

ANÁLISE FATORIAL

JANDYRA MARIA GUIMARÃES FACHEL

DISSERTAÇÃO APRESENTADA AO  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DA  
UNIVERSIDADE DE SÃO PAULO  
PARA OBTENÇÃO DO GRAU DE MESTRE  
EM  
ESTATÍSTICA

ORIENTADOR:

PROF. DR. CLÓVIS DE ARAUJO PERES

- SÃO PAULO, NOVEMBRO DE 1976 -

Aos meus pais  
e ao Fernando

---

## PREFÁCIO

Ao escrevermos este trabalho procuramos apresentar a teoria estatística da Análise Fatorial, com ênfase nos problemas de estimação e testes de hipóteses, de forma clara e de acordo com os recentes desenvolvimentos do objeto.

No capítulo I expomos o que é Análise Fatorial, relacionando-a com outras técnicas de Análise Multivariada e incluímos uma nota histórica sobre o assunto.

O modelo teórico de Análise Fatorial é apresentado no capítulo II, que é complementado por uma breve exposição sobre Análise de Componentes Principais, comparando-se, então, os dois modelos.

O capítulo III trata de estimação e testes de hipóteses no modelo fatorial. Nele são apresentados, com detalhes, o método de estimação de máxima verossimilhança e os testes da razão de verossimilhança em Análise Fatorial. Outros métodos de estimação são também abordados.

Os diversos modelos de rotação e transformação dos fatores, com exemplos, são mostrados no capítulo IV.

No último capítulo apresentamos dois exemplos da aplicação da técnica de Análise Fatorial

\*  
\* \* \*

Finalmente, gostaríamos de externar nossos agrade-

cimentos a todos que direta ou indiretamente contribuíram para a realização deste trabalho, em especial

ao Professor Clóvis de Araujo Peres, que além de orientador, foi um amigo em todos os momentos, transmitindo-nos com dedicação seus conhecimentos e experiência;

aos Professores Pedro A. Morettin e Herbert G. Callhau por seus valiosos comentários;

a Nicolas Botto Correa e a José Alberto Moreno;

aos Professores do Departamento de Estatística do Instituto de Matemática e Estatística da Universidade de São Paulo;

aos colegas Mariane Streibel e João Beal Vargas;

aos colegas do Departamento de Estatística da Universidade Federal do Rio Grande do Sul;

ao sr. João Baptista Esteves de Oliveira pelo eficiente trabalho de datilografia.

## Í N D I C E

|  |    |
|--|----|
| CAP. I - INTRODUÇÃO. . . . .   | 1  |
| 1.1 - O Problema. . . . .  | 1  |
| 1.2 - O que é Análise Fatorial. . . . .                                    | 2  |
| 1.3 - Nota Histórica. . . . .  | 4  |
| CAP. II - ANÁLISE FATORIAL E SEU MODELO TEÓRICO. . . . .                   | 8  |
| 2.1 - Conceitos Preliminares. . . . .                                      | 8  |
| 2.2 - O Modelo Fatorial e seus Parâmetros . . . . .                        | 11 |
| 2.3 - Análise de Componentes Principais . . . . .                          | 16 |
| CAP. III - ESTIMAÇÃO E TESTES DE HIPÓTESES NO MODELO FATORIAL. . . . .     | 22 |
| 3.1 - Estimacão pelo Método de Máxima Verossimilhança . . . . .            | 22 |
| 3.1.1 - Modelo não restrito. . . . .                                       | 24 |
| 3.1.2 - Modelo restrito. . . . .   | 31 |
| 3.2 - Teste da Razão de Verossimilhança no Modelo Fatorial. . . . .        | 34 |
| 3.2.1 - Modelo não restrito. . . . .                                       | 34 |
| 3.2.2 - Modelo restrito. . . . .   | 37 |
| 3.3 - Outros Métodos de Estimacão . . . . .                                | 39 |
| - Solução do Fator Principal. . . . .                                      | 40 |
| - Método Centróide. . . . .  | 43 |
| 3.4 - Estimacão dos Escores Fatoriais . . . . .                            | 43 |
| 1) Método de Regressão (Método de Thomson). . . . .                        | 44 |
| 2) Método de Bartlett . . . . .  | 46 |
| CAP. IV - TRANSFORMAÇÃO E INTERPRETAÇÃO DOS FATORES. . . . .               | 48 |
| 4.1 - Rotação dos Fatores . . . . .  | 48 |
| 4.2 - Transformacão dos Fatores - Caso Obliquo. . . . .                    | 59 |
| CAP. V - ALGUMAS APLICAÇÕES DA ANÁLISE FATORIAL. . . . .                   | 66 |
| 5.1 - Exemplos do Uso da Análise Fatorial em Pesquisas Agrícolas . . . . . | 66 |
| 5.2 - Modelo de Regionalizacão. . . . .                                    | 72 |
| REFERÊNCIAS BIBLIOGRÁFICAS . . . . .                                       | 79 |

## CAPÍTULO I

### INTRODUÇÃO

#### 1.1 - O PROBLEMA

Nas mais diversas áreas do conhecimento, como por exemplo, Psicologia, Economia, Agricultura, Biologia, Educação, Ciências Sociais, o pesquisador, não raras vezes se depara com observações de várias variáveis para cada elemento de uma amostra de indivíduos, de animais, ou de outros tipos de unidades experimentais.

Em muitas situações é desejável examinar as inter-relações entre as variáveis. Estas inter-relações podem ser medidas ou pelas covariâncias ou pelos coeficientes de correlação entre as variáveis. Se o número de variáveis é grande, o pesquisador pode estar interessado em estruturar e simplificar seus dados de maneira a conservar o máximo de informação expressa pelas variáveis originais.

Uma solução para este problema é encontrar variáveis hipotéticas que sejam combinações lineares das variáveis observadas e que, mesmo em menor número possam ser mais convenientemente estudadas.

Duas técnicas estatísticas de Análise Multivariada são comumente utilizadas para tratar este problema: Análise

de Componentes Principais e Análise Fatorial. Com frequência, estas duas técnicas são confundidas. Nosso trabalho versará fundamentalmente sobre Análise Fatorial, mas com o intuito de esclarecer idéias, distinguiremos os dois métodos. No Capítulo II, seção 2.3, apresentaremos de forma resumida a técnica de Análise de Componentes Principais.

## 1.2 - O QUE É ANÁLISE FATORIAL?

Análise Fatorial é uma técnica de Análise Multivariada que trata das relações internas de um conjunto de variáveis, substituindo um conjunto inicial de variáveis correlacionadas por um conjunto menor de "fatores" (ou variáveis hipotéticas) que podem ser não correlacionados (Fatores Ortogonais) ou Correlacionados (Fatores Oblíquos), e que explicam a maior parte da variância do conjunto original.

Como uma técnica de Análise Multivariada é relevante mostrar como se situa a Análise Fatorial em relação às outras técnicas. Segundo Kendall (1950), as técnicas de Análise Multivariada podem ser distinguidas em:

- a) Análise de Dependência: quando queremos estudar a dependência de uma ou mais variáveis em relação às outras. Consideramos, então, dois subconjuntos, um no qual as variáveis são denominadas independentes e outro em que tratamos das variáveis dependentes.
- b) Análise de Interdependência: quando estamos interessados nas relações de um conjunto de variáveis entre si, sem selecionarmos nenhuma delas, em especial, como variável dependente.

No primeiro tipo de Análise enquadram-se, por exemplo, Análise de Regressão e Análise de Variância Multivariada, enquanto que é no segundo tipo de classificação, salien

tando-se apenas o caráter de interdependência das variáveis, que se enquadram as técnicas de Análise de Componentes Principais e Análise Fatorial.

No caso bivariado vemos claramente esta distinção entre Análise de Dependência e de Interdependência: a correlação entre duas variáveis é uma questão de interdependência e é uma relação simétrica entre as variáveis, enquanto que a regressão de uma variável sobre a outra, trata-se de dependência e não é uma relação simétrica, isto é, a regressão de  $x$  sobre  $y$  não é a mesma que a regressão de  $y$  sobre  $x$ .

Em Psicologia e em outras ciências esta distinção tem seu paralelo. Os psicólogos distinguem entre fidedignidade, que trata de consistência interna ou interdependência, e validade, que trata da concordância com um critério externo ou de dependência.

A Análise Fatorial tem uma peculiaridade própria em relação a outras técnicas estatísticas, por ter sido desenvolvida por pessoas que não eram primariamente estatísticos e sim psicólogos. Por esta razão ela tendeu a adquirir uma linguagem e técnica próprias. O assunto, como veremos no histórico, era limitado à concepção psicológica de fatores mentais, tal que um estatístico comum teria dificuldades em relacioná-lo aos métodos gerais da estatística. Alguns estatísticos nem mesmo acreditam na Análise Fatorial como um todo. É então, útil uma exposição sobre Análise Fatorial como uma técnica estatística, para que os pesquisadores, cujo interesse principal está em aplicações fora do campo da Psicologia possam utilizá-la. Este é um dos objetivos deste trabalho.

Mais recentemente tem sido proposto um uso diversificado da Análise Fatorial, considerado por alguns estudiosos como seu valor principal, que é o de complementar técni

cas estatísticas convencionais. Uma destas técnicas é a Análise de Regressão, e em alguns casos especiais, é mostrado como um melhor subconjunto de variáveis independentes pode ser encontrado usando-se Análise Fatorial, para predizer uma variável dependente. Também, em um trabalho atual, Lawley e Maxwell (1973) publicaram um artigo intitulado: "Regression and Factor Analysis", onde um modelo básico de Análise Fatorial é empregado na estimação dos coeficientes de regressão. Entretanto, é ainda pequeno o número de trabalhos publica-dos a este respeito.

### 1.3 - NOTA HISTÓRICA

A Análise Fatorial surgiu, no início deste século, como um desenvolvimento natural da análise de conjuntos de variáveis correlacionadas.

A técnica de Análise Fatorial foi desenvolvida inicialmente pelo psicólogo Charles Spearman, que em 1904 pu-blicou o artigo intitulado "General Intelligence Objective-ly Determined and Measured". Neste artigo, Spearman exami-na os trabalhos anteriores sobre medida de inteligência e apresenta, pela primeira vez, sua teoria de um fator de inteligência geral, seu famoso fator "g". Esta variável hipotéti-ca explicaria, em geral, as intercorrelações de um conjunto de testes cognitivos.

Bartlett (1962) comenta que o modelo proposto por Spearman, com apenas um fator comum, foi uma primeira apro-ximação a situações reais, mas que, apesar de útil, era inadequado. Assim, mais "fatores comuns" necessitaram ser in-troduzidos no modelo, dando lugar a uma especificação mais geral do mesmo.

Surgiram, então, os conceitos de "fatores múltiplos"

que foram desenvolvidos por L.L.Thurstone, da escola americana, e também por Burt e Thonson, da escola inglesa, que como Spearman, eram psicólogos. Estes foram os precursores mais ativos da Análise Fatorial. Detalhes de suas contribuições e referências aos primeiros artigos podem encontrar-se em seus textos: Thurstone (1935), (1947), Burt (1941) e Thonson (1951).

Como vimos todo o desenvolvimento inicial da Análise Fatorial foi no campo da Psicologia, para o estudo de hipóteses sobre a organização da habilidade mental. A análise partia do exame das matrizes de correlação ou covariância de um conjunto, ou de uma bateria de testes cognitivos. Os métodos para obtenção dos parâmetros do modelo eram métodos empíricos, aproximados. Várias controvérsias surgiram sobre o assunto, que por um longo tempo desencorajaram o interesse mostrado pelos matemáticos e estatísticos nos problemas teóricos envolvidos.

A primeira abordagem rigorosa da Análise Fatorial, do ponto de vista da estatística teórica, tratando do problema de estimação, foi feita por Lawley (1940). Através desta nova abordagem ao problema, Lawley eliminou as indeterminações e decisões subjetivas que acompanhavam outros métodos para a determinação dos coeficientes fatoriais a partir da matriz de covariância ou correlação das variáveis observadas, como por exemplo o Método Centróide de Thurstone.

Usando então os resultados de Inferência Estatística, a partir de suposições impostas ao modelo, a Análise Fatorial passou a ser considerada um método estatístico propriamente dito. Assim, teste de hipóteses, como o teste sobre a adequacidade de um modelo com "m" fatores foram desenvolvidos rigorosamente, pelo princípio da razão de verossimilhança.

A partir de 1940, Lawley publicou outros trabalhos tratando de diversos problemas sobre Análise Fatorial e ainda outras contribuições foram feitas por Howe (1955), Anderson e Rubin (1956) e Rao (1955), para citar apenas algumas.

Com o advento dos computadores eletrônicos, novas técnicas de Análise Fatorial foram desenvolvidas, tornando-se cada vez mais fácil o seu uso, não só no campo da Psicologia mas também nos vários outros ramos da Ciência. Há exemplos de aplicação de Análise Fatorial nas mais diversas áreas do conhecimento e que vem sendo publicados regularmente.

O método de máxima verossimilhança para a estimação dos parâmetros do modelo não havia sido utilizado anteriormente porque o método implicava em cálculos trabalhosos que não eram possíveis sem o uso do computador. Entretanto, a partir de 1966, trabalhos devidos em grande parte ao estatístico sueco Jöreskog proporcionaram procedimentos de estimação rápidos e eficientes. Jöreskog usou inicialmente um método numérico para a minimização de uma função de várias variáveis e posteriormente desenvolveu ainda outros métodos mais adequados.

Os textos modernos sobre Análise Fatorial, considerando sua importância histórica e sua diversificação quanto aos temas são os de Harman (1960) e de Lawley e Maxwell (1971). Nestas duas obras encontram-se referências bibliográficas numerosas sobre o assunto. Em vários textos de Análise Multivariada, como os de Morrison (1967), Press (1972), Afifi e Azen (1972) e outros, encontram-se capítulos com um bom tratamento do assunto.

Vimos então, e Harman (1960) comenta, que foram as novas técnicas de computação, os procedimentos objetivos para determinar as soluções e os testes de hipóteses, os gran

des responsáveis por transportar a Análise Fatorial do abis  
mo de uma visão apenas psicológica para um respeitado ramo  
da Análise Estatística Multivariada.

## CAPÍTULO II

### ANÁLISE FATORIAL E SEU MODELO TEÓRICO

#### 2.1 - CONCEITOS PRELIMINARES

Dado um conjunto de  $p$  variáveis, com  $N$  observações para cada variável, que formam o arranjo de valores

$$(X_{ik}), \quad i=1, \dots, p, \quad k=1, \dots, N,$$

denominado matriz de dados e supondo que as  $p$  variáveis são correlacionadas, pergunta-se se é possível reduzir a dimensão inicial ( $p$ ) do problema através de novas variáveis hipotéticas que mesmo em menor número explicarão a maior parte da variação original. Assim, supõe-se que cada variável  $X_i$ ,  $i=1, \dots, p$  é representada como uma função linear de um menor número ( $m < p$ ) de variáveis hipotéticas, denominadas fatores comuns (por serem comuns a várias variáveis) mais um fator residual ou único para cada variável.

Sejam  $X_1, X_2, \dots, X_p$ , variáveis aleatórias com

$$\text{Var } X_i = \sigma_i^2 = \sigma_{ii} \quad \text{e} \quad \text{Cov}(X_i, X_r) = \sigma_{ir}, \quad i \neq r.$$

Como estamos interessados apenas na estrutura de covariân-

cia dos dados, podemos supor sem perda de generalidade que cada variável tem média zero.

De acordo com o que foi exposto acima, podemos escrever cada variável no modelo fatorial como:

$$X_i = \sum_{j=1}^m \lambda_{ij} f_j + e_i, \quad i=1, \dots, p \quad (2.1.1)$$

onde  $f_j$  é o  $j$ -ésimo fator comum, sendo que o número  $m$  de tais fatores é especificado e onde  $e_i$  é um resíduo representando fontes de variação que afetam apenas a variável  $X_i$ . Os coeficientes  $\lambda_{ij}$  são denominados *cargas fatoriais* e refletem a importância do  $j$ -ésimo fator na composição da  $i$ -ésima variável  $X_i$ .

