fatores comuns forem considerados correlacionados, então estamos perante o modelo oblíquo, o qual não será tratado neste trabalho.

Nas condições do modelo ortogonal, verifica-se que

$$\Sigma = \Lambda\Lambda^T + \Psi. \quad (3.2)$$

De facto,

$$\begin{aligned} \Sigma &= E[(\Lambda\mathbf{F} + \boldsymbol{\epsilon})(\Lambda\mathbf{F} + \boldsymbol{\epsilon})^T] \\ &= E[\Lambda\mathbf{F}(\Lambda\mathbf{F})^T + \Lambda\mathbf{F}\boldsymbol{\epsilon}^T + \boldsymbol{\epsilon}(\Lambda\mathbf{F})^T + \boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] \\ &= \Lambda E(\mathbf{F}\mathbf{F}^T)\Lambda^T + \Lambda E(\mathbf{F}\boldsymbol{\epsilon}^T) + E(\boldsymbol{\epsilon}\mathbf{F}^T)\Lambda^T + E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T) \\ &= \Lambda\Lambda^T + \Psi, \end{aligned}$$

uma vez que  $E(\mathbf{F}\boldsymbol{\epsilon}^T) = Cov(\mathbf{F}, \boldsymbol{\epsilon}) = 0$ .

Tem-se também

$$Cov(\mathbf{X}, \mathbf{F}) = E((\Lambda\mathbf{F} + \boldsymbol{\epsilon})\mathbf{F}^T) = E(\Lambda\mathbf{F}\mathbf{F}^T) + E(\boldsymbol{\epsilon}\mathbf{F}^T) = \Lambda.$$

Note-se que a igualdade  $Cov(\mathbf{X}, \mathbf{F}) = \Lambda$  é equivalente a

$$Cov(X_i, F_j) = \alpha_{ij}, \quad i = 1, 2, \dots, p, \quad j = 1, 2, \dots, m.$$

Assim,  $\alpha_{ij}$  é uma medida da influência da variável  $X_i$  no fator  $F_j$ , pelo que, quando os pesos fatoriais são elevados em valor absoluto, identificam o fator a que cada variável se associa. Mais precisamente, as variáveis que contribuem para a determinação de um fator são aquelas cujos pesos fatoriais são mais elevados nesse fator.

De (3.2), resulta

$$V(X_i) = \alpha_{i1}^2 + \alpha_{i2}^2 + \dots + \alpha_{im}^2 + \psi_i, \quad i = 1, 2, \dots, p.$$

O valor

$$h_i^2 = \alpha_{i1}^2 + \alpha_{i2}^2 + \dots + \alpha_{im}^2$$

designa-se por comunalidade da variável  $X_i$  e corresponde à porção da variância de  $X_i$  que é explicada pelos fatores comuns. O valor  $\psi_i$  corresponde à porção desta variância que é explicada pelo fator específico e chama-se variância específica.

Em geral, a matriz  $\Lambda$  em (3.1) não é única. Efetivamente, se  $T$  for uma matriz ortogonal de ordem  $m$ , tem-se

$$\mathbf{X} = \Lambda\mathbf{F} + \boldsymbol{\epsilon} = \Lambda T T^T \mathbf{F} + \boldsymbol{\epsilon} = \Lambda^* \mathbf{F}^* + \boldsymbol{\epsilon},$$

com  $\Lambda^* = \Lambda T$  e  $\mathbf{F}^* = T^T \mathbf{F}$ . Note-se que

$$E(\mathbf{F}^*) = T^T E(\mathbf{F}) = 0 \quad \text{e} \quad Cov(\mathbf{F}^*) = T^T Cov(\mathbf{F}) T = I.$$

Assim, tanto  $\mathbf{F}$  como  $\mathbf{F}^*$  satisfazem as condições do modelo ortogonal. Além disso, ambas as matrizes  $\Lambda$  e  $\Lambda^*$  dão origem à matriz  $\Sigma$ . De facto,

$$\Sigma = \Lambda \Lambda^T + \Psi = \Lambda T T^T \Lambda^T + \Psi = \Lambda^* (\Lambda^*)^T + \Psi.$$

Observando ainda que as comunalidades,  $h_i^2$ ,  $i = 1, 2, \dots, p$ , correspondem aos elementos diagonais de  $\Lambda \Lambda^T$  e que  $\Lambda \Lambda^T = \Lambda^* (\Lambda^*)^T$ , conclui-se que os valores de  $h_i^2$  e, consequentemente os valores de  $V(X_i)$ , não são afetados pela escolha de  $T$ .

### 3.2. Adequação da aplicação do modelo fatorial

A aplicação da análise fatorial requer que as variáveis  $X_1, \dots, X_p$  sejam correlacionadas, pois caso contrário, tal aplicação não faz sentido. Se as variáveis não forem correlacionadas duas a duas, então a matriz de correlações é a matriz identidade. Faz então sentido testar a hipótese  $H_0 : C = I$  contra a alternativa  $H_1 : C \neq I$ . Para testar estas hipóteses, temos o teste de esfericidade de Bartlett. Sob a hipótese  $H_0$ , a distribuição da estatística deste teste converge em lei para uma variável aleatória com distribuição do quiquadrado com  $\frac{p(p-1)}{2}$  graus de liberdade, sob a hipótese da normalidade do vetor aleatório  $\mathbf{X} = [X_1 \dots X_p]^T$ . Tal estatística é dada por

$$\chi^2 = -\left(n - 2 - \frac{2p + 5}{6}\right) \ln |R|.$$

Ao nível de significância  $\alpha$ ,  $\alpha \in ]0, 1[$ , rejeita-se  $H_0$  se o valor observado da estatística de teste for superior ao quantil de ordem  $1 - \alpha$  daquela distribuição do quiquadrado. Este teste é muito sensível à violação da hipótese da normalidade de  $\mathbf{X}$  (cf. [5]).

O procedimento mais utilizado para avaliar o grau de adequação dos dados à aplicação da análise fatorial é a estatística de Kaiser-Meyer-Olkin, que é dada por

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

onde  $r_{ij}$  e  $v_{ij}$  são, respetivamente, os elementos na posição  $(i, j)$  da matriz de correlações amostral,  $R$ , e da matriz  $V = UR^{-1}U$ , na qual  $U = \left[(diag R^{-1})^{\frac{1}{2}}\right]^{-1}$ . Note-se que  $diag R^{-1}$  é a matriz cuja diagonal coincide com a diagonal de  $R^{-1}$  mas com os restantes elementos nulos e  $(diag R^{-1})^{\frac{1}{2}}$  é a matriz diagonal cujo  $i$ -ésimo elemento

diagonal é a raiz quadrada do  $i$ -ésimo elemento diagonal de  $\text{diag } R^{-1}$ . A matriz  $V$  é usualmente designada por matriz de correlação anti-imagem.

*Kaiser* propôs a seguinte relação entre o valor do KMO e o uso da análise fatorial

Valor do KMO	Recomendação AF
]0.9, 1]	Excelente
]0.8, 0.9]	Boa
]0.7, 0.8]	Média
]0.6, 0.7]	Aceitável
]0.5, 0.6]	Fraca
$\leq 0.5$	Inaceitável

### 3.3. Extração dos fatores

A expressão "extração dos fatores" é usada habitualmente na literatura para designar a metodologia de estimação da matriz de pesos fatoriais,  $\Lambda$ , e da matriz das variâncias específicas,  $\Psi$ .

Nesta secção apresentam-se dois dos métodos mais usados na extração dos fatores, nomeadamente o método das componentes principais e o método da máxima verosimilhança. No primeiro caso não são exigidas condições que envolvam distribuições de probabilidade, o que já não acontece no segundo caso, no qual se exige a normalidade do vetor aleatório  $\mathbf{X}$ . Neste trabalho, o método das componentes principais é estudado com mais detalhe.