Supomos também que as  $p$  variáveis aleatórias  $e_i$  são não correlacionadas entre si e não correlacionadas com os fatores comuns. Estes fatores  $f_j$  podem ser correlacionados entre si (oblíquos) ou não correlacionados (ortogonais). A variância de  $e_i$ , denominada variância residual (ou especificidade), é denotada por  $\psi_i$ .

As variâncias e covariâncias das variáveis observadas, sob a suposição de fatores não correlacionados, podem, então, ser escritas como

$$\sigma_{ii} = \sigma_i^2 = \sum_{j=1}^m \lambda_{ij}^2 + \psi_i, \quad i=1, 2, \dots, p$$

e

$$\sigma_{ir} = \sum_{j=1}^m \lambda_{ij} \lambda_{rj}, \quad i \neq r \quad r=1, 2, \dots, p.$$

A parte da variância de cada variável, devida só aos fatores comuns é dada por

$$h_i^2 = \sigma_i^2 - \psi_i = \sum_{j=1}^m \lambda_{ij}^2$$

e é denominada *comunalidade* da variável  $X_i$ .

Na prática, os parâmetros do modelo a serem estimados, a partir dos dados experimentais, são os  $\lambda_{ij}$  (cargas fatoriais) e os  $\psi_i$  (especificidades).

Se os fatores  $f_j$  são supostos correlacionados, os coeficientes de correlação entre os fatores são parâmetros adicionais que também requerem estimação, como veremos na próxima seção.

Após as cargas fatoriais terem sido obtidas, o analista deve interpretar os fatores comuns da melhor maneira possível. Com este objetivo faz-se a Rotação dos Fatores.

Finalmente pode-se estimar os valores dos fatores comuns para cada indivíduo, como função das variáveis observadas. Estes valores denominam-se *escores fatoriais*. Segundo Lawley e Maxwell (1971), estimação em Análise Fatorial é um procedimento em dois estágios. Primeiro os parâmetros do modelo são estimados e então eles são usados para determinar os escores fatoriais.

Observando as equações do modelo fatorial vemos que a especificação de cada variável  $X_i$  é similar na forma, à especificação da teoria de Regressão. Em Análise Fatorial, entretanto, os fatores, que fazem o papel das variáveis independentes, não são observáveis.

Anderson e Rubin (1956) distinguem duas espécies de modelos em Análise Fatorial, um considerando os fatores

como variáveis aleatórias, e outro considerando os fatores como quantidades não aleatórias que variam de um indivíduo a outro. Whittle (1953) considera também os fatores como parâmetros e usa os procedimentos de mínimos quadrados para estimá-los, minimizando a soma de quadrados residual. Esta abordagem não é comum nas aplicações práticas e portanto não a trataremos neste trabalho. Deteremo-nos no Método de Estimção de Máxima Verossimilhança desenvolvido por Lawley para a estimação dos parâmetros do Modelo de Análise Fatorial, tanto no caso de fatores ortogonais, como oblíquos.

## 2.2 - O MODELO FATORIAL E SEUS PARÂMETROS

Na seção anterior introduzimos o modelo de Análise Fatorial. Consideraremos agora este modelo com maiores detalhes, concentrando nossa atenção sobre seus parâmetros, do modelo. A partir desta seção usaremos a notação matricial por conveniência. (Os vetores serão denotados em negrito ( $X$ ) e as matrizes por letras maiúsculas).

Sejam  $X$  e  $e$  os vetores colunas com elementos  $X_i$  e  $e_i$  ( $i=1, \dots, p$ ) respectivamente e seja  $f$  o vetor coluna com elementos  $f_j$  ( $j=1, \dots, m$ ). Então as equações 2.1.1 são escritas na forma matricial da seguinte maneira:

$$X = Af + e \quad (2.2.1)$$

onde  $A = [\lambda_{ij}]$  é a matriz  $p \times m$  das cargas fatoriais. Estas equações não podem ser verificadas diretamente pois as  $p$  variáveis observadas são expressas em termos de  $p+m$  outras variáveis não observáveis.

A matriz de covariância de  $X$  é denotada por

$$\Sigma = [\sigma_{ir}]$$

e é uma matriz simétrica, positiva definida e de ordem p. Os vetores aleatórios e e f são supostos não correlacionados e então

$$E(ef') = 0.$$

A matriz de covariância de e é dada por

$$E(ee') = \psi$$

onde  $\psi$  é uma matriz diagonal de ordem p cujos elementos diagonais são  $\psi_1, \psi_2, \dots, \psi_p$ .

Inicialmente consideremos o caso dos fatores não correlacionados (ortogonais) e com variâncias unitárias. Sua matriz de covariância é então dada por

$$E(ff') = I_m$$

onde  $I_m$  é a matriz unitária de ordem m.

Das suposições consideradas acima, temos

$$\Sigma = E(XX') = E[(Af+e)(Af+e)']$$

ou

$$\Sigma = \Lambda\Lambda' + \psi. \quad (2.2.2)$$

Esta representação de  $\Sigma$  é fundamental e os parâmetros desconhecidos a serem estimados são os elementos de  $\Lambda$  e  $\psi$ . Assim, dada a matriz  $\Sigma$  nós perguntamos se para um valor de m, especificado, menos do que p, é possível definir

uma única  $\psi$  com elementos diagonais positivos, e uma única matriz  $\Lambda$ ,  $p \times m$ , satisfazendo a equação 2.2.2, para que possamos obter estimadores consistentes dos parâmetros.

As equações 2.2.1 e 2.2.2 são também satisfeitas se se nós trocarmos  $f$  por  $M'f$  e  $\Lambda$  por  $\Lambda M$ , sendo  $M$  qualquer matriz ortogonal de ordem  $m$ , e  $M'$  sua transposta. Assim existirá uma infinidade de escolhas para a matriz  $\Lambda$ , pois qualquer solução  $\Lambda$  pode ser trocada por  $\Lambda M$ . Deve-se, então, impor algumas restrições sobre  $\Lambda$  para que esta matriz possa ser unicamente determinada, partindo-se da suposição que existe um único  $\psi$ .

Anderson e Rubin (1956) denominam o problema de resolver as equações (2.2.2) de maneira única de problema de identificação. Em seu artigo, eles apresentam várias condições que são ou necessárias ou suficientes para unicidade de  $\Lambda$  e  $\psi$ , não havendo nenhuma condição necessária e suficiente. Nós não as discutiremos integralmente aqui, apenas nos referiremos à algumas restrições que podem ser feitas sobre os elementos de  $\Lambda$ , para identificação, e que serão convenientes para os procedimentos de estimação do Capítulo III.

No caso de fatores ortogonais,  $m > 1$  e supondo que exista uma única matriz diagonal  $\psi$  com elementos positivos, escolhe-se  $\Lambda$  tal que a matriz  $\Lambda' \psi^{-1} \Lambda$ , que é uma matriz  $m \times m$ , seja diagonal e que os elementos diagonais sejam positivos, distintos e arranjados em ordem decrescente de grandeza. Com esta restrição sobre  $\Lambda$ , esta matriz é univocamente definida. Os elementos diagonais da matriz  $\Lambda' \psi^{-1} \Lambda$  são as  $m$  raízes características não nulas da matriz  $\psi^{-1/2} (\Sigma - \psi) \psi^{-1/2}$  que é uma matriz simétrica e de posto  $m$ .

Consideremos agora a questão de unicidade de  $\psi$ . Segundo Lawley e Maxwell, uma condição necessária simples pa-

ra a unicidade de  $\psi$  e que vale para qualquer valor de  $m$ , é que toda coluna de  $\Lambda$  deve ter ao menos três elementos diferentes de zero.

Outras considerações sobre a unicidade dos parâmetros são as abordadas por Lawley e Maxwell (1971): Seja  $H_m$  a hipótese que, para um valor especificado de  $m$ , uma dada matriz de covariância  $\Sigma$  possa ser expressa na forma  $\Sigma = \Lambda\Lambda' + \psi$ .

A condição que  $\Lambda'\psi^{-1}\Lambda$  seja diagonal, para  $m > 1$ , impõe  $\frac{1}{2} m(m-1)$  restrições sobre os parâmetros. Assim o número de parâmetros livres em  $\psi$  e  $\Lambda$  é  $p + pm - \frac{1}{2} m(m-1)$ . Igualando-se os correspondentes elementos das matrizes em ambos os lados da equação  $\Sigma = \Lambda\Lambda' + \psi$ , obtemos  $\frac{1}{2} p(p+1)$  equações distintas. Seja  $C$  a diferença entre o número de equações e o número de parâmetros livres. Então

$$\begin{aligned} C &= \frac{1}{2} p(p+1) - p - pm + \frac{1}{2} m(m-1) = \\ &= \frac{1}{2} [(p-m)^2 - (p+m)] \end{aligned}$$

- a) Se  $C = 0$  teremos tantas equações quantos parâmetros livres, e a hipótese  $H_m$  será, em geral, trivialmente verdadeira e  $\psi$  será univocamente determinado.
- b) Se  $C < 0$ , existem menos equações do que parâmetros livres e  $H_m$  será trivialmente verdadeira mas com uma infinidade de escolhas para  $\psi$  e  $\Lambda$ .
- c) Se  $C > 0$  há mais equações do que parâmetros livres e então  $H_m$  é não trivial, a menos que se façam restrições sobre os elementos de  $\Sigma$ .

Antes de tratarmos o caso dos fatores oblíquos faremos a seguinte observação sobre os parâmetros  $\lambda_{ij}$ . Das su

posições do modelo fatorial, no caso ortogonal, decorre que

$$\text{Cov}(X, f') = E[(\Lambda f + e)f'] = \Lambda E f f' + E(e f') = \Lambda$$

isto é, os parâmetros  $\lambda_{ij}$  representam a covariância da  $i$ -ésima variável  $X_i$  com o  $j$ -ésimo fator comum. Em problemas práticos é comum transformar as variáveis observáveis em variáveis padronizadas com média zero e variância unitária. Desta forma, a matriz  $\Sigma$  será a matriz de correlação dos dados observados e os parâmetros  $\lambda_{ij}$  serão interpretados como os coeficientes de correlação entre a  $i$ -ésima variável  $X_i$  com o  $j$ -ésimo fator comum  $f_j$ .

Consideremos agora o caso em que os fatores  $f_j$ ,  $j = 1, \dots, m$ ,  $m > 1$ , são supostos correlacionados tal que

$$\text{Var } f_j = \phi_{jj} \quad \text{e} \quad \text{Cov}(f_j, f_k) = \phi_{jk}, \quad j \neq k, \quad j = 1, \dots, m.$$

A matriz de covariância de  $f$  é dada então por

$$E f f' = \Phi = [\phi_{jk}],$$

sendo  $\Phi$  uma matriz simétrica de ordem  $m$ .

A matriz de covariância de  $X$  é então dada por

$$\Sigma = \Lambda \Phi \Lambda' + \psi \tag{2.2.3}$$

Os parâmetros do modelo são agora  $\Lambda$ ,  $\psi$  e  $\Phi$ . As equações (2.2.1) e (2.2.3) continuam satisfeitas se nós trocarmos  $f$  por  $Uf$ ,  $\Lambda$  por  $\Lambda U^{-1}$  e  $\Phi$  por  $U\Phi U'$ , onde  $U$  é qualquer matriz não singular de ordem  $m$ .

A questão da unicidade dos parâmetros, neste caso, será agora discutida. A matriz  $U$  tem  $m^2$  elementos, assim ao menos  $m^2$  restrições sobre os elementos de  $\Phi$  e  $\Lambda$  são necessá

rias para que eles possam ser definidas univocamente.

É usual supor que os fatores têm variâncias unitárias, então  $m$  restrições são obtidas igualando-se os elementos diagonais de  $\Phi$  à unidade. As restantes restrições são impostas requerendo que certos elementos de  $\Lambda$  sejam nulos. Segundo Lawley e Maxwell (1971) se a configuração de zeros em  $\Lambda$  é tal que ela possa ser destruída por qualquer transformação não singular dos fatores, então  $\Lambda$  e  $\Phi$  são univocamente determinados, dados  $\Sigma$  e  $\psi$ . No Capítulo III consideraremos o problema de estimação quando as restrições são dadas por elementos zero em posições especificados na matriz  $\Lambda$  (modelo restrito).

### 2.3 - ANÁLISE DE COMPONENTES PRINCIPAIS

A Análise de Componentes Principais é uma técnica de Análise Multivariada introduzida por K. Pearson em 1901 e desenvolvida posteriormente por H. Hotelling, a partir de 1933, pela qual um conjunto de  $p$  variáveis  $X_1, X_2, \dots, X_p$  é transformado linearmente em um número igual de novas variáveis  $Y_1, Y_2, \dots, Y_p$ , não correlacionadas e ordenadas por suas variâncias em ordem decrescente e de tal forma que a variância total do conjunto inicial é preservada. Esta técnica é relativamente bem conhecida e a apresentaremos aqui resumidamente.

Na prática, o método de Componentes Principais é utilizado para encontrarmos combinações lineares com grandes variâncias pois, algumas vezes, as primeiras poucas componentes explicam uma grande proporção da variância total das variáveis originais. Estas componentes principais podem, então, para certos propósitos, serem utilizados para descrever com maior parcimônia a estrutura de dependência das va-

riáveis observáveis.

Em Análise de Componentes Principais nós não partimos de um modelo assim como em Análise Fatorial. A exposição que segue será feita supondo que o leitor está familiarizado com a teoria sobre raízes e vetores características de uma matriz.

A seguir definiremos componentes principais na população.

Seja o vetor aleatório  $X$  de  $p$  componentes com

$$E(X) = 0 \text{ e } EXX' = \Sigma.$$

Supomos que as raízes características da matriz de covariância  $\Sigma$ , denotadas por  $\delta_1, \delta_2, \dots, \delta_p$  são distintas e arranjadas em ordem decrescente de grandeza. Da teoria sobre raízes e vetores característicos sabe-se que a matriz  $\Sigma$  (simétrica e real) pode ser escrita como

$$\Sigma = \Gamma \Delta \Gamma'$$

onde

$$\Delta = \begin{bmatrix} \delta_1 & 0 & \dots & 0 \\ 0 & \delta_1 & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & \delta_p \end{bmatrix}$$

e onde  $\Gamma$  é matriz ortogonal de ordem  $p$ . A  $j$ -ésima coluna de  $\Gamma$  é o vetor característico normalizado associado a raiz  $\delta_j$ .

O vetor de componentes principais  $Y$ , cujas componentes são  $Y_1, Y_2, \dots, Y_p$ , é definido pela transformação li

near ortogonal

$$Y = \Gamma'X$$

tal que a matriz de covariância de  $Y$  é dada por

$$E(YY') = E(\Gamma'XX'\Gamma) = \Gamma'\Sigma\Gamma = \Gamma'(\Gamma\Delta\Gamma')\Gamma = \Delta$$

e portanto as componentes principais  $Y_1, \dots, Y_p$  são não correlacionadas e a variância de  $Y_j$  é  $\delta_j$ . Além disso, a  $j$ -ésima componente de  $Y$ ,  $Y_j$  tem variância máxima entre todas as combinações lineares normalizadas (tais que a soma de quadrados dos coeficientes é igual a 1) não correlacionados com  $Y_1, \dots, Y_{j-1}$ .

Outra propriedade importante na Análise de Componentes Principais é que:

$$\text{tr}(\Sigma) = \text{tr}(\Gamma\Delta\Gamma') = \text{tr}(\Delta\Gamma\Gamma') = \text{tr}\Delta = \sum_{j=1}^p \delta_j.$$

Isto significa que a variância total das variáveis  $X$  é igual a variância total das variáveis  $Y$ , ou seja, a variação total das variáveis originais é preservada por esta transformação linear no método de Componentes Principais.