#### 3.3.1. Método das componentes principais

A solução para o modelo fatorial apresentada pelo método das componentes principais, escolhe para os  $m$  fatores as primeiras  $m$  componentes principais,  $Y_1, \dots, Y_m$  (obtidas da matriz de variâncias-covariâncias,  $\Sigma$ , de  $\mathbf{X}$  ou da correspondente matriz de correlações,  $C$ ), divididas pelo respetivo desvio-padrão. Mais precisamente,  $F_j = \frac{Y_j}{\sqrt{\lambda_j}}$ ,  $j = 1, 2, \dots, m$ , uma vez que, recorde-se,  $V(Y_j) = \lambda_j$ , sendo  $\lambda_j$  o  $j$ -ésimo maior valor próprio de  $\Sigma$  (ou de  $C$ ). Desta forma, tem-se  $V(F_j) = 1$ . Recordemos ainda que, associado ao valor próprio  $\lambda_j$ , temos o vetor próprio normado  $\mathbf{e}_j = [e_{1j} \ e_{2j} \ \dots \ e_{pj}]^T$  e que  $Cov(X_i, Y_j) = \lambda_j e_{ij}$ . Além disso, como foi verificado na

secção 3.1, tem-se  $Cov(X_i, F_j) = \alpha_{ij}$ . Então

$$\begin{aligned}\alpha_{ij} &= Cov(X_i, F_j) = Cov\left(X_i, \frac{Y_j}{\sqrt{\lambda_j}}\right) = \frac{1}{\sqrt{\lambda_j}} Cov(X_i, Y_j) \\ &= \frac{1}{\sqrt{\lambda_j}} \lambda_j e_{ij} = \sqrt{\lambda_j} e_{ij}, \quad i = 1, 2, \dots, p, \quad j = 1, 2, \dots, m.\end{aligned}$$

Sabemos também que  $\Sigma = \Gamma \Sigma_Y \Gamma^T$  onde  $\Sigma_Y$  é uma matriz diagonal na qual constam os valores próprios de  $\Sigma$  por ordem decrescente. Nas colunas da matriz  $\Gamma$  estão os vetores próprios associados àqueles valores próprios. Assim, designando  $\Sigma_Y^{\frac{1}{2}} = diag(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_p})$ , tem-se

$$\Sigma = \Gamma \Sigma_Y \Gamma^T = \Gamma \Sigma_Y^{\frac{1}{2}} \Sigma_Y^{\frac{1}{2}} \Gamma^T = M M^T,$$

com  $M = \Gamma \Sigma_Y^{\frac{1}{2}} = [\sqrt{\lambda_1} \mathbf{e}_1 \dots \sqrt{\lambda_p} \mathbf{e}_p]$ .

A matriz  $\Lambda$  é então estimada a partir das primeiras  $m$  colunas da matriz  $M$ , ou seja,  $\Lambda$  é estimada por

$$\tilde{\Lambda} = [\sqrt{l_1} \mathbf{q}_1 \dots \sqrt{l_m} \mathbf{q}_m], \quad m < p, \quad (3.3)$$

onde  $l_j$  representa o  $j$ -ésimo valor próprio da matriz de variâncias-covariâncias amostral  $S$  (ou da matriz de correlações amostral,  $R$ ) e  $\mathbf{q}_j$  representa o  $j$ -ésimo vetor próprio normado associado a  $l_j$ ,  $j = 1, 2, \dots, m$ . Assim, os coeficientes  $\alpha_{ij}$  são estimados por

$$\tilde{\alpha}_{ij} = \sqrt{l_j} q_{ij}, \quad i = 1, 2, \dots, p, \quad j = 1, 2, \dots, m.$$

As communalidades e as variâncias específicas são então estimadas, respetivamente, por  $\tilde{h}_i^2 = \tilde{\alpha}_{i1}^2 + \tilde{\alpha}_{i2}^2 + \dots + \tilde{\alpha}_{im}^2$  e  $\tilde{\psi}_i = s_{ii} - \tilde{h}_i^2$ , onde  $s_{ii}$  representa o  $i$ -ésimo elemento diagonal da matriz  $S$ ,  $i = 1, 2, \dots, p$ . Note-se que, se for usada a matriz  $R$ , tem-se  $\tilde{\psi}_i = 1 - \tilde{h}_i^2$ .

Como já foi referido, as variáveis cujos pesos fatoriais são mais elevados num determinado fator são as que contribuem para a determinação desse fator. Se for usada a matriz  $R$  para a estimação dos pesos fatoriais, muitas vezes consideram-se significativos os pesos fatoriais superiores ou iguais a 0.5, por serem responsáveis por pelo menos 25% da variância total ([7], pg. 490).

### 3.3.2. Método da máxima verosimilhança

A aplicação deste método exige que a distribuição do vetor aleatório  $\mathbf{X}$  seja normal multivariada,  $N_p(\boldsymbol{\mu}, \Sigma)$ . A função de verosimilhança é então definida pela expressão

(2.10) considerada agora como função de  $\boldsymbol{\mu}$ ,  $\Lambda$  e  $\Psi$ , tendo em conta que  $\Sigma = \Lambda\Lambda^T + \Psi$ . Já vimos que tal função é maximizada relativamente a  $\boldsymbol{\mu}$  em  $\bar{\mathbf{x}}$ .

Pretende-se então maximizar

$$\frac{1}{(\det(\Lambda\Lambda^T + \Psi))^{\frac{n}{2}}} \exp \left[ -\frac{1}{2} \text{tr} \left( (\Lambda\Lambda^T + \Psi)^{-1} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T \right) \right]$$

relativamente a  $\Lambda$  e a  $\Psi$ . Devido à multiplicidade de escolhas para  $\Lambda$ , é imposta a condição  $\Lambda^{-1}\Psi^{-1}\Lambda = D$ , sendo  $D$  uma matriz diagonal. No entanto, esta maximização só é possível usando métodos numéricos. Existem programas eficientes já implementados que, para cada amostra concreta de  $\mathbf{X}$ , permitem obter as estimativas de máxima verosimilhança para  $\Lambda$  e  $\Psi$ , como é o caso do SPSS.

Uma informação mais completa sobre a estimação de  $\Lambda$  e  $\Psi$  através o método da máxima verosimilhança pode ser encontrada em [2] (pgs. 80-85).

### 3.4. Escolha do número de fatores

A escolha do número de fatores é um passo fundamental na análise fatorial. É desejável ter  $m < p$  fatores pois, caso contrário, embora se encontrem  $p$  variáveis não correlacionadas não se diminui o número de variáveis iniciais que é um dos objetivos da análise fatorial. De seguida apresentam-se vários critérios para efetuar tal escolha, todos eles baseados nos valores próprios  $\lambda_j$ ,  $j = 1, 2, \dots, p$ , da matriz de variâncias-covariâncias  $\Sigma$  (ou da matriz de correlações,  $C$ ).

1. **Critério da percentagem de variância total explicada.** Como a variância da  $j$ -ésima componente principal é igual ao  $j$ -ésimo valor próprio de  $\Sigma$ , a percentagem de variância total explicada pelas  $m$  primeiras componentes principais,  $m \leq p$ , é

$$\frac{\lambda_1 + \lambda_2 + \dots + \lambda_m}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \times 100\%.$$

O número de fatores que se consideram é igual ao número de valores próprios que é necessário considerar para perfazer uma determinada percentagem de variância total explicada, usualmente 85%. Note-se que, se for usada a matriz  $C$ , aquela percentagem é dada por  $\frac{\lambda_1 + \lambda_2 + \dots + \lambda_m}{p} \times 100\%$ .

2. **Critério de Kaiser.** Este critério, desenvolvido por Kaiser (1958), considera que o número de fatores deve ser igual ao número de valores próprios maiores ou iguais que a média aritmética dos  $p$  valores próprios. Observe-se que esta média é igual a 1 quando é usada a matriz de correlações.

3. *Scree plot*. Tem por base a interpretação de um gráfico no qual se representam os pontos  $(j, l_j)$ ,  $j = 1, 2, \dots, p$ . Na figura 3.1 apresenta-se um exemplo de um scree plot.