A importância da  $j$ -ésima componente principal é medida por

$$\frac{\delta_j}{\text{tr}\Sigma}, \quad j=1, \dots, p$$

e a proporção da variância total explicada pelas primeiras  $m$  componentes principais é dada por

$$\frac{\delta_1 + \delta_2 + \dots + \delta_m}{\text{tr}\Sigma}$$

Observamos que existe uma relação inversa simples entre as variáveis originais e as componentes deduzidas. Isto é, cada variável  $X_i$  pode ser expressa como uma combinação linear das componentes principais. Assim temos:

$$X = \Gamma Y \quad \text{e} \quad \Sigma = \Gamma \Delta \Gamma'$$

Este conjunto de equações é chamado o modelo de Componentes Principais. Neste modelo a matriz de covariância,  $\Sigma$ , pode ser fatorada da seguinte maneira: seja  $B = \Gamma \Delta^{1/2}$ , então

$$\Sigma = \Gamma \Delta \Gamma' = \Gamma \Delta^{1/2} \Delta^{1/2} \Gamma' = B B'$$

onde  $\Delta^{1/2}$  é a matriz diagonal, cujos elementos diagonais são  $\sqrt{\delta_1}, \sqrt{\delta_2}, \dots, \sqrt{\delta_p}$ . Esta fatoração é única pois foi feita a suposição inicial de que todas as raízes características de  $\Sigma$  são distintas e arranjadas em ordem decrescente de grandeza. Esta matriz  $B$  não deve ser submetida a alguma pós-multiplicação por uma matriz ortogonal (rotação), pois desta forma as novas componentes não teriam mais a propriedade de máxima variância.

Uma importante diferença entre as equações básicas dos modelos de Componentes Principais e de Análise Fatorial é que no primeiro caso não há componente residual. Em outras palavras, o modelo de Componentes Principais atribui toda a variação original às  $p$  componentes.

Um método iterativo direto para obter os elementos  $\alpha_{ij}$  da matriz  $\Gamma$  é dado por Hotelling em seus artigos origi-

nais e encontram-se também em alguns textos de Análise Multi-variada, como o de Morrison (1967).

Em aplicações práticas é comum, quando as variáveis estão medidas em unidades muito diferentes entre si, padronizar cada variável  $X_i$ , já que combinações lineares das variáveis originais poderiam ter pouco significado. Usa-se, então, a matriz de correlação dos dados originais ao invés da matriz de covariância. Neste caso a variância total das variáveis padronizadas é dado por  $p$ . As componentes principais obtidas da matriz de correlação não são as mesmas do que as obtidas da matriz de covariância e esta é uma das desvantagens do método de Componentes Principais.

A correlação entre a  $i$ -ésima variável  $X_i$  e a  $j$ -ésima componente é dada por

$$\text{Corr}(X_i, Y_j) = \frac{\alpha_{ij} \sqrt{\delta_j}}{\sigma_{ii}}, \quad i, j = 1, \dots, p$$

onde  $\sigma_{ii}$  é o desvio padrão da variável  $X_i$ . Se a matriz de correlação é usada para se fazer a Análise de Componentes Principais, então  $\text{Corr}(X_i, Y_j) = \alpha_{ij} \sqrt{\delta_j}$  e a proporção da variância total atribuída a  $j$ -ésima componente é dada por

$$\frac{\delta_j}{p}.$$

Supomos, agora, que  $\Sigma$  não é conhecida, como ocorre frequentemente em pesquisa aplicada. Então o parâmetro  $\Sigma$  deve ser estimado e a Análise de Componentes Principais deve ser feita com base em dados amostrais. Supomos que uma amostra de  $N(>p)$  conjuntos de observações das  $p$  variáveis em  $X$  foram obtidas. Destes dados os estimadores usuais, não vie

sados, das variâncias e covariâncias são encontrados, formando uma matriz S, simétrica e positiva definida, dada por

$$S = \frac{1}{N-1} \left( \sum_{\alpha} X_{\alpha} X_{\alpha}' - N \bar{X} \bar{X}' \right)$$

onde  $X_{\alpha}$  ( $\alpha=1, \dots, N$ ) é o vetor coluna representando o  $\alpha$ -ésimo conjunto de observações e  $\bar{X}$  é o vetor de médias amostral.

As componentes principais para a amostra são definidas exatamente como para a população, usando S em lugar de  $\Sigma$  e elas têm propriedades análogas. Além disso, se supormos que X tem distribuição normal multivariada e seguindo a teoria de estimação obteremos estimadores de máxima verossimilhança para  $\Delta$  e  $\Gamma$ . Estudos sobre isto foram feitos originalmente por Girshick (1936) e posteriormente por Anderson (1958).

É possível também encontrar-se as variâncias e covariâncias assintóticas das raízes características  $\delta_j$  de S e dos coeficientes das componentes principais. Estes resultados foram obtidos por Girshick (1939) e encontram-se em alguns dos textos modernos sobre Componentes Principais. Variâncias e covariâncias assintóticas das raízes características de uma matriz de correlação amostral R foram também encontradas por Girshick (1939).

## CAPÍTULO III

### ESTIMAÇÃO E TESTES DE HIPÓTESES

#### NO MODELO FATORIAL

Numerosos métodos aproximados para a estimação dos parâmetros do modelo fatorial foram propostos desde os primórdios da Análise Fatorial. Destes métodos os mais conhecidos são a Solução do Fator Principal e o Método Centróide. Na seção 3.3 faremos comentários sobre esses métodos. Mas para um enfoque rigoroso ao problema da estimação em Análise Fatorial o Método de Máxima Verossimilhança, proposto mais recentemente, é o mais indicado e o apresentaremos aqui com detalhes.

#### 3.1 - ESTIMAÇÃO PELO MÉTODO DE MÁXIMA VEROSSIMILHANÇA

Os procedimentos de máxima verossimilhança, em geral, são bem conhecidos em Inferência Estatística. Os estimadores obtidos por este método satisfazem a maioria das propriedades frequentemente usadas para selecionar estimadores entre várias escolhas possíveis, como as propriedades de consistência, eficiência e suficiência.

O método de máxima verossimilhança para Análise Fa

torial foi introduzido por Lawley em 1940 e não era muito usado anteriormente por suas dificuldades computacionais. Atualmente, no entanto, devido aos trabalhos de Jöreskog já se encontram procedimentos rápidos e eficientes para a obtenção dos estimadores por este método. Mas a principal vantagem de utilizar o método de máxima verossimilhança para Análise Fatorial é, talvez, a possibilidade de se desenvolverem testes de hipóteses com o objetivo de testar a adequacidade do modelo, o que não era possível através de outros métodos, devido às suas características não estatísticas.

Além das suposições habituais do modelo fatorial, expostas no Capítulo II, suporemos que os vetores aleatórios  $f$  e  $e$  têm distribuição normal multivariada com vetores média zero e com matrizes de covariância  $\Phi$  e  $\Psi$  respectivamente (quando os fatores são não correlacionados,  $\Phi = I_m$ ). Da suposição de normalidade e das suposições feitas no Capítulo II segue que os fatores  $f$  e  $e$  são mutuamente independentes. Como  $X$  é expresso em termos de  $f$  e  $e$  conforme a equação (2.1.1), temos que o vetor das variáveis observáveis é também normal com média zero e matriz de covariância  $\Sigma$ .

Descreveremos, a seguir, o método de máxima verossimilhança para obter estimadores dos parâmetros  $\Lambda$  e  $\Psi$ , quando os fatores são supostos ortogonais e dos parâmetros  $\Lambda$ ,  $\Psi$  e  $\Phi$  quando os fatores são supostos oblíquos. Supomos que o número  $m$ , de fatores comuns, é conhecido e supomos também que estes parâmetros possam ser definidos univocamente, de acordo com o que foi exposto no Capítulo II. Vimos que o problema da unicidade dos parâmetros, no modelo fatorial, é resolvido após a imposição de certas restrições. Estas restrições podem ser apenas por conveniências essencialmente matemáticas, quando, então, o modelo é denominado *Modelo Não Restrito*, segundo a nomenclatura de Lawley & Maxwell (1971), ou

podem ser restrições que além de especificar o número de fatores, especificam os valores de certos elementos em  $\Lambda$ ,  $\Phi$  e até mesmo em  $\Psi$  de acordo com algum significado intrínscico aos problemas teóricos da área em estudo. Um modelo deste segundo tipo é denominado *Modelo Restrito*. Veremos a seguir o problema estimação para estes dois tipos de modelos.

O método de estimação de máxima verossimilhança para o Modelo Não-Restrito é apresentado apenas no caso de fatores ortogonais, isto é, quando os fatores são supostos não correlacionados, enquanto que para o Modelo Restrito, os fatores, quando  $m > 1$ , podem ser correlacionados.

Pelo método de máxima verossimilhança os estimadores dos parâmetros do modelo fatorial são expressos como soluções de equações matriciais implícitas que podem ser resolvidos apenas por processos iterativos. A origem destas equações é o que apresentaremos a seguir

### 3.1.1 - MODELO NÃO RESTRITO

Consideremos o modelo fatorial, no qual os fatores são não correlacionados, com as suposições feitas na seção 2.2, além das suposições de normalidade discutidas acima,

$$X = \Lambda f + e$$

$$\Sigma = \Lambda \Lambda' + \Psi$$

Supomos, também, que os parâmetros são definidos univocamente, isto é, que exista um único  $\Psi$ , com elementos diagonais positivos e um único  $\Lambda$  tal que  $\Lambda' \Psi^{-1} \Lambda$  é uma matriz diagonal, cujos elementos diagonais são positivos, distintos e arranjados em ordem decrescente de grandeza. Nesta seção nós mostraremos o problema de estimar  $\Lambda$  e  $\Psi$ , a partir da matriz

de covariância amostral.

Supomos que uma amostra aleatória de  $N(>p)$  observações de  $X$  é obtida. Sejam  $X_1, X_2, \dots, X_N$  os vetores colunas que representam estas  $N$  observações. Seja  $S = (s_{ij})$  a matriz de covariância amostral, definida por

$$S = \frac{1}{n} \sum_{k=1}^N (X_k - \bar{X})(X_k - \bar{X})'$$

onde

$$n = N-1 \quad \text{e} \quad \bar{X} = \frac{1}{N} \sum_{k=1}^N X_k, \text{ tal que } E(S) = \Sigma.$$

Como  $X$  é  $N(0, \Sigma)$ , os elementos de  $nS$  seguem a distribuição de Wishart com  $n$  graus de liberdade (Anderson, 1958). A função densidade desta distribuição é dada por

$$L = C |S|^{\frac{1}{2}(n-p-1)} |\Sigma|^{-\frac{1}{2}n} \exp(-1/2 \text{tr} \Sigma^{-1} S)$$

onde  $C$  é uma constante.

De acordo com o modelo fatorial, toda a informação da amostra está contida em  $S$ , portanto a função de verossimilhança da amostra é dada por  $L$ , definida acima.

O logaritmo da função de verossimilhança é então da do por

$$L^* = \log_e L = -\frac{1}{2} n [\log_e |\Sigma| + \text{tr}(\Sigma^{-1} S)]$$

onde se omitiu uma função das observações. Como  $\Sigma = \Lambda \Lambda' + \Psi$ , te mos

$$L^* = -\frac{1}{2} n [\log_e |\Lambda \Lambda' + \Psi| + \text{tr}(\Lambda \Lambda' + \Psi)^{-1} S].$$

Lawley e Maxwell (1971) sugerem que é mais conveniente minimizar a função

$$F(\Lambda, \Psi) = \log_e |\Sigma| + \text{tr}(\Sigma^{-1}S) - \log_e |S| - p$$

o que é equivalente a maximizar  $L$ , pois  $\log_e L$  é  $-\frac{1}{2}n$  vezes  $F$  mais uma função das observações apenas.

Então, derivando a função  $F$  em relação aos elementos de  $\Lambda$  e  $\Psi$  por meio de processos comuns de derivação de matrizes e igualando a zero, obtemos:

$$\frac{\partial F}{\partial \Lambda} = 2\Sigma^{-1}(\Sigma - S)\Sigma^{-1}\Lambda = 0$$

$$\frac{\partial F}{\partial \Psi} = \text{diag}[\Sigma^{-1}(\Sigma - S)\Sigma^{-1}] = 0$$

As equações do sistema acima podem ser simplificadas. Para a primeira temos:

$$(\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1})\Lambda = 0$$

ou

$$S\Sigma^{-1}\Lambda = \Lambda \quad (3.1.1)$$

Apesar da expressão obtida ser muito simples ela envolve a inversão de uma matriz  $p \times p$ . A fim de evitar este problema, podemos transformá-la em uma outra que envolve apenas a inversão de uma matriz diagonal. Para isto usamos um artifício, que consiste em pós-multiplicar ambos os membros da expressão  $\Sigma = \Lambda\Lambda' + \Psi$  por  $\Psi^{-1}\Lambda$ . Assim obtemos:

$$\Sigma\Psi^{-1}\Lambda = \Lambda(\Lambda'\Psi^{-1}\Lambda + I)$$

$$\Sigma \psi^{-1} \Lambda (\Lambda' \psi^{-1} \Lambda + I)^{-1} = \Lambda$$

ou

$$\Sigma^{-1} \Lambda = \psi^{-1} \Lambda (\Lambda' \psi^{-1} \Lambda + I)^{-1} \quad (3.1.2)$$

Substituindo este resultado na expressão 3.1.1, obtemos

$$S \psi^{-1} \Lambda (\Lambda' \psi^{-1} \Lambda + I)^{-1} = \Lambda$$

Finalmente

$$S \psi^{-1} \Lambda = \Lambda (\Lambda' \psi^{-1} \Lambda + I)$$

Para a segunda equação do sistema, temos

$$\text{diag}(\Sigma^{-1}) = \text{diag}(\Sigma^{-1} S \Sigma^{-1});$$

pré e pós multiplicando as matrizes de ambos os membros pela matriz diagonal  $(\Sigma^{-1} \Lambda \Lambda')$ , obtemos

$$\text{diag}(\Sigma^{-1} \Lambda \Lambda' \Sigma^{-1} \Lambda \Lambda') = \text{diag}(S - S \Sigma^{-1} \Lambda \Lambda' - \Lambda \Lambda' \Sigma^{-1} S + \Lambda \Lambda' \Sigma^{-1} S \Sigma^{-1} \Lambda \Lambda')$$

Usando (3.1.1) temos que

$$\text{diag}(\Sigma) = \text{diag}(S)$$

$$\text{diag}(\Lambda \Lambda' + \psi) = \text{diag } S$$

Assim, as equações de máxima verossimilhança na forma simplificada são dadas por:

$$S \psi^{-1} \Lambda = \Lambda (I + \Lambda' \psi^{-1} \Lambda) \quad (3.1.2)$$

$$\text{diag}(\Lambda\Lambda' + \psi) = \text{diag } S \quad (3.1.3)$$

Estas equações podem ser resolvidas apenas por processos iterativos para então se obter  $\hat{\Lambda}$  e  $\hat{\psi}$ .

Lawley & Maxwell dizem que para achar um mínimo da função  $F$  é conveniente usar um procedimento em dois estágios. Primeiro encontra-se um mínimo condicional de  $F$  para um dado  $\psi$  que é denotado por  $f(\psi)$  e então se obtém  $\psi$  que minimiza esta  $f(\psi)$ . É possível, sob certas condições, obter um  $\Lambda$  tal que  $\frac{\partial F}{\partial \Lambda} = 0$  e tal que  $\Lambda'\psi^{-1}\Lambda$  é diagonal. Assim além de minimizar  $F$  para dado  $\psi$ , este processo satisfaz as condições impostas sobre o verdadeiro parâmetro  $\Lambda$  para sua unicidade.

Rao & Maxwell sugeriram um método para a solução numérica das equações de máxima verossimilhança que parece ter boas propriedades de convergência. Este processo iterativo está exposto em Morrison (1967).

O método computacional iterativo desenvolvido por Jöreskog para resolver as equações de máxima verossimilhança converge rapidamente e os valores dos parâmetros podem ser obtidos tão precisamente quanto se queira. Este método e um programa de computador em linguagem FORTRAN IV foi publicado por Jöreskog (1966). O programa é denominado UMLFA-"Unrestricted Maximum Likelihood Factor Analysis". Este programa pode ser obtido no "University of Chicago Computation Center", através do código de programa UCSM718.

Uma característica satisfatória do método de máxima verossimilhança para Análise Fatorial é que, quando a variável é multiplicada por uma constante os estimadores das cargas fatoriais para esta variável são multiplicados pela mesma constante e os estimadores de variância específica são

multiplicados pelo quadrado da constante (Anderson & Rubin, 1956). Assim ao realizarmos os cálculos numéricos, podemos padronizar todas as variáveis  $X_i$  e substituir  $S$  pela matriz de correlação amostral.

Antes de verificarmos a propriedade acima, vamos ver o que acontece com o modelo, quando a variável  $X$  é multiplicada por uma matriz diagonal  $D$ , cujos elementos diagonais são constantes não necessariamente iguais e diferentes de zero. Seja

$$DX = Z$$

Assim,

$$EZZ' = DED = D\Lambda(D\Lambda)' + D\psi D = \Sigma_Z$$

Usando a seguinte notação,

$$\Sigma_Z = \Lambda_Z \Lambda_Z' + \psi_Z$$

temos

$$\psi_Z = D\psi D \text{ e } \Lambda_Z \Lambda_Z' = D\Lambda(D\Lambda)',$$

onde  $\Sigma_Z$  é a matriz de covariância,  $\Lambda_Z$  a matriz de cargas fatoriais e  $\psi_Z$  a matriz das variâncias específicas para as variáveis transformadas. Com relação à unicidade dos parâmetros verificamos que a restrição imposta à matriz  $\Lambda$  quando a variável é  $X$  implica a mesma restrição quando a variável é  $Z$  pois

$$(D\Lambda)' (D\psi D)^{-1} D\Lambda = \Lambda' \psi^{-1} \Lambda.$$

Vejamos agora como o método de estimação é afetado

quando transformamos as unidades de medida das variáveis originais. Se  $DX = Z$ , a matriz de covariância amostral é dada por DSD e as equações para obtenção dos estimadores de máxima verossimilhança são, então, dadas por

$$\begin{aligned} DSD\psi_z^{-1}\Lambda_z &= \Lambda_z (I + \Lambda_z'\psi_z^{-1}\Lambda_z) \\ \text{diag}(\psi_z + \Lambda_z\Lambda_z') &= \text{diag}(DSD) \end{aligned}$$

Claramente  $\hat{\Lambda}_z = D\hat{\Lambda}$  e  $\hat{\psi}_z = D\hat{\psi}D$  são as soluções deste sistema quando  $\hat{\Lambda}$  e  $\hat{\psi}$  são as soluções de (3.1.2) e (3.1.3). Esta propriedade, verificada acima, salienta a vantagem do método de máxima verossimilhança para Análise Fatorial.

Outro comentário a fazer é sobre os elementos da matriz diagonal  $\Lambda'\psi^{-1}\Lambda$ , a qual nos referimos frequentemente. Se as variáveis são transformadas de maneira que sua variância específica se torne unitária, então

$$\frac{\lambda_{ij}}{\psi_i}$$

representa a parte da variância de  $X_i$  devida ao  $j$ -ésimo fator. Considerando todas as  $p$  variáveis, a variância total em  $X$ , devida ao  $j$ -ésimo fator é então dada por

$$\sum_i \frac{\lambda_{ij}^2}{\psi_i},$$

que é o  $j$ -ésimo elemento diagonal de  $\Lambda'\psi^{-1}\Lambda$ . Fazer a restrição de que  $\Lambda'\psi^{-1}\Lambda$  seja diagonal e com os elementos diagonais arranjados em ordem decrescente de grandeza significa então, que nossa escolha dos fatores é tal que o primeiro fator tem uma contribuição máxima à variância de  $X$ , o segundo tem uma contribuição máxima sujeito a ser não correlacionado com o primeiro e assim por diante.

Finalmente, faremos um rápido comentário sobre a

distribuição assintótica dos estimadores de máxima verossimilhança. Supondo, como antes, que as observações são extraídas de uma população  $N(0, \Sigma)$ , Anderson & Rubin (1956) mostram que  $\sqrt{n}(\hat{\Lambda} - \Lambda)$  e  $\sqrt{n}(\hat{\Psi} - \Psi)$ , onde  $\hat{\Lambda}$  e  $\hat{\Psi}$  são os estimadores de máxima verossimilhança do modelo não-restrito, são assintoticamente, normalmente distribuídos. Expressões para as covariâncias e variância assintóticas dos estimadores podem ser encontradas em Lawley & Maxwell (1971).

### 3.1.2 - MODELO RESTRITO

Nesta secção nós consideraremos o método de máxima verossimilhança para estimar os parâmetros do modelo fatorial, quando os fatores podem ou não ser correlacionados ( $m > 1$ ). Os parâmetros a serem estimados são então  $\Lambda$ ,  $\psi$  e  $\Phi$ , onde  $\Phi$  é a matriz de covariância dos fatores. As condições que determinam os parâmetros univocamente foram comentadas na secção 2.2 e são do tipo que definem o Modelo Restrito considerado no início deste capítulo.

Como vimos, o conjunto de restrições que devem ser feitas sobre os elementos de  $\Lambda$ , para obter unicidade, consiste em requerer que certos elementos desta matriz sejam nulos. Thurstone (1947) foi originalmente quem postulou uma configuração de zeros para a matriz  $\Lambda$ , através do Princípio da Estrutura Simples, composto por cinco regras. Estas regras, no entanto, não especificavam precisamente quais cargas fatoriais deveriam ser nulas, o que implicava em certa indeterminação, inconveniente quando procedimentos de estimação estão sendo desenvolvidos.

Anderson & Rubin (1956) abordam o problema mais objetivamente, formulando condições denominadas: "Elementos zeros em posições especificadas". Estas condições requerem alguma informação a priori, como por exemplo, que algumas va-

riáveis não dependem de alguns fatores. Assim, as condições são  $\lambda_{ij} = 0$  para certos pares  $(i,j)$ , ou seja que o  $j$ -ésimo fator não afeta a  $i$ -ésima variável. Os coeficientes da  $j$ -ésima coluna de  $\Lambda$  são únicos, exceto pela multiplicação de um fator escala, se (a) existe ao menos  $m-1$  elementos zero e (b) o posto de  $\Lambda^{(j)}$  é  $m-1$ , onde  $\Lambda^{(j)}$  é uma submatriz de  $\Lambda$  que tem elementos zero na  $j$ -ésima coluna.

Lawley & Maxwell (1971) apresentam outros métodos para especificar a posição exata de zeros em  $\Lambda$  e formulam também condições para definir os parâmetros univocamente. A técnica mais recente para a estimação por máxima verossimilhança no modelo restrito está exposta na obra acima citada, mas a abordagem é no sentido computacional do método, apresentando o processo iterativo. Como nosso intuito é apenas apresentar as equações para obtenção dos estimadores, seguiremos, então, Anderson & Rubin.

Sejam  $X_1, X_2, \dots, X_N$ ,  $N$  observações da  $N(0, \Sigma)$ , onde  $\Sigma = \Lambda\Phi\Lambda' + \Psi$ . A matriz  $\Phi$  não é necessariamente diagonal com a restrição que seus elementos diagonais sejam unitários. Seja

$$\lambda_{ij} = 0 \quad \begin{array}{l} i=i(1,j), \dots, i(p_j,j) \\ j=1, \dots, m, \end{array}$$

isto é, na  $j$ -ésima coluna de  $\Lambda$ , existem  $p_j$  elementos nulos e eles estão nas linhas numeradas por  $i(1,j), \dots, i(p_j,j)$ .

Seja ainda  $\Gamma$  uma matriz  $p \times m$  que tem elementos zero, onde  $\Lambda$  não tem, isto é

$$\gamma_{ij} = 0 \quad \begin{array}{l} i \neq i(1,j), \dots, i(p_j,j) \\ j=1, \dots, m. \end{array}$$

Seja, como antes,  $S$ , a matriz de covariância amostral. O logaritmo da função de verossimilhança, como na seção 3.1.1 é dado por

$$\log_e L = -\frac{1}{2}n[\log_e |\Sigma| + \text{tr}\Sigma^{-1}S].$$

Agora, entretanto,  $\Sigma = \Lambda\phi\Lambda' + \psi$ . Devemos procurar valores  $\hat{\Lambda}$ ,  $\hat{\phi}$  e  $\hat{\psi}$  que maximizam a função  $G = \log_e L$ . Deriva-se então a função  $G$  em relação apenas aos parâmetros livres de  $\Lambda$ ,  $\psi$  e  $\phi$ , isto é, em relação aos elementos não restritos.

O sistema de equações que se obtém, igualando a zero estas derivadas, é simplificado de maneira análoga ao que foi feito na secção anterior. Os detalhes deste procedimento são encontrados em Anderson & Rubin (1956, pág. 140).

As equações para obtenção dos estimadores de máxima verossimilhança no modelo fatorial restrito são dados por

$$\text{diag}(S - \Lambda\phi\Lambda' - \psi) = 0$$

$$\Gamma'\Lambda = 0$$

$$(\phi^{-1} + \Lambda'\psi^{-1}\Lambda)\Gamma'\psi = \Lambda'\psi^{-1}S - \Lambda'\psi^{-1}\Lambda\phi\Lambda' - \Lambda'.$$

Estas equações não podem ser resolvidas algébricamente; vários processos iterativos tem sido propostos para obtenção dos estimadores e trabalhos recentes têm mostrado que estes procedimentos nem sempre convergem.

Jöreskog e Lawley sugerem um método, usando as derivadas de segunda ordem da função de verossimilhança, que tem boas propriedades de convergência. Este método esta exposto em Lawley & Maxwell (1971) e o programa do computador para o método é devido a Jöreskog e Gruvaneus (1967). O programa é denominado "RMLFA - Restricted Maximum Likelihood Factor Analysis". Neste método, que é muito geral, nenhuma distinção essencial é feita entre os modelos ortogonal e oblíquo. É possível, inclusive, ter casos nos quais um grupo de fatores podem ser correlacionados, enquanto que os fato-

res restantes são supostos não correlacionados com este grupo. O programa foi escrito em linguagem FORTRAN IV e permite que quaisquer valores possam ser especificados em  $\Lambda$ ,  $\phi$  e  $\psi$ . Este programa pode ser obtido através do código UCSM 722 do "University of Chicago Computation Center", Chicago.

No modelo restrito, quando  $\Lambda$  é identificado por elementos zeros especificados. Anderson & Rubin comentam que os estimadores obtidos pelo método de máxima verossimilhança são também assintoticamente normalmente distribuídos.

### 3.2 - TESTES DA RAZÃO DE VEROSSIMILHANÇA NO MODELO FATORIAL

Nas secções anteriores nós discutimos o problema da estimação dos parâmetros, supondo que o modelo é adequado aos dados em estudo, em particular nós supomos que o número  $m$  de fatores comuns é conhecido. A principal vantagem de usar o método de estimação de máxima verossimilhança é que se pode testar, através do critério da razão de verossimilhança apropriado, as hipóteses de que  $\Sigma$ , a matriz de covariância da população pode ser escrita como  $\Sigma = \Lambda\Lambda' + \Psi$  quando os  $m$  fatores são supostos não correlacionados, ou como  $\Sigma = \Lambda\phi\Lambda' + \psi$ , quando os fatores são correlacionados, onde  $\Lambda$  tem  $m$  colunas. Ou seja, testa-se a hipótese de que existam  $m$  fatores comuns para  $m$  especificado. Há outras suposições no modelo, tal como normalidade, linearidade de efeito de fatores, etc. que podem ser questionadas, mas não as consideraremos aqui. Vejamos os testes da razão de verossimilhança no Modelo Não Restrito e no Modelo Restrito.

#### 3.2.1 - MODELO NÃO RESTRITO

Sobre as suposições feitas para o modelo fatorial não restrito, e para um determinado valor de  $m$  seja  $H_m: \Sigma = \Lambda\Lambda' + \psi$ ,

a hipótese nula a ser testada. O critério da razão de verossimilhança apropriado para esta hipótese é devido a Lawley.

Seja  $\Omega$  um conjunto de todas as matrizes  $\Sigma$ , de ordem  $p$ , simétricas e positivas definidas. Seja  $\omega = \{\Sigma \in \Omega / \Sigma = \Lambda \Lambda' + \psi\}$ . Sejam  $L_\Omega$  e  $L_\omega$  os máximos da função de verossimilhança  $L$  em  $\Omega$  e  $\omega$  respectivamente. Usando os resultados de Inferência Estatística [Anderson (1968)], temos que  $L$  atinge seu máximo, em  $\Omega$ , quando  $\Sigma = S$ . Na secção 3.1.1 vimos que

$$\log_e L = -\frac{1}{2} n [\log |\Sigma| + \text{tr}(\Sigma^{-1} S)]$$

portanto

$$\log_e L_\Omega = -\frac{1}{2} n (\log |S| + p).$$

Considerando os resultados anteriores temos que  $L$  atinge seu valor máximo em  $\omega$  quando  $\Sigma$  é igual a  $\hat{\Sigma} = \hat{\Lambda} \hat{\Lambda}' + \hat{\psi}$ , onde  $\hat{\Lambda}$  e  $\hat{\psi}$  são os estimadores de máxima verossimilhança obtido pelo método descrito anteriormente. Assim

$$\log_e L_\omega = -\frac{1}{2} n [\log |\hat{\Sigma}| + \text{tr}(\hat{\Sigma}^{-1} S)].$$

A razão de verossimilhança é  $\lambda = \frac{L_\omega}{L_\Omega}$  mas sua distribuição exata não pode ser determinada. Sabe-se, entretanto, que para grandes amostras,  $-2 \log_e \lambda$  tem distribuição aproximadamente  $\chi^2$ , se  $H_m$  é verdadeira. Então, temos

$$\begin{aligned} -2 \log_e \lambda &= -2 \log_e L_\omega + 2 \log_e L_\Omega = \\ &= n [\log |\hat{\Sigma}| + \text{tr}(S \hat{\Sigma}^{-1}) - \log |S| - p]. \end{aligned} \quad (3.2.1)$$

Se  $\hat{\Sigma}$  é obtido com muita precisão pelas equações de máxima verossimilhança para o modelo não restrito, através de processos iterativos, então, temos que

$$\text{tr}(S\hat{\Sigma}^{-1}) = \text{tr}(I_p) = p.$$

Assim a expressão (3.2.1) pode ser simplificada

$$-2 \log_e \lambda = n[\log|\hat{\Sigma}| - \log|S|].$$

Observamos, no entanto, que se o algoritmo de cálculo para se obter  $\hat{\Sigma}$ , não converge rapidamente a uma solução precisa, a expressão (3.2.1) deve ser usada.

Bartlett (1962) introduziu uma modificação na expressão (3.2.1) obtida acima para que a sua convergência à distribuição limite,  $\chi^2$ , fosse mais rápida. Ele sugeriu que ao invés de  $n$  fosse usado o fator multiplicativo

$$n - \frac{2p+5}{6} - \frac{2}{3}m.$$

Assim, o critério a ser usado para testar  $H_m$  é:

$$U_m = (n - \frac{2p+5}{6} - \frac{2}{3}m)[\log|\hat{\Sigma}| - \log|S|]$$

Sob  $H_m$ ,  $U_m$  tem distribuição assintótica  $\chi_r^2$ , onde

$$r = \frac{1}{2}[(p-m)^2 - (p+m)].$$

Observamos que o número de graus de liberdade  $r$  é obtido pela diferença entre o número de parâmetros livres em  $\Omega$  e o número de parâmetros livres em  $\omega$ . Em  $\Omega$  temos  $p(p+1)/2$  parâmetros livres; em  $\omega$  temos  $mp$  parâmetros em  $\Lambda$  e  $p$  parâmetros em  $\Psi$  a serem estimados, mas a condição de que  $\Lambda'\psi^{-1}\Lambda$  seja uma matriz diagonal impõe  $m(m-1)/2$  restrições a mais para o problema, então, em  $\omega$ ,  $\Sigma$  tem  $mp+p-m(m-1)/2$  parâmetros livres.