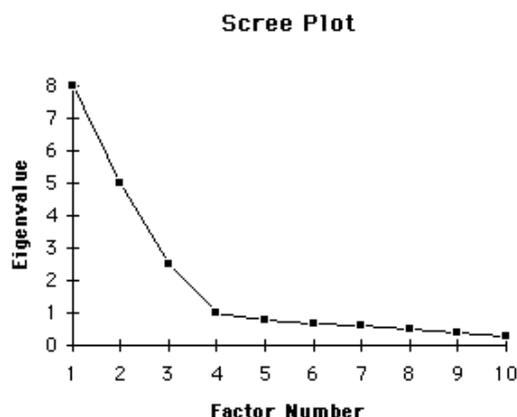


Figura 3.1: Exemplo de um *scree plot*.

Nestes gráficos temos uma linha poligonal que decresce rapidamente nos primeiros fatores. Estes assumem um papel de maior importância na análise fatorial no sentido em que explicam a maior parte da variância total. Por este critério, o número ótimo de fatores é obtido quando a variação da explicação entre fatores consecutivos passa a ser pequena. Por exemplo, no caso correspondente à figura 3.1 devem ser considerados 4 fatores.

No exemplo que se segue utilizamos o método das componentes principais para estimar os pesos fatoriais, usando o primeiro critério enunciado para escolher o número de fatores.

**Exemplo 3.** *Determinado curso tem três disciplinas obrigatórias, matemática ( $D_1$ ), inglês ( $D_2$ ) e português ( $D_3$ ). Seja  $X_i$  a classificação obtida por um aluno na disciplina  $D_i$ ,  $i = 1, 2, 3$ . Registaram-se os resultados obtidos por 5 alunos a essas 3 disciplinas (escala 0 – 20) no seguinte quadro:*

Número Estudante	1	2	3	4	5
$X_1$	6	14	19	7	18
$X_2$	12	8	17	16	13
$X_3$	10	8	18	15	11

Verifica-se facilmente que a matriz de correlação é dada por

$$R = \begin{bmatrix} 1 & 0.095 & 0.239 \\ 0.095 & 1 & 0.949 \\ 0.239 & 0.949 & 1 \end{bmatrix}$$

Para estimar a matriz de pesos fatoriais,  $\Lambda$ , vamos utilizar o método das componentes principais. Os valores próprios da matriz  $R$  são  $l_1 = 2.0032$ ,  $l_2 = 0.9576$  e  $l_3 = 0.0391$ . Os correspondentes vetores próprios normados são  $\mathbf{q}_1 = [0.2238 \ 0.6809 \ 0.6974]^T$ ,  $\mathbf{q}_2 = [0.9570 \ -0.2397 \ -0.0861]^T$  e  $\mathbf{q}_3 = [0.1160 \ 0.6921 \ -0.7124]^T$ .

Sendo  $\mathbf{Z} = [Z_1 \ Z_2 \ \dots \ Z_3]^T$  o vetor das variáveis standardizadas, as componentes principais são dadas por

$$Y_1^* = \mathbf{q}_1^T \mathbf{Z} = 0.2238Z_1 + 0.6809Z_2 + 0.6974Z_3$$

$$Y_2^* = \mathbf{q}_2^T \mathbf{Z} = 0.9570Z_1 - 0.2397Z_2 - 0.0861Z_3$$

$$Y_3^* = \mathbf{q}_3^T \mathbf{Z} = 0.1160Z_1 + 0.6921Z_2 - 0.7124Z_3.$$

As percentagens de variância total explicadas pela primeira, segunda e terceira componentes principais são respectivamente

$$\frac{l_1}{l_1 + l_2 + l_3} \times 100\% = 66.77\%$$

$$\frac{l_2}{l_1 + l_2 + l_3} \times 100\% = 31.92\%$$

$$\frac{l_3}{l_1 + l_2 + l_3} \times 100\% = 1.3\%.$$

Como a percentagem de variância total explicada pelas duas primeiras componentes principais é 98.69%, as variáveis originais podem ser substituídas por  $Y_1$  e  $Y_2$  sem perda significativa de informação.

De acordo com o critério da percentagem de variância total explicada, vamos ter dois fatores. O modelo de análise fatorial neste exemplo é dado por

$$Z_1 = \alpha_{11}F_1 + \alpha_{12}F_2 + \epsilon_1$$

$$Z_2 = \alpha_{21}F_1 + \alpha_{22}F_2 + \epsilon_2$$

$$Z_3 = \alpha_{31}F_1 + \alpha_{32}F_2 + \epsilon_3.$$

Atendendo a (3.3), a matriz  $\tilde{\Lambda}$  é

$$\tilde{\Lambda} = \begin{bmatrix} \sqrt{2.0032} \begin{bmatrix} 0.2238 \\ 0.6809 \\ 0.6974 \end{bmatrix} & \sqrt{0.9576} \begin{bmatrix} 0.9570 \\ -0.2397 \\ -0.0861 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 0.3168 & 0.9463 \\ 0.9637 & -0.2346 \\ 0.9871 & -0.0843 \end{bmatrix}.$$

Portanto, utilizando os pesos fatoriais estimados da forma apresentada, o modelo é estimado por

$$Z_1 = 0.3168F_1 + 0.9463F_2 + \epsilon_1$$

$$Z_2 = 0.9637F_1 - 0.2346F_2 + \epsilon_2$$

$$Z_3 = 0.9871F_1 - 0.0843F_2 + \epsilon_3.$$

Analisando a matriz  $\tilde{\Lambda}$  verifica-se que os pesos fatoriais de  $F_1$  em  $Z_2$  e  $Z_3$  (próximos de 1) se destacam do peso fatorial de  $F_1$  em  $Z_1$ , mais próximo de zero. Os pesos fatoriais de  $F_2$  em  $Z_2$  e  $Z_3$  são próximo de zero ao passo que o de  $Z_1$  é mais próximo de 1. Assim, como seria de esperar, as disciplinas de português e inglês formam um fator (disciplinas da área de letras) e a disciplina de matemática constitui um fator distinto.

Após empregar um dos critérios descritos anteriormente para obter um possível número  $m$  de fatores, é possível testar a adequação do modelo de análise fatorial se o vetor aleatório  $\mathbf{X}$  puder ser considerado normalmente distribuído. Testa-se então a hipótese  $H_0 : \Sigma = \Lambda\Lambda^T + \Psi$  contra a alternativa  $H_1 : \Sigma$  é uma outra matriz definida positiva. A correspondente estatística de teste é dada por

$$\left( n - \frac{2p + 4m + 11}{6} \right) \ln \left( \frac{|\hat{\Lambda}\hat{\Lambda}^T + \hat{\Psi}|}{|\hat{\Sigma}|} \right),$$

onde  $\hat{\Lambda}$  e  $\hat{\Psi}$  representam os estimadores de máxima verosimilhança de  $\Lambda$  e  $\Psi$ , respectivamente, e  $\hat{\Sigma} = \frac{n-1}{n}S$  (estimador de máxima verosimilhança de  $\Sigma$  apresentado no capítulo anterior). Esta estatística converge em lei para uma variável aleatória com distribuição do quiquadrado com  $\frac{1}{2}[(p-m)^2 - p - m]$  graus de liberdade. Assim, a aplicação prática do teste pode ser feita desde que a dimensão da amostra,  $n$ , e a diferença  $n - p$  sejam suficientemente grandes. Este teste aceita a hipótese nula a um nível de significância  $\alpha$ ,  $\alpha \in ]0, 1[$ , se o valor observado da estatística de teste for inferior ou igual ao quantil de ordem  $1 - \alpha$  da referida distribuição do quiquadrado. Finalmente, note-se que, como o parâmetro da lei do quiquadrado tem que ser positivo, o teste só poderá ser aplicado se  $m$  e  $p$  verificarem  $\frac{1}{2}[(p-m)^2 - p - m] > 0$ , o que é equivalente a  $m < \frac{1}{2}(2p + 1 - \sqrt{8p + 1})$  (cf. [3], pgs. 416-419).

### 3.5. Rotação dos fatores

A solução encontrada para o modelo de análise fatorial nem sempre é facilmente interpretável, ou seja, os pesos fatoriais das variáveis nos fatores comuns podem ter

valores que não permitem a determinação dos fatores.

Já foi referido que qualquer rotação ortogonal da matriz dos pesos fatoriais,  $\Lambda$ , gera a matriz de variâncias-covariâncias (ou a matriz de correlações, se for esta a matriz usada na sua determinação). Por outro lado, os  $m$  fatores podem ser interpretados como um sistema de eixos ortogonais dois a dois. Neste sistema de eixos, representemos cada um dos  $p$  pontos  $(\alpha_{1i}, \alpha_{i2}, \dots, \alpha_{im})$ ,  $i = 1, \dots, p$ . Com o objetivo de tornar os pesos fatoriais elevados ainda mais elevados e os pesos fatoriais baixos ainda mais baixos, facilitando a interpretação, podem ser efetuadas rotações dos fatores. Relativamente aos novos eixos fatoriais os  $p$  pontos acima referidos têm outras coordenadas, que correspondem aos pesos fatoriais depois da rotação.

Existem dois tipos de rotações, a ortogonal e a oblíqua. A rotação ortogonal é a opção mais utilizada, sendo assim designada por manter a ortogonalidade dos fatores. A rotação oblíqua não mantém a ortogonalidade dos fatores. A figura 3.2 ilustra o efeito da rotação dos fatores no caso  $m = 2$  e  $p = 10$ .

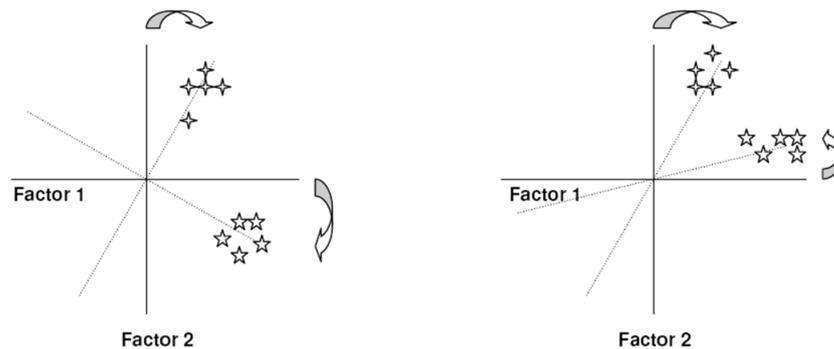


Figura 3.2: Exemplo do efeito da rotação (ortogonal e oblíqua) de dois fatores

Dentro das rotações ortogonais existem três métodos, nomeadamente Quartimax, Varimax e Equimax. Destes, o método Varimax é o mais utilizado e, por conseguinte, será abordado de seguida de forma mais aprofundada.

O método Varimax foi proposto por Kaiser em 1968. Sendo  $T$  uma matriz ortogonal,  $L = \Lambda T$  é uma matriz que representa a matriz dos pesos fatoriais rotacionada. Seja  $\theta$  o ângulo de rotação correspondente e  $l_{ij} = l_{ij}(\theta)$ ,  $i = 1, 2, \dots, p$  e  $j = 1, 2, \dots, m$ , o elemento genérico da matriz  $L$ . No método Varimax pretende-se encontrar o ângulo  $\theta$  que maximiza a função  $\phi(\theta)$  dada por

$$\phi(\theta) = \sum_{j=1}^m \sum_{i=1}^p (d_{ij}^2 - \bar{d}_j)^2 = \sum_{j=1}^m \sum_{i=1}^p d_{ij}^4 - p \sum_{j=1}^m \bar{d}_j^2, \quad (3.4)$$

onde

$$d_{ij} = \frac{l_{ij}}{h_i} \text{ e } \bar{d}_j = \frac{1}{p} \sum_{i=1}^p d_{ij}^2.$$

Deste modo,  $d_{ij}^2$  representa o quadrado do peso fatorial  $l_{ij}$  normalizado pela comunalidade da variável  $X_i$ ,  $h_i^2$ , e, para cada  $j = 1, \dots, m$ ,  $\bar{d}_j$  representa a média dos valores  $d_{ij}^2$ ,  $i = 1, \dots, p$ . Quando se toma o máximo de  $\phi$ , impõe-se que a diferença  $d_{ij}^2 - \bar{d}_j$ ,  $i = 1, 2, \dots, p$  e  $j = 1, 2, \dots, m$ , seja o maior possível, ou seja, conseguem-se pesos fatoriais extremos, obtendo-se assim variância máxima.

Vamos analisar com mais pormenor o caso  $m = 2$ . Para este caso, sabe-se que a matriz

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

é ortogonal e faz a rotação dos eixos no sentido horário de um ângulo  $\theta$ . Tem-se então  $l_{i1} = \alpha_{i1} \cos \theta - \alpha_{i2} \sin \theta$  e  $l_{i2} = \alpha_{i1} \sin \theta + \alpha_{i2} \cos \theta$ ,  $i = 1, 2, \dots, p$ . Assim,

$$d_{i1} = \frac{\alpha_{i1} \cos \theta - \alpha_{i2} \sin \theta}{h_i} \text{ e } d_{i2} = \frac{\alpha_{i1} \sin \theta + \alpha_{i2} \cos \theta}{h_i}, \quad i = 1, 2, \dots, p.$$

De (3.4), para  $m = 2$  tem-se

$$\begin{aligned} \phi(\theta) &= \sum_{j=1}^2 \sum_{i=1}^p d_{ij}^4 - p \sum_{j=1}^2 \bar{d}_j^2 \\ &= \sum_{i=1}^p \left[ \left( \frac{\alpha_{i1} \cos \theta - \alpha_{i2} \sin \theta}{h_i} \right)^4 + \left( \frac{\alpha_{i1} \sin \theta + \alpha_{i2} \cos \theta}{h_i} \right)^4 \right] - \\ &\quad - p \sum_{i=1}^p \left\{ \left[ \frac{1}{p} \left( \frac{\alpha_{i1} \cos \theta - \alpha_{i2} \sin \theta}{h_i} \right)^2 \right]^2 + \left[ \frac{1}{p} \left( \frac{\alpha_{i1} \sin \theta + \alpha_{i2} \cos \theta}{h_i} \right)^2 \right]^2 \right\} \\ &= \frac{1}{4} [\cos(4\theta)A + \sin(4\theta)B + C] \end{aligned} \quad (3.5)$$

onde, considerando  $G_{a,b} = \sum_{i=1}^p \frac{\alpha_{i1}^a \alpha_{i2}^b}{h_i^{a+b}}$ ,

$$\begin{aligned} A &= (G_{0,4} + G_{4,0} - 6G_{2,2} - G_{0,2}^2 - G_{2,0}^2 + 2G_{0,2}G_{2,0} + 4G_{1,1}^2) \\ B &= 4(G_{1,3} - G_{3,1} - G_{1,1}G_{0,2} + G_{1,1}G_{2,0}) \\ C &= p \left[ 3(G_{2,0} + G_{0,2})^2 - (3G_{0,2}^2 + 3G_{2,0}^2 + 2G_{0,2}G_{2,0} + 4G_{1,1}^2) \right] \end{aligned}$$

Uma vez que  $(A^2 + B^2)^{\frac{1}{2}} > 0$ , podemos multiplicar e dividir o segundo membro de (3.5) por esta constante, obtendo-se

$$\phi(\theta) = \frac{1}{4}(A^2 + B^2)^{\frac{1}{2}} \left[ \cos(4\theta) \frac{A}{(A^2 + B^2)^{\frac{1}{2}}} + \sin(4\theta) \frac{B}{(A^2 + B^2)^{\frac{1}{2}}} \right] + \frac{1}{4}C. \quad (3.6)$$

Como  $-1 \leq \frac{A}{(A^2 + B^2)^{\frac{1}{2}}} \leq 1$  e  $-1 \leq \frac{B}{(A^2 + B^2)^{\frac{1}{2}}} \leq 1$ , existe um  $\alpha$  tal que  $\cos \alpha = \frac{A}{(A^2 + B^2)^{\frac{1}{2}}}$  e  $\sin \alpha = \frac{B}{(A^2 + B^2)^{\frac{1}{2}}}$ . Substituindo em (3.6) obtém-se

$$\begin{aligned} \phi(\theta) &= \frac{1}{4}(A^2 + B^2)^{\frac{1}{2}} [\cos(4\theta) \cos \alpha + \sin(4\theta) \sin \alpha] + \frac{1}{4}C \\ &= \frac{1}{4}(A^2 + B^2)^{\frac{1}{2}} \cos(4\theta - \alpha) + \frac{1}{4}C \end{aligned}$$

O máximo da função  $\phi(\theta)$  é atingido em  $\theta = \alpha/4$ , sendo o valor de  $\alpha$  obtido de

$$\tan \alpha = \frac{\sin \alpha}{\cos \alpha} = \frac{B}{A}.$$

Para encontrar o quadrante de uma possível rotação basta analisar os sinais do seno e cosseno na equação anterior.