Desta maneira o teste-critério usado rejeita a hipótese nula, ao nível de significância  $\alpha$ , se  $U_m \geq \chi_{\alpha, r}^2$ .

Da propriedade de invariância dos estimadores de  $\underline{m\bar{a}}$

xima verossimilhança segue que o mesmo valor do teste estatístico seria obtido para uma solução fatorial em termos da matriz de correlação.

Lawley & Maxwell (1971) sugerem ainda outros critérios para se obter uma aproximação ao teste da razão de verossimilhança exposto nesta secção. Um destes critérios é

$$T_m = (n - \frac{1}{6}(2p+5) - \frac{2}{3}m) \sum_{i < j} \sum (S_{ij} - \hat{\sigma}_{ij})^2 / \hat{\psi}_i \hat{\psi}_j$$

a qual tem distribuição assintótica  $\chi_r^2$  e onde

$$\hat{\sigma}_{ij} = \sum_{h=1}^m \hat{\lambda}_{ih} \hat{\lambda}_{jh}, \quad i \neq j.$$

Outro critério baseado nas raízes características da matriz  $\hat{\psi}^{-1/2} S \hat{\psi}^{-1/2}$  e a estatística para este critério é dada por

$$T_m^* = \frac{1}{2} [(\hat{\theta}_{m+1} - 1)^2 + \dots + (\hat{\theta}_p - 1)^2]$$

onde  $\hat{\theta}_{m+1}, \dots, \hat{\theta}_p$  são as  $p-m$  menores raízes características da matriz citada. A distribuição assintótica deste critério é também  $\chi_r^2$ .

### 3.2.2 - MODELO RESTRITO

Consideremos agora o modelo fatorial com as suposições feitas para o modelo restrito, onde os fatores podem ser correlacionados. A hipótese a ser testada neste caso é

$$H_m: \Sigma = \Lambda \Phi \Lambda' + \psi.$$

O critério da razão de verossimilhança pode ser obtido como anteriormente e temos

$$-2 \log_e \lambda = n[\log|\hat{\Sigma}| + \text{tr}(S\hat{\Sigma}^{-1}) - \log|S| - p]$$

mas agora  $\hat{\Sigma} = \hat{\Lambda}\hat{\phi}\hat{\Lambda}' + \hat{\psi}$ , onde os elementos não restritos de  $\Lambda, \phi$  e  $\psi$  são estimados pelo método de máxima verossimilhança, enquanto que os elementos restritos tomam seus valores especificados. Sob a hipótese  $H_m$ , quando  $n$  é grande,  $-2 \log_e \lambda$  tem distribuição aproximadamente  $\chi^2$  com  $d$  graus de liberdade, onde

$$\begin{aligned} d &= p(p+1)/2 - (p+pm+m(m+1)/2-\ell) = \\ &= p^2 - \frac{1}{2}(p+m)(p+m+1) + \ell, \end{aligned}$$

sendo  $\ell$  o número de parâmetros fixos. Uma melhor aproximação à distribuição  $\chi^2$  é obtida se na expressão de  $-2 \log_e \lambda$ ,  $n$  é substituído por um fator multiplicativo conveniente. Lawley sugere usar o fator  $n-(2p+5)/6$ .

Os processos iterativos para o método de máxima verossimilhança no modelo restrito não tem propriedades de convergência tão boas quanto para o modelo não restrito. Este fato não permite a simplificação na estatística do teste que foi feita no modelo não restrito.

Portanto, o critério para testar a hipótese  $H_m$  usa a estatística

$$U_m^* = [n-(2p+5)/6](\log|\hat{\Sigma}| + \text{tr}(S\hat{\Sigma}^{-1}) - \log|S| - p)$$

a qual tem distribuição  $\chi_d^2$ , onde o número de graus de liberdade é

$$d = p^2 - \frac{1}{2}(p+m)(p+m+1) + \ell.$$

Como vimos, as hipóteses estabelecidas acima espe-

pecificam o número de fatores no modelo. Frequentemente, o pesquisador não conhece este número e às vezes ele nem mesmo está em condições de postular um valor específico  $m$ . Na prática, o problema não é usualmente o de testar uma hipótese particular, mas sim o de determinar o número de fatores tal que o modelo correspondente se ajuste adequadamente aos dados.

Uma das maneiras de resolver este problema é através de procedimentos sequenciais. Assim, parte-se de algum valor pequeno  $m_0$ , que pode ser 1, e então se estima os parâmetros com  $m = m_0$ . Testa-se a hipótese nula para  $m = m_0$  e se a hipótese é rejeitada testa-se a hipótese com  $m = m_0 + 1$  e assim por diante, com  $m$  crescendo uma unidade em cada etapa, até que a hipótese não seja rejeitada para algum  $m$ .

Usar uma sequência de testes de razão de verossimilhança é computacionalmente difícil pois em cada estágio de ve-se calcular  $\hat{\lambda}$  e  $\hat{\psi}$  (e  $\hat{\phi}$ ) para um valor de  $m$ . Além disso, os critérios  $U_m$  obtidos desta forma não são independentes, e o nível de significância verdadeiro do teste pode ser muito diferente do valor nominal usado em cada estágio do procedimento sequencial. Infelizmente não há procedimento para o qual exista uma teoria estatística adequada, nem mesmo uma teoria assintótica para esse problema.

### 3.3 - OUTROS MÉTODOS DE ESTIMAÇÃO

Antes de ter sido desenvolvido o método de máxima verossimilhança para Análise Fatorial, o problema da determinação das cargas fatoriais e variâncias únicas era resolvido por vários outros métodos, todos matemáticos em essência, sem características estatísticas. Entre estes métodos estão: Solução Bi-Fatorial, Solução do Fator Principal, Método Centróide, Solução de Grupos Múltiplos e outros. Estes

métodos ou soluções fatoriais, como comenta Harman (1960) podiam ser distinguidos com base na contribuição total de um fator  $f_j$  para as variâncias de todas as variáveis  $X_i$ . Esta contribuição total de  $f_j$  é definida como

$$V_j = \sum_{i=1}^p \lambda_{ij}^2.$$

Algumas soluções forneciam fatores com contribuições decrescentes em relação à communalidade total, outras forneciam fatores com contribuições aproximadamente iguais à variância total de  $X$ .

Nesta secção exporemos a Solução do Fator Principal por ser ainda muito utilizada e apenas faremos comentários sobre o Método Centróide.

#### - SOLUÇÃO DO FATOR PRINCIPAL

A Solução do Fator Principal ou Método dos Fatores Principais originou-se no trabalho de Hotelling e é extremamente relacionada com a Análise de Componentes Principais apresentada no Capítulo II.

Do ponto de vista algébrico, a seleção dos fatores, por este método, é equivalente a escolher um conjunto de fatores em ordem decrescente de sua contribuição à communalidade total das variáveis. A análise começa com um fator  $f_1$  cuja contribuição às communalidades das variáveis é a maior possível; um segundo fator  $f_2$ , não correlacionado com  $f_1$ , com uma contribuição máxima à communalidade residual é encontrado. E o processo continua até que a communalidade total tenha sido analisada.

Na prática, as communalidades verdadeiras ou o seu complemento, as variâncias específicas de cada variável, não

são conhecidas e elas devem ser estimadas a partir dos dados amostrais. Vários métodos para estimar as comunalidades tem sido propostos, mas nenhum tem se mostrado superior a outro. Os estimadores de comunalidade de cada variável  $X_i$  mais comuns são:

- a) o quadrado do coeficiente de correlação múltipla entre a variável  $X_i$  e todas as outras
- b) o maior valor absoluto da linha  $i$  da matriz de correlação amostral
- c) estimadores obtidos de análises preliminares por processos iterativos.

Após obter-se os estimadores das comunalidades de cada variável, substituímos a diagonal principal da matriz de correlação amostral pelas comunalidades estimadas. A matriz assim obtida é denominada matriz de correlação reduzida, e será denotada por  $R$ .

A partir da matriz de correlação reduzida, aplica-se o método de componentes principais, como vimos no Capítulo II. Escolhe-se então as  $m$  primeiras maiores raízes características de  $R$  e os  $m$  vetores características correspondentes, obtendo-se, então, a matriz das cargas fatoriais estimadas pela Solução dos Fatores Principais e que é dada por:

$$\hat{A} = \hat{\Gamma} \hat{\Delta}^{1/2}$$

onde  $\hat{\Delta}$  é a matriz diagonal de ordem  $m$ , cujos elementos diagonais são as  $m$  primeiras raízes características e  $\hat{\Gamma}$  é a matriz  $p \times m$ , cujas colunas são os vetores característicos correspondentes.

Em algumas aplicações, quando as comunalidades estimadas não diferem muito da unidade, a matriz de correlação não alterada, isto é, com unidades na diagonal, pode ser

tomada como ponto de partida da análise.

Se as comunalidades verdadeiras fossem inseridas na diagonal da matriz de correlação, o posto da matriz de correlação alterada deveria ser reduzido, isto é, deveria ser menor do que  $p$ , quando as suposições do modelo fatorial são válidas. Idealmente, então, a questão do número de fatores a serem extraídos poderia ser resolvido examinando-se o posto da matriz. Entretanto, erros aleatórios de medida e a ausência de um processo de estimação das comunalidades conveniente, fazem com que não possamos determinar o número  $m$  desta forma.

Para decidir o número  $m$  de fatores comuns, vários critérios são sugeridos na literatura. Os mais comuns são: (Afifi e Azen, 1972).

- a) A estrutura física pode sugerir um determinado número de fatores comuns intrínsecos.
- b) Quando a matriz de correlação não alterada é usada, retêm-se apenas os fatores correspondentes às raízes características da matriz que são maiores do que a unidade.
- c) Escolhe-se o número de fatores os quais explicam uma proporção especificada da comunalidade total ou da variância total.

O critério b) está de acordo com considerações de significância estatística e fideignidade psicométrica entre outras coisas, mas não é imune, aparentemente, ao criticismo. Este critério é devido a Kaiser.

Existem programas de computador para o Método do Fator Principal, como por exemplo o sub-programa "FACTOR" do SPSS. Este programa assim como outros, oferecem ao usuário várias opções e abrangem também outros métodos para Análise Fatorial.

Algumas vezes uma solução pelo Método dos Fatores Principais serve como a análise final, podendo os fatores serem interpretados adequadamente. Outras vezes, será desejável transformar tal solução a um sistema de referência diferente (ou ortogonal, ou oblíquo) por meio dos processos de rotação que veremos no próximo Capítulo. A solução dos fatores principais fornece sempre, inicialmente, fatores ortogonais entre si. Fatores oblíquos serão obtidos só após algum método de transformação<sup>ortogonal</sup>.

#### - MÉTODO CENTRÓIDE

O Método Centróide ou Método da Soma Simples para obter as cargas fatoriais era inicialmente muito usado porque os cálculos requeridos eram relativamente simples. Nós não discutiremos o método em detalhes e ele pode ser encontrado em Thurstone (1947) ou Harman (1960). Este método utiliza uma matriz  $Q$ ,  $p \times m$ , de posto  $m$ , com elementos 1 ou -1, e os resultados obtidos não diferem muito dos obtidos pela Solução do Fator Principal. Com as facilidades computacionais atuais não há justificativa para se voltar às aproximações como as dadas pela Solução Centróide e o seu valor, tem, então, se tornado apenas histórico.

#### 3.4 - ESTIMAÇÃO DOS ESCORES FATORIAIS

Em muitas aplicações, principalmente quando Análise Fatorial é preliminar a algum outro tipo de análise multivariada, ou quando o seu uso principal é para construção de índices, é conveniente, além de estimar os parâmetros do modelo fatorial, procurar descrever os fatores em termos das variáveis observadas. Para isto estimam-se os valores de cada fator para cada indivíduo. Estes valores são denominados

escores fatoriais.

Os escores fatoriais não são parâmetros do modelo, são valores atribuídos às variáveis hipotéticas e por isto não podem ser estimados no sentido estatístico usual.

Em Análise de Componentes Principais, as componentes eram definidas como funções lineares das variáveis observadas e então, os valores de cada componente para cada indivíduo podiam ser facilmente encontrados. Em Análise Fatorial, os fatores não são funções lineares das variáveis observadas apenas e os escores de um indivíduo sobre eles não podem ser encontrados da mesma maneira. É necessário, então, introduzir um princípio de mínimos quadrados para se obter razoáveis estimadores dos escores fatoriais.

Dois métodos para a estimação dos escores fatoriais serão apresentados aqui: o Método de Regressão, devido a Thomson (1951) e o Método de Bartlett. Descreveremos os métodos supondo que os fatores são correlacionados, por ser mais geral. No caso de fatores não correlacionados e padronizados, os resultados são obtidos do mesmo modo, considerando-se  $\Phi = I$ .

Para qualquer dos dois métodos supomos que os parâmetros do modelo fatorial,  $\Lambda$ ,  $\psi$  e  $\Phi$  foram estimados por algum dos procedimentos anteriormente discutidos, desprezando-se erros de amostragem

#### 1) MÉTODO DE REGRESSÃO (MÉTODO DE THOMSON)

Seja  $X = (X_1, X_2, \dots, X_p)'$  o vetor das observações. Seja  $f = (f_1, f_2, \dots, f_m)'$  o vetor dos escores fatoriais e seja  $e = (e_1, \dots, e_p)'$  o vetor dos resíduos. As suposições do modelo fatorial discutidas no Capítulo II são mantidas. Assim temos:

$$E(ff') = \phi$$

$$E(Xf') = E[(\Lambda f + e)f'] = \Lambda E(ff') = \Lambda \phi$$

$$E(XX') = \Sigma = \Lambda \phi \Lambda' + \Psi.$$

onde  $\Lambda$ ,  $\psi$ ,  $\phi$  são constantes por terem sido estimados conforme afirmação anterior.

O método de Thomson, baseado na regressão de  $f$  sobre  $X$  é equivalente, segundo Lawley e Maxwell (1971) a encontrar para cada  $j$ ,  $j=1, \dots, m$ , uma função linear das observações que dará um bom preditor de  $f_j$  dado por

$$\hat{f}_j = a_j' X = X' a_j$$

onde  $a_j$  é um vetor de ordem  $p$ , escolhido de tal maneira que a variância de  $(\hat{f}_j - f_j)$  é mínima. Temos

$$\text{Var}(\hat{f}_j - f_j) = E(X' a_j - f_j)^2$$

Derivando esta expressão em relação a  $a_j$  e igualando a zero obteremos a seguinte expressão para  $\hat{f}$  que é o vetor cujas componentes são  $\hat{f}_1, \hat{f}_2, \dots, \hat{f}_m$ :

$$\hat{f} = \Lambda' \Sigma^{-1} X \tag{3.4.1}$$

Para evitar a inversão de uma matriz  $p \times p$ ,  $(\Sigma^{-1})$ , podemos escrever esta expressão de forma alternativa. Para isto fazemos uso da seguinte identidade:

$$\Lambda' \Sigma^{-1} = (\Lambda' \psi^{-1} \Lambda \phi + I)^{-1} \Lambda' \psi^{-1} \tag{3.4.2}$$

que pode ser verificada pré-multiplicando-se a expressão  $\Sigma = \Lambda \phi \Lambda' + \psi$  por  $\Lambda' \psi^{-1}$ . Substituindo-se (3.4.2) em (3.4.1) teremos:

$$\hat{f} = \phi(I + \Lambda' \psi^{-1} \Lambda \phi)^{-1} \Lambda' \psi^{-1} X$$

que é a expressão para se obter os estimadores dos escores fatoriais para o caso de fatores correlacionados (oblíquos). Para  $\phi = I$ , isto é, quando os fatores são não correlacionados (ortogonais), temos:

$$\hat{f} = (I + \Lambda' \psi^{-1} \Lambda)^{-1} \Lambda' \psi^{-1} X.$$

## 2) MÉTODO DE BARTLETT

Este método, desenvolvido por Bartlett (1938) adota o princípio de mínimos quadrados. Os escores são obtidos de tal forma que a soma de quadrados dos resíduos padronizados seja mínima, em relação aos elementos de  $f$ . Assim

$$\sum_{i=1}^p \frac{e_i^2}{\psi_i} = e' \psi^{-1} e = (X - \Lambda f)' \psi^{-1} (X - \Lambda f)$$

é derivado em relação ao vetor  $f$  e obtém-se:

$$\hat{f}^* = (\Lambda' \psi^{-1} \Lambda)^{-1} \Lambda' \psi^{-1} X$$

que é a expressão para se obter os estimadores dos escores fatoriais, tanto no caso de fatores correlacionados, como no caso de fatores não correlacionados.