### 3.6. Scores fatoriais

Apesar de um fator ser uma variável não observável obtida a partir de um conjunto de variáveis observáveis, um indivíduo da amostra pode possuir um *score* em cada um dos  $m$  fatores obtidos. Um score,  $f_{ij}$ , pode então ser interpretado como uma estimativa do valor do fator  $F_i$  para o  $j$ -ésimo indivíduo da amostra,  $i = 1, 2, \dots, m$ ,  $j = 1, 2, \dots, n$ . No entanto, é de salientar que os scores não são estimativas de parâmetros no sentido usual, mas sim estimativas do valor não observado de cada um dos fatores (que são variáveis aleatórias), para cada indivíduo da amostra. Além disso, o problema de estimação complica-se pelo facto de, relativamente a cada indivíduo  $j$ , as quantidades não observadas,  $m$  valores  $f_{ij}$  e  $p$  valores dos correspondentes fatores específicos, superarem os  $p$  valores observados. Por este motivo, na obtenção dos scores há a necessidade de considerar estimativas de  $\Lambda$  e  $\Psi$  como os verdadeiros valores destas matrizes.

Os scores fatoriais são frequentemente usados para propósitos de diagnósticos e também para efetuar uma análise posterior.

Neste texto serão abordados dois métodos para determinar os scores fatoriais: o método dos mínimos quadrados ponderados e o método de regressão.

#### 3.6.1. Método dos mínimos quadrados ponderados

Como vimos em (3.1), o modelo fatorial pode ser escrito na forma matricial

$$\mathbf{X} - \boldsymbol{\mu} = \Lambda \mathbf{F} + \boldsymbol{\epsilon}.$$

Considerem-se os fatores específicos  $\boldsymbol{\epsilon}^T = [\epsilon_1 \ \epsilon_2 \ \dots \ \epsilon_p]$  como erros. Como as variâncias específicas  $V(\epsilon_i) = \psi_i$ ,  $i = 1, 2, \dots, p$ , não são necessariamente iguais, tomam-se os erros normalizados,  $\frac{\epsilon_i}{\sqrt{\psi_i}}$ ,  $i = 1, \dots, p$ , e minimiza-se a soma dos quadrados dos erros

$$\sum_{i=1}^p \frac{\epsilon_i^2}{\psi_i} = \boldsymbol{\epsilon}^T \boldsymbol{\Psi}^{-1} \boldsymbol{\epsilon} = (\mathbf{x} - \boldsymbol{\mu} - \Lambda \mathbf{f})^T \boldsymbol{\Psi}^{-1} (\mathbf{x} - \boldsymbol{\mu} - \Lambda \mathbf{f}), \quad (3.7)$$

relativamente ao vetor  $\mathbf{f}$ , de dimensão  $m$ . Admitindo que  $\boldsymbol{\mu}$ ,  $\Lambda$  e  $\boldsymbol{\Psi}$  são conhecidos, a solução deste problema de minimização é o vetor  $\mathbf{f}$  que verifica

$$\mathbf{f} = \left( \Lambda^T \boldsymbol{\Psi}^{-1} \Lambda \right)^{-1} \Lambda^T \boldsymbol{\Psi}^{-1} (\mathbf{x} - \boldsymbol{\mu}).$$

(cf. [3], pg. 430). Consideram-se agora como verdadeiros valores para  $\Lambda$  e  $\boldsymbol{\Psi}$  as estimativas obtidas para estas matrizes por um dos métodos referidos, as quais representamos por  $\Lambda^*$  e  $\boldsymbol{\Psi}^*$ , respetivamente. Além disso considera-se como verdadeiro valor de  $\boldsymbol{\mu}$  o vetor média da amostra,  $\bar{\mathbf{x}}$ . Desta forma os scores fatoriais para o  $j$ -ésimo indivíduo da amostra são as  $m$  componentes do vetor

$$\mathbf{f}_j^* = \left( \Lambda^{*T} (\boldsymbol{\Psi}^*)^{-1} \Lambda^* \right)^{-1} \Lambda^{*T} (\boldsymbol{\Psi}^*)^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}), \quad (3.8)$$

onde  $\mathbf{x}_j$  representa o vetor das  $p$  observações correspondentes ao  $j$ -ésimo indivíduo,  $j = 1, 2, \dots, n$ .

### 3.6.2. Método da regressão

Este método requer a normalidade conjunta de  $\mathbf{F}$  e de  $\boldsymbol{\epsilon}$ . Tendo em conta as hipóteses consideradas para modelo ortogonal,  $\mathbf{X} - \boldsymbol{\mu} = \Lambda \mathbf{F} + \boldsymbol{\epsilon}$ , apresentado na secção 3.1, tem-se  $\mathbf{X} - \boldsymbol{\mu} \sim N_p(0, \Lambda \Lambda^T + \boldsymbol{\Psi})$ .

Prova-se também que a distribuição conjunta de  $\mathbf{X}$  e  $\mathbf{F}$  é  $N_{m+p}(0, \Sigma')$  onde

$$\Sigma' = \left[ \begin{array}{c|c} \Lambda \Lambda^T + \boldsymbol{\Psi} & \Lambda \\ \hline \Lambda^T & I \end{array} \right]$$

Além disso, verifica-se que a distribuição condicional de  $\mathbf{F}$  dado  $\mathbf{X} = \mathbf{x}$ ,  $\mathbf{F}|\mathbf{x}$ , é normal multivariada com vetor média e matriz de variâncias-covariâncias dados, respetivamente, por

$$m_{\mathbf{F}|\mathbf{x}} = \Lambda^T (\Lambda \Lambda^T + \boldsymbol{\Psi})^{-1} (\mathbf{x} - \boldsymbol{\mu}) \text{ e } Cov_{\mathbf{F}|\mathbf{x}} = I - \Lambda^T (\Lambda \Lambda^T + \boldsymbol{\Psi})^{-1} \Lambda.$$

De acordo com este método os scores fatoriais são obtidos encontrando estimativas dos coeficientes na regressão multivariada dos fatores nas variáveis, ou seja,

estimativas de  $\Lambda^T(\Lambda\Lambda^T + \Psi)^{-1}$ . Desta forma, tomando como verdadeiras as estimativas de  $\Lambda$  e  $\Psi$  obtidas pelo método da máxima verosimilhança, os scores fatoriais para o  $j$ -ésimo indivíduo da amostra são as  $m$  componentes do vetor

$$\hat{\mathbf{f}}_j = \hat{\Lambda}^T \left( \hat{\Lambda}\hat{\Lambda}^T + \hat{\Psi} \right)^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}), \quad j = 1, 2, \dots, n. \quad (3.9)$$

Para uma informação mais completa sobre a obtenção dos scores fatoriais pode consultar-se [3], pgs. 429-434.

### 3.6.3. Relação entre as estimativas obtidas pelos dois métodos

É possível relacionar os scores factoriais obtidos pelo método dos mínimos quadrados ponderados e pelo método da regressão, quando, em (3.8), usamos  $\Gamma^* = \hat{\Gamma}$  e  $\Psi^* = \hat{\Psi}$ . De facto, nestas condições, denotando  $\mathbf{f}_j^{MQ}$  e  $\mathbf{f}_j^R$  os scores obtidos para o  $j$ -ésimo indivíduo da amostra,  $j = 1, 2, \dots, n$ , pelo método dos mínimos quadrados e pelo método da regressão, respetivamente, tem-se

$$\mathbf{f}_j^R = (I + \hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda} \mathbf{f}_j^{MQ} \quad (3.10)$$

Para verificar a validade desta relação necessitamos do resultado do lema 6, em cuja prova se usam os lemas 4 e 5. Apresentam-se a seguir estes três lemas e respetivas provas.

**Lema 4.**  $(I + \hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda} = I - (I + \hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda})^{-1}$

*Demonstração.* Multiplicando à esquerda ambos os membros da igualdade trivial  $\hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda} = (I + \hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda}) - I$  por  $(I + \hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda})^{-1}$ , obtém-se

$$(I + \hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda} = (I + \hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda})^{-1} [(I + \hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda}) - I],$$

que é equivalente ao que se pretende provar.  $\square$

**Lema 5.**  $(\hat{\Lambda}\hat{\Lambda}^T + \hat{\Psi})^{-1} = \hat{\Psi}^{-1} - \hat{\Psi}^{-1} \hat{\Lambda} (I + \hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}^T \hat{\Psi}^{-1}$

*Demonstração.* Vamos verificar que a igualdade enunciada é equivalente a uma igualdade trivial, o que prova a sua veracidade. Em primeiro lugar, multiplicando aquela igualdade, à direita, por  $\hat{\Lambda}\hat{\Lambda}^T + \hat{\Psi}$ , resulta a igualdade equivalente

$$I = \hat{\Psi}^{-1} \hat{\Lambda}\hat{\Lambda}^T + I - \hat{\Psi}^{-1} \hat{\Lambda} (I + \hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}^T \hat{\Psi}^{-1} (\hat{\Lambda}\hat{\Lambda}^T + \hat{\Psi}),$$

que, por seu lado, é equivalente a

$$0 = \hat{\Psi}^{-1} \hat{\Lambda}\hat{\Lambda}^T - \hat{\Psi}^{-1} \hat{\Lambda} \left[ (I + \hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda}\hat{\Lambda}^T + (I + \hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}^T \right].$$

Usando agora o lema 4, esta igualdade pode ser escrita na forma

$$0 = \widehat{\Psi}^{-1}\widehat{\Lambda}\widehat{\Lambda}^T - \widehat{\Psi}^{-1}\widehat{\Lambda} \left\{ \left[ I - \left( I + \widehat{\Lambda}^T\widehat{\Psi}^{-1}\widehat{\Lambda} \right)^{-1} \right] \widehat{\Lambda}^T + \left( I + \widehat{\Lambda}^T\widehat{\Psi}^{-1}\widehat{\Lambda} \right)^{-1} \widehat{\Lambda}^T \right\},$$

que, como facilmente se verifica, é uma igualdade verdadeira.  $\square$

**Lema 6.**  $\widehat{\Lambda}^T(\widehat{\Lambda}\widehat{\Lambda}^T + \widehat{\Psi})^{-1} = (I + \widehat{\Lambda}^T\widehat{\Psi}^{-1}\widehat{\Lambda})^{-1}\widehat{\Lambda}^T\widehat{\Psi}^{-1}$

*Demonstração.* Começamos por multiplicar ambos os membros da igualdade presente no lema 5, à esquerda, por  $\widehat{\Lambda}^T$  e, seguidamente, pomos  $\widehat{\Lambda}^T\widehat{\Psi}^{-1}$  em evidência, à direita. Resulta então

$$\widehat{\Lambda}^T(\widehat{\Lambda}\widehat{\Lambda}^T + \widehat{\Psi})^{-1} = \left[ I - \widehat{\Lambda}^T\widehat{\Psi}^{-1}\widehat{\Lambda} \left( I + \widehat{\Lambda}^T\widehat{\Psi}^{-1}\widehat{\Lambda} \right)^{-1} \right] \widehat{\Lambda}^T\widehat{\Psi}^{-1}.$$

Como facilmente se verifica, a matriz  $I + \widehat{\Lambda}^T\widehat{\Psi}^{-1}\widehat{\Lambda}$  é simétrica, pelo que, transpondo ambos os membros da igualdade do lema 4, obtemos

$$\widehat{\Lambda}^T\widehat{\Psi}^{-1}\widehat{\Lambda} \left( I + \widehat{\Lambda}^T\widehat{\Psi}^{-1}\widehat{\Lambda} \right)^{-1} = I - \left( I + \widehat{\Lambda}^T\widehat{\Psi}^{-1}\widehat{\Lambda} \right)^{-1}.$$

Usando este resultado na igualdade anterior, obtém-se de imediato o resultado pretendido.  $\square$

Vamos então verificar a igualdade (3.10). Usando o lema 6 e atendendo a (3.9), tem-se, para cada  $j = 1, 2, \dots, n$ ,

$$\begin{aligned} \mathbf{f}_j^R &= \widehat{\Lambda}^T \left( \widehat{\Lambda}\widehat{\Lambda}^T + \widehat{\Psi} \right)^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}) \\ &= \left( I + \widehat{\Lambda}^T\widehat{\Psi}^{-1}\widehat{\Lambda} \right)^{-1} \widehat{\Lambda}^T\widehat{\Psi}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}). \end{aligned}$$

Por outro lado, de (3.8), tem-se

$$\widehat{\Lambda}^T\widehat{\Psi}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}) = \widehat{\Lambda}^T\widehat{\Psi}^{-1}\widehat{\Lambda} \mathbf{f}_j^{MQ}.$$

Conjugando esta igualdade com a anterior, obtém-se a relação (3.10).

## Capítulo 4

# Exemplo de aplicação da Análise Fatorial

Neste capítulo, apresentamos um exemplo de aplicação da análise fatorial. Vamos usar o ficheiro *World.95*, do SPSS, que contém informação sobre 108 países dos cinco continentes em 26 variáveis. Neste exemplo apenas consideramos variáveis quantitativas, pelo que excluímos do estudo as quatro variáveis qualitativas que surgem no ficheiro mencionado. Exluímos também outras três variáveis que são logaritmos de variáveis consideradas e, portanto, redundantes. Os *missings* são tratados pelo procedimento *listwise*, no qual apenas se consideram os registos que têm informação sobre todas as variáveis, opção standard do *software* estatístico utilizado, o SPSS. Desta forma, são incluídos na análise apenas 59 países. As variáveis consideradas no estudo são: *Percentagem de pessoas que habitam nas cidades*, *Longevidade da população do sexo feminino*, *Longevidade da população do sexo masculino*, *Percentagem da população literada*, *Percentagem da população do sexo feminino literada*, *Percentagem da população do sexo masculino literada*, *Taxa de mortalidade infantil*, *Taxa de mortalidade*, *Taxa de natalidade*, *Taxa de fecundidade*, *Média de calorias diárias ingeridas por pessoa*, *Percentagem de crescimento populacional por ano*, *Quociente entre taxas de natalidade e mortalidade*, *Casos sida no total da população*, *Taxa de sida*, *Produto interno bruto per capita*, *Percentagem de crescimento da produção agrícola*, *População em milhares* e *Densidade populacional*. Relativamente ao ficheiro original, todas as variáveis que são taxas são uniformizadas para mil habitantes.

O nosso objetivo é determinar os fatores subjacentes e dar alguma interpretação aos mesmos. Veremos que há variáveis que não podem ser consideradas normais, pelo que não podemos admitir a normalidade do vetor aleatório formado pelas variáveis acima listadas. Com tal, usaremos apenas o método das componentes principais para estimar os pesos fatoriais.