Para o caso de fatores não correlacionados e quando a matriz  $\Lambda' \psi^{-1} \Lambda$  é diagonal, observa-se que os estimadores de  $f_j$  pelos dois métodos descritos diferem apenas por um fator escala.

Lawley & Maxwell mostram que os estimadores obtidos pelo método de regressão são viesados, enquanto que os obtidos pelo método de Bartlett são não viesados. No entan-

to, os estimadores do método de regressão tem menor variância do que os de Bartlett.

## CAPÍTULO IV

### TRANSFORMAÇÃO E INTERPRETAÇÃO

#### DOS FATORES

Após as cargas fatoriais terem sido obtidas, a próxima etapa na análise é a interpretação dos fatores. Para procurar uma melhor interpretação dos fatores é prática comum fazer uma rotação ou uma transformação dos fatores. Vários métodos de rotação e transformação dos fatores são propostos na bibliografia sobre Análise Fatorial. Exporremos os principais métodos, diferenciando-os com o intuito de esclarecer os pesquisadores que utilizam a técnica.

#### 4.1 - ROTAÇÃO DOS FATORES

Nos capítulos anteriores já nos referimos à falta de unicidade das soluções obtidas por Análise Fatorial, isto é, o conjunto das cargas fatoriais, obtidas por qualquer método de solução fatorial, quando o número de fatores comuns é maior do que um, não é único, pois outros conjuntos equivalentes podem ser encontrados por transformações ortogonais das cargas. Em outras palavras, se nós multiplicarmos a matriz de cargas fatoriais,  $p \times m$ ,  $A$ , por uma matriz ortogonal  $M$ , de ordem  $m$ , a decomposição da matriz de covariância  $\Sigma$  não é única pois se  $M$  é ortogonal:

$$(AM)(AM)' + \Psi = AMM'\Lambda' + \Psi = \Lambda\Lambda' + \Psi = \Sigma.$$

Assim, mesmo que os elementos de  $AM$  sejam diferentes das cargas originais, sua habilidade em gerar as covariâncias observadas é inalterada.

Na expressão  $X = \Lambda f + e$  se nós trocarmos  $f$  por  $M'f$ , além de  $\Lambda$  por  $AM$ , observamos que a expressão não se altera pois  $M$  é ortogonal. Na terminologia de Análise Fatorial, temos o que se chama *notação dos fatores*.

Uma figura geométrica ajudará a intuição, como em todo problema de análise multivariada. Podemos considerar as variáveis como pontos em um espaço  $m$ -dimensional (o espaço de fatores), no qual os eixos ortogonais representam os fatores, e as coordenadas destes pontos são as cargas fatoriais. Para facilitar a interpretação dos fatores é conveniente, em muitos casos, mudar o sistema de referência original, o que corresponde a rotação dos fatores como definido acima. Fazendo tal transformação de coordenadas, a configuração geométrica é inalterada.

Harman (1960) explica este fato dizendo que a elaborada fórmula consistindo de seis termos:

$$AX^2 + BY^2 + CXY + DX + EY + F = 0$$

representa uma configuração geométrica (uma elipse) em um sistema (arbitrário) de referência, enquanto a expressão

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

representa precisamente a mesma configuração em outro sistema de referência, selecionado para fazer a equação tão simples quanto possível. A única diferença é que os matemáti-

cos tratam com a configuração geométrica, usando o sistema de referência como ferramenta. Os psicólogos, no entanto tratam com a interpretação do sistema de referência, usando a configuração de pontos meramente como um veículo para obter um eixo de referência que melhor satisfaz as necessidades teóricas e práticas do problema em pesquisa.

Assim, então, o objetivo da rotação dos fatores é obter uma matriz de cargas mais facilmente interpretável ou mais identificável com a natureza das variáveis observadas.

A seguir apresentaremos um exemplo de rotação gráfica, que, cremos, tornará mais claro o problema rotação dos fatores. Este exemplo está exposto no texto de Lawley e Maxwell (1971) e segue o tipo de exemplos clássicos em Análise Fatorial, tratando com testes de natureza cognitiva.

A tabela 4.1.1 mostra os estimadores de máxima verossimilhança das cargas fatoriais para três fatores, obtidos de uma análise dos coeficientes de correlação entre os escores dos 10 testes para uma amostra de 292 crianças.

TABELA 4.1.1 - CARGAS FATORIAIS OBTIDAS PELO MÉTODO DE MÁXIMA VEROSIMILHANÇA PARA UM CONJUNTO DE TESTES COGNITIVOS

| T E S T E S                       | CARGAS FATORIAIS |                 |                  | COMUNALIDADE |
|-----------------------------------|------------------|-----------------|------------------|--------------|
|                                   | I <sub>0</sub>   | II <sub>0</sub> | III <sub>0</sub> |              |
| 1. Compreensão . . . . .          | 0,788            | -0,152          | -0,352           | 0,768        |
| 2. Aritmética . . . . .           | 0,874            | 0,381           | 0,041            | 0,911        |
| 3. Semelhanças . . . . .          | 0,814            | -0,043          | -0,213           | 0,710        |
| 4. Vocabulário . . . . .          | 0,798            | 0,170           | -0,204           | 0,707        |
| 5. Memorização . . . . .          | 0,641            | 0,070           | -0,042           | 0,418        |
| 6. Complementação de figuras . .  | 0,755            | -0,298          | 0,067            | 0,663        |
| 7. Arranjo de figuras . . . . .   | 0,782            | -0,221          | 0,028            | 0,661        |
| 8. Planejamento de blocos . . . . | 0,767            | -0,091          | 0,358            | 0,725        |
| 9. Montagem de objetos . . . . .  | 0,733            | -0,384          | 0,229            | 0,737        |
| 10. Simbologia numérica . . . . . | 0,771            | -0,101          | 0,071            | 0,610        |

Ao observarmos as cargas fatoriais da tabela acima vemos que todas as cargas correspondendo ao fator  $I_0$  são positivas e relativamente altas. Este fator pode então ser interpretado como um fator de inteligência geral. A interpretação dos fatores  $II_0$  e  $III_0$  não é tão imediata, pelo número de cargas negativas que aparecem e pelo pequeno tamanho das cargas. Examinaremos o problema, então, através da representação geométrica para procurar classificar a interpretação dos dois últimos fatores. O ideal seria fazer um modelo tri-dimensional, no qual os eixos ortogonais representam os fatores, mas como isto não é possível aqui, devemos nos contentar em olhar os fatores dois por vez. Para começar, vejamos a representação dos fatores  $I_0$  e  $III_0$  na Fig. 4.1.1.

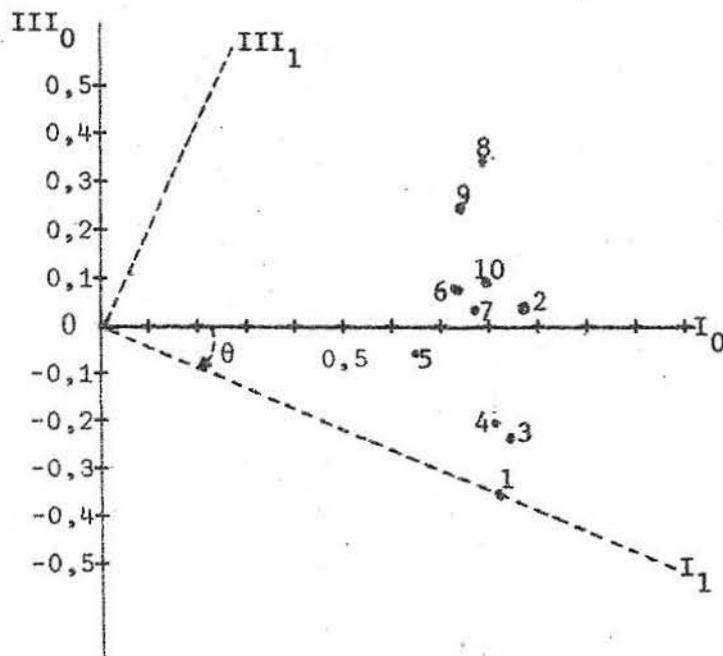


Figura 4.1.1

Ao examinarmos os pontos num gráfico devemos verificar se eles não estão concentrados em grupos. O exame da fig. 4.1.1 mostra que se os eixos são rotacionados ortogonalmen

te na direção horária, de tal forma que o primeiro fator passe através do ponto teste 1, então todos os pontos estarão no primeiro quadrante e todos os sinais das cargas se tornarão positivos. Em sua nova posição os fatores são denotados por  $I_1$  e  $III_1$  (os subscritos referem-se ao número de rotações). O fator I, pode ainda ser denominado *inteligência geral*, já que o teste 1 é um teste de compreensão. Para se obter as cargas sobre estes novos eixos ou fatores, deve-se pôs multiplicar a matriz de cargas originais,  $10 \times 2$ , correspondendo só aos fatores  $I_0$  e  $III_0$  por uma matriz ortogonal da forma

$$\begin{bmatrix} \cos\theta & \text{sen}\theta \\ -\text{sen}\theta & \cos\theta \end{bmatrix}$$

O ângulo  $\theta$  pode ser determinado ou medindo-se diretamente na figura, ou por cálculo. No exemplo  $\theta$  é aproximadamente  $24^\circ$  e as novas cargas sobre os fatores  $I_1$  e  $III_1$  são dadas na tabela 4.1.2.

TABELA 4.1.2 - CARGAS SOBRE OS FATORES ROTADOS

| TESTE | a) PRIMEIRA ROTAÇÃO |         | b) SEGUNDA ROTAÇÃO |        |         |
|-------|---------------------|---------|--------------------|--------|---------|
|       | $I_1$               | $III_1$ | $I_1$              | $II_1$ | $III_1$ |
| 1     | 0,863               | 0,000   | 0,863              | -0,106 | 0,109   |
| 2     | 0,781               | 0,394   | 0,781              | 0,548  | 0,000   |
| 3     | 0,830               | 0,138   | 0,830              | 0,069  | 0,127   |
| 4     | 0,812               | 0,139   | 0,812              | 0,018  | 0,219   |
| 5     | 0,602               | 0,223   | 0,602              | 0,209  | 0,105   |
| 6     | 0,622               | 0,369   | 0,622              | 0,058  | 0,471   |
| 7     | 0,703               | 0,345   | 0,703              | 0,094  | 0,399   |
| 8     | 0,554               | 0,640   | 0,554              | 0,397  | 0,510   |
| 9     | 0,576               | 0,508   | 0,576              | 0,098  | 0,629   |
| 10    | 0,675               | 0,379   | 0,675              | 0,202  | 0,363   |

Agora fazemos o gráfico para os fatores  $II_0$  e  $III_1$  (figura 4.1.2) e observa-se que fazendo uma rotação ortogo-

nal, agora no sentido anti-horário, pode-se novamente eliminar quase todas cargas negativas. Neste caso é conveniente fazer passar o eixo do fator II através do ponto-teste 2, que é o teste de Aritmética, pois isto pode auxiliar na interpretação deste novo fator. A matriz de rotação agora é do tipo

$$\begin{bmatrix} \cos\theta & -\text{sen}\theta \\ \text{sen}\theta & \cos\theta \end{bmatrix}$$

O ângulo  $\theta$  na figura é aproximadamente  $46^\circ$  e as novas cargas sobre os fatores, denotados agora por  $II_1$  e  $III_2$ , junto com as cargas sobre  $I_1$  são dadas na tabela 4.1.2 (b)

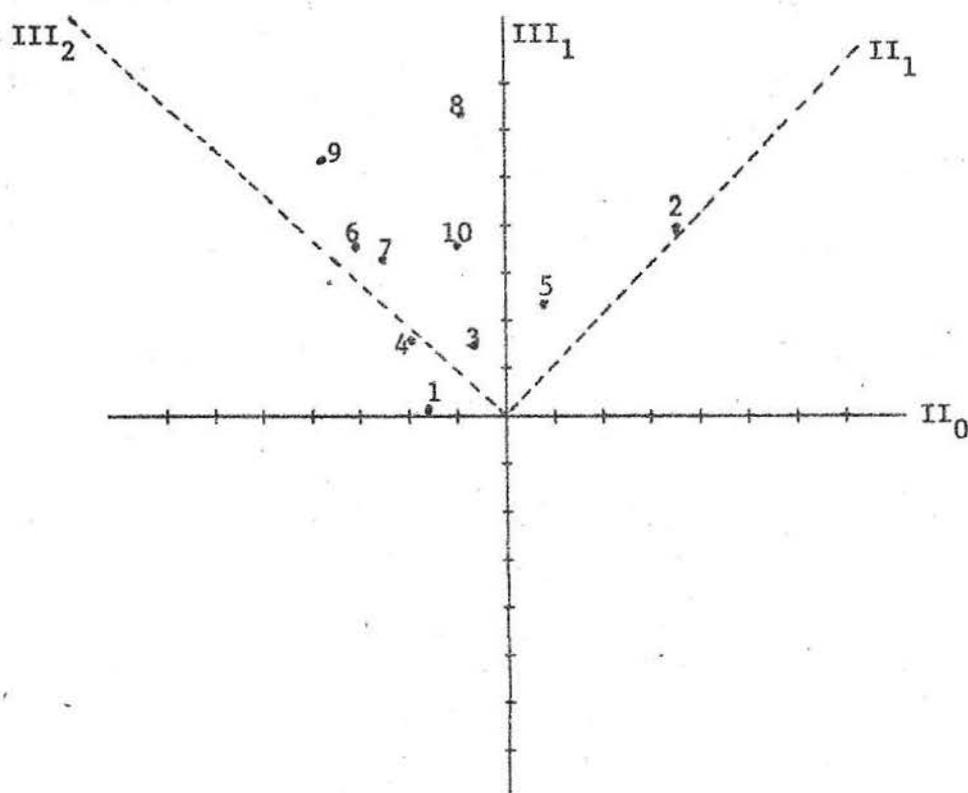


Figura 4.1.2

Finalmente olhamos o novo conjunto de cargas, mos-

trado na tabela 4.1.2,(b) para ver se os fatores podem ser interpretados com maior significado. O fator I, pode ainda ser descrito como um fator de *inteligência geral*. Sobre o fator  $III_2$ , as maiores cargas são para os últimos cinco testes. A semelhança entre estes testes, após o efeito do fator  $I_1$  ter sido removido, mostra tratar-se da habilidade para visualizar, tal que este terceiro fator poderia ser denominado *visualização*. Sobre o fator  $II_1$ , as maiores cargas são para Aritmética (2) e Planejamento de Blocos (8). Como o teste Planejamento de Blocos não trata com números, seria imprudente ver o fator como sendo habilidade numérica, e uma melhor descrição dele deve ser *conhecimento mecânico* ou mesmo *memória*.

Este exemplo ilustra bem o tipo de problema de interpretar os resultados em uma Análise Fatorial. Mas, o fato de que outras matrizes ortogonais poderiam ter sido usadas para fazer a rotação dos fatores, deixa claro o elemento subjetivo do problema. Além disso dois ou mais pesquisadores trabalhando com os mesmos dados poderiam obter diferentes soluções. Isto fez com que os analistas fatoriais buscassem métodos, que, em algum sentido pré-determinado, conduzissem a um conjunto de resultados únicos.