### 4.1. Análise preliminar dos dados

A normalidade de cada uma das variáveis que entra no estudo não garante a normalidade do correspondente vetor aleatório,  $\mathbf{X}$ . No entanto, se uma das variáveis não for normal, o vetor não é normal. Como queremos testar a normalidade de cada variável isoladamente, utilizamos o procedimento *pairwise* para tratar os *missings*. De acordo com este procedimento usamos todos os dados disponíveis para cada variável. Tendo em conta que a dimensão das amostras em estudo varia entre 75 e 108, o teste mais indicado para testar a normalidade é o teste de Kolmogorov-Smirnov com correção de Lilliefors. Os resultados correspondentes encontram-se na figura 4.1. Verificamos que, ao nível de significância 0.05, apenas uma delas pode ser considerada normal e, mesmo esse caso tem associado um p-valor muito baixo (0.052).

**Tests of Normality**

	Kolmogorov-Smirnov <sup>a</sup>		
	Statistic	df	Sig.
Percentagem de pessoas que habitam nas cidades	,086	107	,052
Longevidade da população do sexo feminino	,172	108	,000
Longevidade da população do sexo masculino	,164	108	,000
Percentagem da população literada	,171	106	,000
Percentagem da população do sexo feminino literada	,144	85	,000
Percentagem da população do sexo masculino literada	,169	85	,000
Taxa de mortalidade infantil	,168	108	,000
Taxa de mortalidade	,136	108	,000
Taxa de natalidade	,161	108	,000
Taxa de fecundidade	,161	107	,000
Média de calorias diárias ingeridas por pessoa	,143	75	,001
Percentagem de crescimento populacional por ano	,101	108	,009
Quociente entre taxas de natalidade e mortalidade	,142	108	,000
Casos sida no total da população	,423	106	,000
Taxa de sida	,311	106	,000
Produto interno bruto per capita	,208	108	,000
Percentagem de crescimento da produção agrícola	,128	106	,000
População em milhares	,373	108	,000
Densidade populacional	,385	108	,000

a. Lilliefors Significance Correction

Figura 4.1: Teste de Kolmogorov-Smirnov.

Podemos assim concluir que o vetor  $\mathbf{X}$  não pode ser considerado normal.

Como vimos no capítulo anterior, não faz sentido aplicar a análise fatorial se as correlações entre as variáveis forem fracas. Sendo violado o pressuposto de normalidade do vetor  $\mathbf{X}$ , não devemos usar o teste de esfericidade de Bartlett, pelo que recorreremos ao valor da estatística KMO para avaliar a adequação dos dados à aplicação da análise fatorial. Como se pode observar na figura 4.2, tal valor é 0.821, pelo que faz sentido aplicar a análise fatorial.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,821
Bartlett's Test of Sphericity	Approx. Chi-Square	1566,726
	df	171
	Sig.	,000

Figura 4.2: Valor KMO e teste de esfericidade de Bartlett.

## 4.2. Aplicação da análise fatorial

Vamos utilizar apenas o método das componentes principais na estimação da matriz dos pesos fatoriais, uma vez que, como vimos, o vetor  $\mathbf{X}$  não deve ser considerado normal, logo não faz sentido aplicar o método da máxima verosimilhança. Usamos para tal a matriz de correlações da amostra,  $R$ . Começamos por determinar os seus valores próprios, lembrando que a variância da  $j$ -ésima componente principal amostral coincide com o  $j$ -ésimo maior deles, sendo assim possível calcular a percentagem de variância total explicada por cada uma das componentes principais amostrais. Na tabela 4.3 apresentam-se os resultados referidos acima.

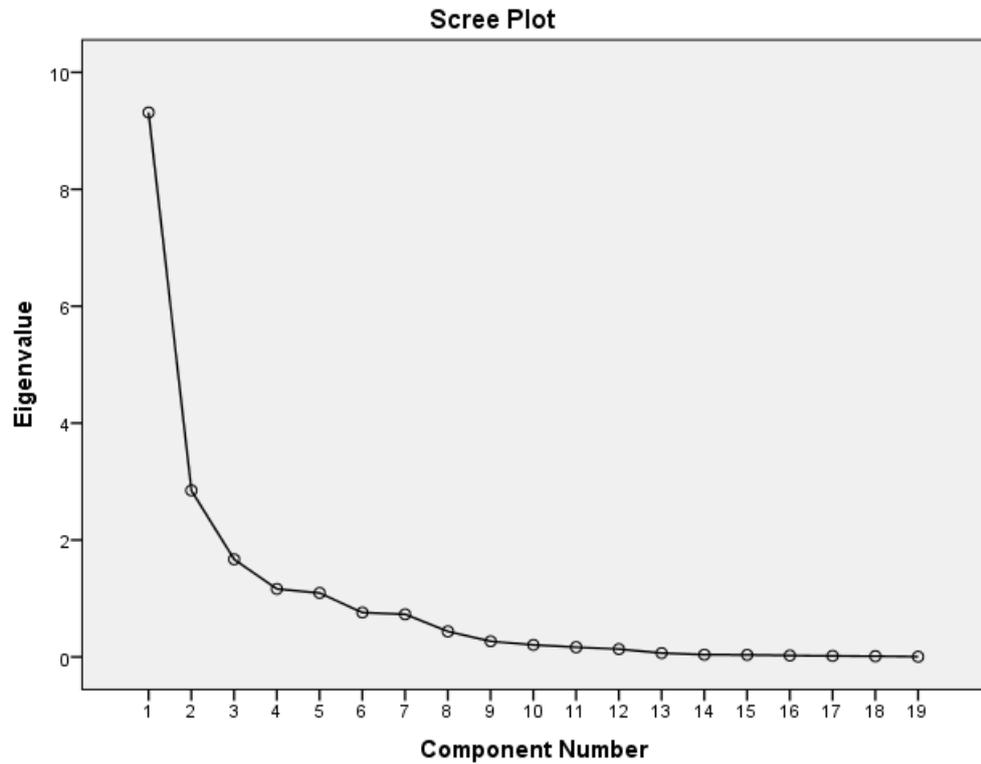
**Total Variance Explained**

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	9,314	49,019	49,019
2	2,848	14,990	64,009
3	1,671	8,793	72,802
4	1,164	6,126	78,928
5	1,095	5,765	84,693
6	,760	4,000	88,693
7	,730	3,843	92,536
8	,438	2,305	94,840
9	,269	1,413	96,254
10	,206	1,084	97,338
11	,167	,879	98,218
12	,135	,709	98,926
13	,067	,353	99,279
14	,040	,212	99,491
15	,036	,191	99,682
16	,025	,133	99,815
17	,018	,094	99,909
18	,013	,069	99,978
19	,004	,022	100,000

Extraction Method: Principal Component Analysis.

Figura 4.3: Valores próprios e percentagens de variância explicada das 19 componentes principais amostrais.

Analisando a tabela 4.3, por exemplo o valor  $l_1 = 9.314$  representa a variância amostral da primeira componente principal. Dividindo pela soma de todas as variâncias, obtemos a percentagem de variância que a primeira componente explica em relação ao total, aproximadamente 49%. De acordo com o critério de Kaiser retêm-se 5 componentes principais, tantas quantos os valores próprios maiores que um. Para confirmar o número de fatores a reter pode utilizar-se o *scree plot*, gráfico que se apresenta na figura seguinte.

Figura 4.4: *Scree plot*.

Verifica-se que a linha poligonal decresce rapidamente nos primeiros 5 fatores, os quais explicam a maior parte da variância total, aproximadamente 85%.

O passo seguinte é determinar a matriz de pesos fatoriais que, no SPSS, é dada na tabela *component matrix*, de forma a construir e interpretar os fatores. No quadro que se segue apresentamos os pesos fatoriais estimados pelo método das componentes principais.

**Component Matrix**

	Component				
	1	2	3	4	5
Percentagem de pessoas que habitam nas cidades	,754	-,262	,328	,141	-,122
Longevidade da população do sexo feminino	,956	-,219	-,028	,031	,051
Longevidade da população do sexo masculino	,930	-,265	-,048	,064	,115
Percentagem da população literada	,912	,079	,012	-,293	-,126
Percentagem da população do sexo feminino literada	,907	,089	-,004	-,291	-,148
Percentagem da população do sexo masculino literada	,869	,099	-,011	-,282	-,119
Taxa de mortalidade infantil	-,949	,121	-,047	,048	-,002
Taxa de mortalidade	-,738	,563	,159	-,018	-,172
Taxa de natalidade	-,910	-,273	,173	-,030	,023
Taxa de fecundidade	-,888	-,234	,224	,029	,044
Média de calorias diárias ingeridas por pessoa	,757	,054	,240	,220	,149
Percentagem de crescimento populacional por ano	-,564	-,724	,249	,028	,176
Quociente entre taxas de natalidade e mortalidade	,093	-,883	,110	-,056	,221
Casos sida no total da população	,180	,441	,587	-,032	,520
Taxa de sida	-,363	,495	,551	-,319	-,049
Produto interno bruto per capita	,628	,324	,494	,358	,150
Percentagem de crescimento da produção agrícola	-,037	,499	-,457	,106	,179
População em milhares	,063	,249	-,468	,057	,671
Densidade populacional	,173	,117	,000	,766	-,319

Figura 4.5: Matriz de pesos fatoriais.

No entanto, tal como já foi referido, a interpretação de cada um dos fatores é mais fácil recorrendo aos pesos fatoriais rotacionados. Os valores dos pesos fatoriais rotacionados segundo o critério Varimax são apresentados na tabela seguinte, denominada *rotated component matrix*.

	Component				
	1	2	3	4	5
Percentagem de pessoas que habitam nas cidades	<b>,678</b>	,318	,206	,310	,284
Longevidade da população do sexo feminino	<b>,911</b>	,320	,061	-,042	,170
Longevidade da população do sexo masculino	<b>,873</b>	,386	,062	-,091	,178
Percentagem da população literada	<b>,958</b>	-,060	,068	,084	-,088
Percentagem da população do sexo feminino literada	<b>,956</b>	-,080	,048	,086	-,081
Percentagem da população do sexo masculino literada	<b>,916</b>	-,081	,057	,059	-,086
Taxa de mortalidade infantil	<b>-,922</b>	-,210	-,119	-,028	-,111
Taxa de mortalidade	<b>-,691</b>	-,637	,129	,105	-,075
Taxa de natalidade	<b>-,902</b>	,169	-,084	,213	-,201
Taxa de fecundidade	<b>-,898</b>	,159	-,011	,211	-,144
Média de calorias diárias ingeridas por pessoa	<b>,653</b>	,148	,403	-,034	,303
Percentagem de crescimento populacional por ano	-,614	<b>,687</b>	-,055	,253	-,144
Quociente entre taxas de natalidade e mortalidade	,043	<b>,892</b>	-,111	,160	-,130
Casos sida no total da população	,085	-,104	<b>,895</b>	-,106	-,112
Taxa de sida	-,317	-,483	<b>,487</b>	,325	-,330
Produto interno bruto per capita	,484	-,062	<b>,688</b>	,037	,425
Percentagem de crescimento da produção agrícola	-,021	-,411	-,066	<b>-,569</b>	,068
População em milhares	,028	,015	,102	<b>-,847</b>	-,098
Densidade populacional	,049	-,117	-,027	,053	<b>,844</b>

Figura 4.6: Matriz de pesos fatoriais rotacionada segundo critério Varimax.

Recorde-se que os pesos fatoriais mais elevados em valor absoluto identificam o fator a que cada variável se associa. Assim, considerando apenas os pesos fatoriais em valor absoluto superiores a 0.45 (realçados a negrito na tabela *rotated component matrix*), podemos considerar os fatores que se descrevem a seguir.

- Fator 1: Este fator é constituído por 11 variáveis, das quais 7 são fortemente correlacionadas positivamente e 4 são fortemente correlacionadas negativamente. No primeiro grupo incluem-se as variáveis *Percentagem de pessoas que habitam nas cidades*, *Longevidade da população do sexo feminino*, *Longevidade da população do sexo masculino*, *Percentagem da população literada*, *Percentagem da população do sexo feminino literada*, *Percentagem da população do sexo masculino literada* e *Média de calorias diárias ingeridas por pessoa* e no

segundo grupo incluem-se as variáveis *Taxa de mortalidade infantil*, *Taxa de mortalidade*, *Taxa de natalidade* e *Taxa de fecundidade*. Estudos das ciências sociais indicam que as pessoas que habitam nas cidades são, em média, mais instruídas e têm elevada esperança média de vida. Além disso, nas cidades o número médio de calorias ingeridas por pessoa é grande, em consequência do popularizar das refeições à base de *fast food*. Por outro lado, nos mesmos locais, as taxas de mortalidade da população geral e infantil, assim como as taxas de natalidade e fecundidade, são baixas. Este motivo explica as correlações negativas das variáveis referidas. Um nome sugestivo para este fator é *Dicotomia cidade/campo*.

- Fator 2: *Percentagem de crescimento populacional por ano* e *quociente entre taxa de natalidade e taxa de mortalidade* são as variáveis que têm pesos fatoriais mais elevados neste fator. Era de esperar que as variáveis em questão fossem fortemente correlacionadas e constituíssem um fator, uma vez que, se o quociente referido for menor que 1 indica que existem mais mortes que nascimentos conduzindo ao decréscimo da população. Se esse quociente for maior que 1 indica o oposto. Este fator denominar-se-á *Demografia*.
- Fator 3: As variáveis que apresentam pesos fatoriais mais elevados neste fator são *Casos sida no total da população*, *Taxa de sida* e *Produto interno bruto per capita*. O produto interno bruto por pessoa é um indicador fiável do índice de desenvolvimento de um país. Neste sentido, quando este valor é baixo, as condições sociais no país em questão são propícias à delinquência, conduzindo ao aumento dos casos de HIV. Vamos denominar este fator de *Qualidade de vida*.
- Fator 4: As variáveis que apresentam pesos fatoriais mais elevados em valor absoluto neste fator são *Percentagem de crescimento da produção agrícola* e *População em milhares*. Em países com produções agrícolas abundantes e com tendência crescente, o número de habitantes é, em geral, mais elevado que na situação oposta. Este fator será denominado *Importância da produção agrícola para a população*.
- Fator 5: A única variável que entra neste fator é *Densidade populacional*. Esta variável não é fortemente correlacionada com as restantes variáveis que integram o exemplo. Desta forma, este fator designar-se-á *Densidade populacional*.

Em resumo, as 19 variáveis originais passam a constituir 5 fatores relacionados com a dicotomia cidade/campo, a demografia, a qualidade de vida, a importância da produção agrícola para a população e a densidade populacional. Verifica-se que o número de variáveis é relativamente pequeno, resultando em fatores com poucas variáveis, nomeadamente o quinto fator.

Para finalizar, observamos que a interpretação e rotulagem dos fatores pode ser muito subjetiva. Pode até haver mais do que uma solução para o problema, na medida em que podem ser considerados mais ou menos fatores e, para um mesmo número de fatores, podemos usar mais do que um método de rotação da matriz de pesos fatoriais. Neste sentido, o ideal é o estabelecimento de um diálogo contínuo entre pesquisador e estatístico para que o primeiro possa apontar possíveis soluções e o segundo possa informar os limites e possibilidades das técnicas estatísticas.



# Bibliografia

- [1] T. W. Anderson, An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, Inc., New York, 1958.
- [2] W. R. Dillon, M. Goldstein, *Multivariate Analysis, Methods and Applications*, John Wiley & Sons, Inc., New York, 1984.
- [3] R. A. Johnson, D. W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice-Hall International, Inc., New Jersey, 1992.
- [4] D. N. Lawley, Test of Significance for the Latent Roots of Covariance and Correlation Matrices, *Biometrika*, **43**, 128-136, 1956.
- [5] J. Maroco, *Análise Estatística, com utilização do SPSS*, Edições Sílabo, 3ª Edição, Lisboa, 2007.
- [6] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, Inc., New York, 1982.
- [7] M. H. Pestana, J. N. Gageiro, *Análise de dados para Ciências Sociais - A complementaridade do SPSS*, 4ª Ed., Edições Sílabo, Lisboa, 2005.