Thurstone formulou, inicialmente, o conceito de Estrutura Simples, cuja idéia geral é de que muitos testes (variáveis) não dependeriam de todos os fatores, e então os coeficientes ou as cargas fatoriais destes fatores deveriam ser nulas. Assim, como explicam Anderson e Rubin (1956), dada uma matriz de cargas  $\Lambda$ , se consideraria todas as rotações, isto é todas as matrizes da classe  $\Lambda M$ , onde  $M$  é ortogonal, e se escolheria aquela com maior número de coeficientes nulos. Esta matriz, considerada como a que fornece a estrutura mais simples, proporcionaria uma interpretação mais expressiva dos fatores. Thurstone também sugeriu estrutura simples quan

do fatores correlacionados são considerados. Ele formulou cinco regras para se escolher a matriz de carga final, formando o seu Princípio da Estrutura Simples e que estão expostas nas obras de Thurstone (1947), Harman (1960) e outros.

Morrison (1967) comenta que estrutura simples satisfazendo as condições de Thurstone raramente existem em soluções fatoriais extraídas de dados reais.

Os princípios de estrutura simples foram empregados originalmente nos métodos gráficos de transformação dos fatores e mais recentemente tem sido adaptados a procedimentos analíticos. Estes procedimentos, ou métodos analíticos além de simplificar a estrutura das cargas fatoriais, fornecem soluções únicas. Eles foram desenvolvidos tanto para rotações ortogonais dos fatores como transformações oblíquas.

Veremos inicialmente os métodos analíticos de rotações ortogonais: o Método Quartimax e o Método Varimax.

Seja  $M$  uma matriz ortogonal de ordem  $m$ , cuja  $j$ -ésima coluna é denotada por  $m_j$  ( $j=1, \dots, m$ ). Seja, como usualmente,  $A$  a matriz  $p \times m$  das cargas fatoriais não rotadas. A  $i$ -ésima coluna de  $A'$  será denotada pelo vetor  $m \times 1$ ,  $l_i$ . A matriz  $\Gamma = [\gamma_{ij}]$  das cargas rotadas é dada por

$$\Gamma = AM$$

e então

$$\gamma_{ij} = l_i' m_j.$$

A matriz  $\Gamma$  é comumente denominada estrutura fatorial cujos elementos,  $\gamma_{ij}$ , representam a correlação entre as variáveis e os fatores rotados.

Após a rotação ortogonal ter sido feita, a comunalidade de cada variável  $X_i$ , permanece a mesma, isto é

$$h_i^2 = \sum_{j=1}^m \lambda_{ij}^2 = \sum_{j=1}^m \gamma_{ij}^2$$

pois como vimos, devido a ortogonalidade da matriz  $M$ , temos

$$\Lambda\Lambda' = (\Lambda M)(M'\Lambda') = \Gamma\Gamma'.$$

O objetivo do método analítico Quartimax é determinar a transformação ortogonal  $M$  que transforma a matriz  $\Lambda$  na nova matriz fatorial  $\Gamma$  para a qual a variância dos quadrados das cargas fatoriais é um máximo. Então, a matriz de rotação  $M$ , deveria ser escolhida tal que maximizasse a expressão abaixo em relação aos elementos de  $M$  (note que  $\gamma_{ij} = \ell'_{im_j}$ ):

$$Q = \frac{1}{mp} \sum_{i=1}^p \sum_{j=1}^m \gamma_{ij}^4 - (\bar{\gamma}^2)^2$$

onde

$$\bar{\gamma}^2 = \frac{1}{mp} \sum_{i=1}^p \sum_{j=1}^m \gamma_{ij}^2.$$

Como  $\bar{\gamma}^2$  permanece constante sob transformação ortogonal, e termos constantes não afetam o processo de maximização, este critério é equivalente a maximização da soma das quarta potência das cargas fatoriais. Daí o nome Quartimax. Este critério é devido a Nenhaus e Wrigley (1954), mas outros autores trabalhando independentemente chegaram a critérios equivalentes.

Pela solução Quartimax, os grandes valores das car

gas tendem a se tornar maiores e os pequenos tendem a ser menores em valor absoluto, comparado com os tamanhos originais. O método Quartimax diz respeito à simplificação da descrição de cada linha (correspondendo às variáveis) da matriz de cargas fatoriais. Em contraposição, Kaiser (1958) deu mais ênfase à simplificação das colunas (correspondente aos fatores) da matriz de cargas, em uma tentativa para satisfazer os requisitos para estrutura simples e estabeleceu o critério Varimax.

O critério Varimax é uma modificação do método Quartimax e é o que mais se aproxima da estrutura simples. Kaiser definiu a simplicidade de um fator  $j$  como a variância de suas cargas ao quadrado, isto é

$$V_j = \frac{1}{p} \sum_{i=1}^p (\gamma_{ij}^2)^2 - \frac{1}{p^2} \left( \sum_{i=1}^p \gamma_{ij}^2 \right)^2, \quad j=1, \dots, m.$$

Quando a variância atinge um máximo, o fator tem maior interpretabilidade ou simplicidade, no sentido de que as cargas deste fator tendem ou à unidade, ou à zero. O critério de máxima simplicidade de uma matriz fatorial completa é definido como a maximização da soma destas simplicidades.

Como este critério dá igual peso às variáveis com comunalidades grandes ou pequenas, Kaiser sugeriu que antes de iniciar o processo de maximização, as cargas deveriam ser divididas pela raiz quadrada da comunalidade correspondente, o que é equivalente a normalizar os vetores  $\ell_i$ . Após a matriz  $M$  ter sido obtida, as cargas finais deveriam ser multiplicadas novamente pela raiz quadrada da comunalidade. A este critério varimax modificado, Kaiser denominou *Critério Varimax Normal* e é o mais utilizado.

No método original de Kaiser, os fatores eram rotados em pares até que as cargas convergissem para seus valores finais. Horst em 1965 sugeriu o que ele denominou *Solução Varimax de fatores simultâneos* pela qual em cada iteração todos os fatores são rotados simultaneamente. Este método tem a vantagem de não acumular erros de arredondamento e de ser mais rápido do que os anteriores. O processo de cálculo para este método está exposto muito claramente no texto de Lawley e Maxwell (1971).

Pelo processo Varimax, os fatores também são rotados de forma que as novas cargas tendem a ser relativamente maiores ou menores em tamanho absoluto comparado com seus tamanhos originais, como para o processo Quartimax.

O método Varimax parece ter suas limitações, segundo Lawley e Maxwell. A primeira é que o padrão de cargas obtido pode mudar muito se fatores adicionais são incluídos na rotação. A segunda é que ele não é muito útil em casos onde existe um fator geral dominante.

Como exemplo, na tabela 4.1.3, apresentaremos os resultados.  
TABELA 4.1.3 - ROTAÇÃO VARIMAX DAS CARGAS DA TABELA 4.1.1

| VARIÁVEL | CARGAS FATORIAIS |       |       | COMUNALIDADE |
|----------|------------------|-------|-------|--------------|
|          | I                | II    | III   |              |
| 1        | 0,759            | 0,329 | 0,289 | 0,768        |
| 2        | 0,340            | 0,849 | 0,274 | 0,911        |
| 3        | 0,633            | 0,450 | 0,327 | 0,710        |
| 4        | 0,657            | 0,345 | 0,397 | 0,707        |
| 5        | 0,370            | 0,453 | 0,276 | 0,418        |
| 6        | 0,464            | 0,263 | 0,615 | 0,663        |
| 7        | 0,485            | 0,332 | 0,561 | 0,660        |
| 8        | 0,183            | 0,472 | 0,684 | 0,725        |
| 9        | 0,354            | 0,209 | 0,754 | 0,737        |
| 10       | 0,409            | 0,423 | 0,514 | 0,610        |

| MATRIZ DE ROTAÇÃO M |        |       |        |
|---------------------|--------|-------|--------|
|                     | 0,560  | 0,633 | 0,534  |
|                     | -0,311 | 0,758 | -0,573 |
|                     | -0,768 | 0,155 | 0,622  |

sultados da aplicação da solução Varimax para fatores simultâneos para os dados do exemplo anterior (apresentados na tabela 4.1.1), como consta no texto de Lawley & Maxwell.

Estes resultados são menos facilmente interpretáveis do que os resultados obtidos por rotação gráfica anteriormente, devido a presença de um fator geral, o fator de inteligência. Deve ser notado, entretanto, que na bibliografia encontram-se vários exemplos da aplicação do método Varimax com ótimos resultados para a interpretação de fatores.

Os métodos Quartimax e Varimax fazem parte das opções oferecidas para Rotação de Fatores do subprograma FACTOR do SPSS.

#### 4.2 - TRANSFORMAÇÃO DOS FATORES - CASO OBLÍQUO

Em muitas aplicações a configuração das cargas fatoriais pode ser mais simplificada ainda se os fatores são transformados em fatores oblíquos ou correlacionados. Métodos analíticos, assim como gráficos, foram também elaborados para este caso. Vamos nos referir, nesta secção, aos primeiros métodos analíticos sugeridos, voltando nossa maior atenção a um método bem mais recente, devido a Hendrickson e White (1964).

Os trabalhos iniciais sobre métodos analíticos para obtenção de fatores finais oblíquos foram desenvolvidos principalmente por Carrol em 1953 e por Kaiser e Dickman em 1959.

Nos métodos analíticos, para o caso oblíquo, o problema é determinar uma matriz de transformação,  $U$ , não necessariamente ortogonal que transforme a matriz de cargas fatoriais iniciais em uma nova matriz  $V$ ,  $p \times m$ , tal que

$$V = AU$$

Esta matriz  $V = [v_{ij}]$  é denominada *estrutura fatorial de referência*, ou simplesmente, estrutura de referência e as novas cargas "fatoriais" representam a correlação entre as variáveis e os eixos (fatores) de referência. No caso de fatores não correlacionados, os eixos de referência são os próprios fatores rotados.

Existem vários tipos de soluções oblíquas, que podem ser resumidos nos seguinte:

Uma solução oblíqua obtida sob a condição que o critério

$$K = \frac{\sum_{i=1}^p \sum_{j=1}^m v_{ij}^4}{\left( \sum_{i=1}^p \sum_{j=1}^m v_{ij}^2 \right)^2}$$

seja um máximo é chamada solução Oblimax, cujo termo foi sugerido por Saunders.

Além deste critério existe uma classe de soluções, denominados métodos Oblimin (indireto), que envolvem a minimização do critério geral:

$$B = \sum_{j < k = 1}^m \left[ \sum_{i=1}^p v_{ij}^2 v_{ik}^2 - \gamma \sum_{i=1}^p v_{ij}^2 \sum_{i=1}^p v_{ik}^2 \right]$$

onde  $\gamma$  é uma constante arbitrária que pode assumir qualquer valor entre 0 e 1. Quando  $\gamma = 1$ , o critério é denominado Covarimin (que é o análogo oblíquo do método Varimax). Para  $\gamma = 0$ , temos o critério Quartimin (análogo ao método Quartimax). Se  $\gamma = 1/2$ , o critério é denominado Biquartimin. Além destes valores para  $\gamma$ , qualquer outro valor entre 0 e 1 poderia ser selecionado para  $\gamma$ , com uma correspondente variação do cri-

tério. Carroll sugere que os resultados são geralmente mais satisfatórios pelo método Biquartimin.

Ainda um outro procedimento analítico envolvendo fatores oblíquos e um critério minimizante é o de Kaiser e Dickman e como explica Harman, mesmo que ele não seja um caso especial da classe Oblimin, existe muita semelhança na expressão formal, ou seja que

$$D = \sum_{j < k = 1}^m \left[ \sum_{i=1}^p \left( \frac{v_{ij}^2}{h_i^2} \right) \left( \frac{v_{ik}^2}{h_i^2} \right) / \left( \sum_{i=1}^p \frac{v_{ij}^2}{h_i^2} \right) \left( \sum_{i=1}^p \frac{v_{ik}^2}{h_i^2} \right) \right]$$

seja um mínimo e para o qual se considera as cargas normalizadas.

Ao se fazer a escolha de uma forma de solução, uma importante consideração a fazer é o grau de correlação entre os fatores finais. Nos estudos empíricos, feitos por Harman, o critério Quartimin mostrou produzir fatores altamente correlacionados. Pelo critério Covarimin, os fatores finais mostraram ser quase ortogonais. Os critérios Biquartimin e o de Kaiser-Dickman produzem fatores não muito correlacionados, mas também não quase ortogonais. O critério Oblimax também leva a fatores finais altamente correlacionados.

Harman comenta também que não há experiência suficiente, para mostrar que exista superioridade de um destes métodos, para justificar a exclusão dos outros. Os critérios Biquartimin e Kaiser-Dickman têm a vantagem de produzir fatores nem tão oblíquos e nem tão ortogonais, mas a vantagem de um em relação ao outro, depende da natureza dos dados em consideração: se os dados são particularmente simples, ou extremamente complexos, o Critério de Kaiser-Dickman é melhor, mas se os dados são moderadamente complexos, o critério Biquartimin é superior.

É conveniente observar que o método Oblimin comentado acima é o método Oblimin Indireto. Para algumas modificações no método resulta o método Oblimin Direto que é o que consta no Subprograma FACTOR do SPSS:

Finalmente, exporemos com detalhes matemáticos, o método analítico mais recente para o problema de fazer uma transformação a fatores oblíquos. O método é denominado Promax e é devido a Hendrickson e White (1964) e tem mostrado bons resultados na prática. Seguiremos a abordagem feita por Lawley & Maxwell (1971).

O método Promaxa parte de uma matriz de cargas fatoriais rotadas pelo método Varimax,  $\Gamma = [\gamma_{ij}]$ . Desta matriz é construída outra matriz  $p \times m$ ,  $Q = [q_{ij}]$  cujos elementos são definidos por

$$q_{ij} = |\gamma_{ij}^{k-1}| \cdot \gamma_{ij}$$

onde  $k$  é algum inteiro satisfazendo  $k > 1$ . Então  $q_{ij} = 0$ , se  $\gamma_{ij} = 0$ , caso contrário  $q_{ij}$  tem o mesmo sinal que  $\gamma_{ij}$  e tem o mesmo valor absoluto que  $(\gamma_{ij})^k$ . O valor escolhido para  $k$  é obtido por tentativas, mas a experiência tem mostrado que valores maiores do que quatro (4), apesar de poderem simplificar a configuração das cargas, produzem fatores muito correlacionados.

Deve-se agora determinar uma matriz  $U$ , da seguinte maneira: sejam  $q_j$  e  $u_j$  a  $j$ -ésima coluna de  $Q$  e  $U$ , respectivamente. Para cada valor de  $j$ , nós escolhemos  $u_j$  tal que ele minimize a expressão

$$(q_j - \Gamma u_j)' (q_j - \Gamma u_j).$$

Derivando esta expressão em relação ao vetor  $u_j$ , temos:

$$-2\Gamma'(q_j - \Gamma u_j)$$

e igualando a zero, obtemos

$$(\Gamma'\Gamma)u_j = \Gamma'q_j.$$

Para todos os valores de  $j$ , temos finalmente

$$(\Gamma'\Gamma)U = \Gamma'Q$$

ou

$$U = (\Gamma'\Gamma)^{-1}\Gamma'Q.$$

Esta matriz  $U$  tem o efeito, em geral, de aumentar as cargas, em tamanho absoluto, que são relativamente grandes e de diminuir as pequenas.

Na prática é conveniente fazer uma mudança de escala nas colunas de  $U$  tal que os fatores transformados tenham variância unitária. Para isto determinamos a matriz diagonal  $D$ , com elementos diagonais positivos, que satisfaz

$$D^2 = \text{dia}[(U'U)^{-1}].$$

A matriz de transformação  $M$ , que substitui  $U$ , é então dada por

$$M = UD$$

e a matriz de cargas transformadas é

$$\Lambda^* = \Gamma M.$$

Nós agora temos

$$\begin{aligned} \Gamma\Gamma' &= \Lambda^*M^{-1}M'^{-1}\Lambda^* \\ &= \Lambda^*(M'M)^{-1}\Lambda^* \\ &= \Lambda^*\Phi\Lambda^* \end{aligned}$$

onde

$$\Phi = (M'M)^{-1} = D^{-1}(U'U)^{-1}D^{-1}$$

Pelas equações do modelo fatorial oblíquo, está claro que  $\Phi$  é a matriz de covariância dos fatores transformados e da definição de  $D$ , é evidente que  $\Phi$  tem elementos diagonais unitários. Então  $\Phi$  é também a matriz de correlação para os novos fatores, os quais de fato foram padronizados.

A aplicação do método Promax, com  $k = 2$ , para o exemplo dos 10 testes cognitivos, a partir das cargas Varimax da tabela 4.1.3 mostra os seguintes resultados, segundo Lawley & Maxwell (1971).

TABELA 4.2.1 - TRANSFORMAÇÃO PROMAX  
DAS CARGAS DA TABELA 4.1.3

| TESTE | CARGAS FATORIAIS |       |       |
|-------|------------------|-------|-------|
|       | I                | II    | III   |
| 1     | 0,737            | 0,139 | 0,086 |
| 2     | -0,012           | 0,924 | 0,074 |
| 3     | 0,522            | 0,320 | 0,137 |
| 4     | 0,590            | 0,163 | 0,230 |
| 5     | 0,207            | 0,410 | 0,141 |
| 6     | 0,362            | 0,060 | 0,539 |
| 7     | 0,364            | 0,155 | 0,460 |
| 8     | -0,086           | 0,388 | 0,646 |
| 9     | 0,234            | 0,011 | 0,732 |
| 10    | 0,236            | 0,304 | 0,408 |

| MATRIZ DE TRANSFORMAÇÃO (M) |        |        | MATRIZ DE CORRELAÇÃO DOS FATORES ( $\Phi$ ) |       |       |
|-----------------------------|--------|--------|---|-------|-------|
| 1,222                       | -0,279 | -0,239 | 1,000                                       | 0,621 | 0,461 |
| -0,459                      | 1,277  | -0,183 |   | 1,000 | 0,481 |
| -0,136                      | -0,238 | 1,134  |   |       | 1,000 |

Observa-se que esta configuração das cargas é muito mais simples do que a da tabela 4.2.1, havendo um grande número de cargas pequenas. A interpretação para os fatores II e III é a mesma que anteriormente, apenas o fator I pode agora ser interpretado como um *fator verbal*.

## CAPÍTULO V

### ALGUMAS APLICAÇÕES DA ANÁLISE FATORIAL

Como vimos, a Análise Fatorial tem aplicação nos mais diversos campos da ciência. Já nos referimos à alguns exemplos, parcialmente. Neste capítulo mostraremos como Análise Fatorial pode ser utilizada num problema na área da Agricultura, apresentando os resultados nas diversas etapas do método. Em um outro exemplo ilustraremos o problema da de terminação de um modelo de regionalização, onde Análise Fatorial foi utilizada para a construção de um índice sócio-e conômico que permitiu obter uma hierarquização de cidades com a finalidade de estabelecer regiões prioritárias para in vestimentos na área de saneamento básico.

Várias outras aplicações da técnica de Análise Fatorial poderiam ter sido apresentadas, mas, infelizmente não é possível nos referirmos a todas. Acreditamos que estes dois exemplos serão suficientes para dizer da importância e apli cabilidade do método.

#### 5.1 - EXEMPLO DO USO DA ANÁLISE FATORIAL EM PESQUISAS AGRÍCOLAS

Este exemplo é parte de uma pesquisa sobre as cau sas dos diferentes rendimentos do fumo, em uma determinada

região agrária do Rio Grande do Sul caracterizada por grande diversificação quanto à capacidade do uso do solo e geomorfologia. O trabalho é de autoria do geógrafo José Alberto Moreno da U.G.C. - CEMAPA - Secretaria de Agricultura do Estado do Rio Grande do Sul e os resultados foram publicados no Boletim Geográfico do Rio Grande do Sul, Ano 18, nº 16, 1973 sob o título *Pesquisa das Causas dos Diferentes Rendimentos do Fumo*.

A técnica de Análise Fatorial foi aplicada para analisar as intercorrelações entre 17 variáveis, relativas a produção de fumo, com o objetivo de identificar um menor número de fatores que contivessem aproximadamente o mesmo total de informação expresso pelas variáveis originais.

As informações utilizadas são relativas à safra do fumo de 1971, de 83 propriedades produtoras de fumo, cujos proprietários são fornecedores de fumo em folha à Companhia de Cigarros Souza Cruz.

As 17 variáveis analisadas são as seguintes:

- Var 01 - Área cultivada com fumo na propriedade
- Var 02 - Número de pés de fumo por m<sup>2</sup>
- Var 03 - Número de pés de fumo plantados
- Var 04 - Fumo produzido na propriedade (em Kg)
- Var 05 - Rendimento bruto com a venda do fumo (em Cr\$)
- Var 06 - Preço médio por Kg pago pela Companhia Souza Cruz
- Var 07 - Rendimento Kg/hectare
- Var 08 - Rendimento Cr\$/hectare
- Var 09 - Fertilizantes aplicados (em Kg)
- Var 10 - Inseticidas e fungicidas empregadas (em Kg)
- Var 11 - Quantidade de lenha consumida (em m<sup>3</sup>)
- Var 12 - Mão de obra empregada (dias-homem)
- Var 13 - Renda líquida
- Var 14 - Quantidade de inseticidas por pé plantado

Var 15 - Quantidade de adubo por pē plantado

Var 16 - Quantidade de lenha consumida por Kg de fumo

Var 17 - Mão de obra por hectare colhidos (dias-homem)

O método de Análise Fatorial empregado foi a Solução dos Fatores Principais descrito no Capítulo III<sup>(1)</sup>. Feito o processamento dos dados, utilizando-se o subprograma "FACTOR" do SPSS, destacamos os seguintes resultados. Na tabela 5.1.1 apresentamos a matriz de correlação obtida.

TABELA 5.1.1 - MATRIZ DE CORRELAÇÃO

|    | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 11    | 12    | 13    | 14    | 15    | 16    | 17    |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1  | 1,00  | 0,13  | 0,99  | 0,74  | 0,64  | -0,20 | -0,28 | -0,31 | 0,94  | 0,71  | 0,98  | 0,07  | 0,08  | 0,01  | -0,08 | 0,19  | 0,22  |
| 2  | 0,15  | 1,00  | 0,01  | 0,10  | 0,18  | 0,13  | -0,07 | 0,01  | 0,94  | 0,03  | 0,02  | 0,03  | 0,20  | 0,03  | 0,09  | -0,10 | -0,08 |
| 3  | 0,99  | 0,01  | 1,00  | 0,74  | 0,62  | -0,32 | -0,27 | -0,32 | 0,95  | 0,71  | 0,99  | 0,98  | 0,06  | 0,01  | -0,09 | 0,20  | 0,24  |
| 4  | 0,74  | 0,10  | 0,74  | 1,00  | 0,94  | -0,08 | 0,39  | 0,31  | 0,63  | 0,33  | 0,71  | 0,70  | 0,66  | -0,24 | -0,27 | -0,48 | -0,45 |
| 5  | 0,64  | 0,18  | 0,62  | 0,94  | 1,00  | 0,27  | 0,43  | 0,46  | 0,54  | 0,27  | -0,60 | 0,59  | 0,61  | -0,22 | -0,20 | -0,52 | -0,41 |
| 6  | -0,20 | 0,13  | -0,22 | -0,08 | 0,27  | 1,00  | 0,14  | 0,53  | -0,16 | -0,12 | -0,22 | -0,22 | 0,47  | 0,04  | 0,18  | -0,18 | -0,17 |
| 7  | -0,28 | -0,07 | -0,27 | 0,39  | 0,43  | 0,14  | 1,00  | 0,91  | -0,35 | -0,40 | -0,29 | -0,30 | 0,75  | -0,29 | -0,23 | -0,93 | -0,92 |
| 8  | -0,31 | 0,01  | -0,32 | 0,31  | 0,48  | 0,53  | 0,91  | 1,00  | -0,37 | -0,39 | -0,34 | -0,35 | 0,84  | -0,23 | -0,14 | -0,66 | -0,86 |
| 9  | 0,94  | 0,04  | 0,95  | 0,63  | 0,54  | -0,16 | -0,35 | -0,37 | 1,00  | 0,73  | 0,95  | 0,94  | -0,03 | 0,09  | 0,22  | 0,29  | 0,33  |
| 10 | 0,71  | 0,03  | 0,71  | 0,33  | 0,27  | -0,12 | -0,40 | -0,39 | 0,73  | 1,00  | 0,72  | 0,73  | -0,22 | 0,69  | 0,12  | 0,36  | 0,42  |
| 11 | 0,98  | 0,20  | 0,99  | 0,71  | 0,60  | -0,22 | -0,29 | -0,34 | 0,95  | 0,72  | 1,00  | 0,99  | 0,03  | 0,03  | -0,06 | 0,23  | 0,26  |
| 12 | 0,07  | 0,03  | 0,08  | 0,70  | 0,59  | -0,22 | -0,30 | -0,35 | 0,94  | 0,73  | 0,99  | 1,00  | 0,00  | 0,04  | -0,04 | 0,23  | 0,30  |
| 13 | 0,08  | 0,20  | 0,05  | 0,66  | 0,61  | 0,47  | 0,75  | 0,64  | -0,03 | -0,22 | 0,03  | 0,00  | 1,00  | -0,34 | -0,25 | -0,82 | -0,83 |
| 14 | 0,01  | 0,03  | 0,01  | -0,24 | -0,22 | 0,04  | -0,23 | -0,23 | 0,99  | 0,69  | -0,03 | 0,04  | -0,34 | 1,00  | 0,27  | 0,30  | 0,34  |
| 15 | -0,08 | 0,09  | -0,09 | -0,27 | -0,20 | 0,18  | -0,23 | -0,14 | 0,22  | 0,12  | -0,06 | -0,04 | -0,25 | 0,27  | 1,00  | 0,24  | 0,28  |
| 16 | 0,19  | -0,10 | 0,20  | -0,46 | -0,32 | -0,18 | -0,93 | -0,86 | 0,29  | 0,36  | 0,23  | 0,23  | -0,82 | 0,30  | 0,24  | 1,00  | 0,96  |
| 17 | 0,22  | -0,08 | 0,24  | -0,45 | -0,49 | -0,17 | -0,92 | -0,86 | 0,33  | 0,42  | 0,26  | 0,30  | -0,83 | 0,34  | 0,28  | 0,96  | 1,00  |

O programa tem várias opções para determinação do número de fatores a serem extraídos. Escolheu-se a opção que considera apenas os fatores associados às raízes características maiores ou igual a 1,0. A variância total das variáveis padronizadas, é 17 e corresponde ao traço da matriz de

TABELA 5.1.2

| FATOR | RAÍZ CARACTERÍSTICA | PERCENTAGEM DA VARIÂNCIA TOTAL EXPLICADA PELO FATOR | PERCENTAGEM ACUMULADA |
|-------|---------------------|---|-----------------------|
| 1     | 6,80                | 40,00   | 40,00                 |
| 2     | 5,54                | 32,60   | 72,61                 |
| 3     | 1,60                | 9,40  | 82,01                 |
| 4     | 1,10                | 6,46  | 88,47                 |

(1) - Salienciamos no Capítulo III que o método de máxima verossimilhança é o que tem melhores propriedades, infelizmente os programas de computador para este método não estão ainda disponíveis em nosso país.

correlação. A tabela 5.1.2 mostra as raízes características e percentagem da variância explicada pelos 4 fatores.

Observa-se então que 88,47% da variância total das variáveis originais é explicada por 4 fatores. A matriz das cargas fatoriais (Matriz Fatorial) e as comunalidades para cada variável são dados na Tabela 5.1.3.

TABELA 5.1.3 - MATRIZ FATORIAL E COMUNALIDADES

| VARIÁVEL | F A T O R E S |       |       |       | COMUNALIDADES |
|----------|---------------|-------|-------|-------|---------------|
|          | 1             | 2     | 3     | 4     |               |
| 1        | 0,94          | 0,31  | -0,04 | 0,09  | 0,9894        |
| 2        | 0,05          | 0,13  | 0,36  | 0,61  | 0,5211        |
| 3        | 0,95          | 0,30  | -0,08 | 0,00  | 0,9989        |
| 4        | 0,51          | 0,84  | -0,11 | -0,03 | 0,9787        |
| 5        | 0,39          | 0,90  | 0,14  | 0,07  | 0,9866        |
| 6        | -0,27         | 0,24  | 0,69  | 0,23  | 0,6595        |
| 7        | -0,52         | 0,76  | -0,06 | -0,26 | 0,9192        |
| 8        | -0,56         | 0,76  | 0,24  | -0,12 | 0,9632        |
| 9        | 0,95          | 0,18  | 0,09  | 0,06  | 0,9466        |
| 10       | 0,82          | -0,07 | 0,36  | -0,40 | 0,9669        |
| 11       | 0,95          | 0,27  | -0,07 | 0,01  | 0,9804        |
| 12       | 0,95          | 0,25  | -0,06 | 0,01  | 0,9687        |
| 13       | -0,22         | 0,93  | 0,17  | 0,09  | 0,9503        |
| 14       | 0,20          | -0,38 | 0,61  | -0,59 | 0,9046        |
| 15       | 0,06          | -0,32 | 0,59  | 0,16  | 0,4797        |
| 16       | 0,46          | -0,83 | 0,00  | 0,18  | 0,9174        |
| 17       | 0,50          | -0,81 | 0,05  | 0,12  | 0,9230        |

O programa fornece ainda a matriz fatorial rotada pelo método VARIMAX, caso nenhum outro método de rotação seja especificado. Analisando-se as matrizes fatoriais, com o objetivo de interpretar os fatores, decidimos usar apenas a matriz fatorial inicial por fornecer uma interpretação mais simples dos fatores, neste exemplo.

Observamos aqui, que na pesquisa original, a interpretação dos fatores foi feita a partir da matriz fatorial rotada pelo método VARIMAX e as análises e conclusões poste

riores basearam-se nos fatores rotados, que diferem um pouco dos obtidos neste exemplo.

As cargas fatoriais, quando a Análise Fatorial parte de uma matriz de correlação, são os coeficientes de correlação entre as variáveis e os fatores. Então, para interpretar cada fator nós olhamos para as variáveis com grandes cargas em valor absoluto, isto é para as variáveis altamente correlacionadas com o fator. Neste exemplo, observando-se a matriz fatorial, temos os seguintes fatores, tomando-se apenas as cargas  $\geq 0,60$  em valor absoluto.

FATOR 1 — envolvendo as variáveis:

- 1 - área cultivada
- 3 - Número de pés plantados
- 9 - Quantidade de adubo empregado
- 10 - Quantidade de inseticida empregado
- 11 - Lenha consumida
- 12 - Mão de obra

FATOR 2 — envolvendo as variáveis:

- 4 - Fumo produzido (Kg)
- 5 - Rendimento bruto (Cr\$)
- 7 - Rendimento Kg/hectare
- 8 - Rendimento Cr\$/hectare
- 13 - Renda líquida
- 16 - Lenha consumida por Kg de fumo (correlação negativa)
- 17 - Mão de obra por hectare colhido (correlação negativa)

FATOR 3 — representado pelas variáveis:

- 6 - Preço médio/Kg pago pela Companhia Souza Cruz
- 14 - Gramas inseticida por pé de fumo

FATOR 4 — envolvendo a variável:

- 2 - Pés de fumo por  $m^2$

Assim, vemos que o fator 1 que explica 40% da va-

riância total das variáveis originais é o fator *Investimento*. O fator 2, explicando 32,6% da variância total diz respeito à *Rendimento*. O terceiro fator é o fator *Qualidade do fumo*, explicando 9,4% e o último fator é interpretado como fator *Espaçamento*, explicando apenas 6,46% da variância total.

Uma das desvantagens do método de Análise Fatorial é que, às vezes, os fatores não são facilmente interpretáveis, mesmo após a rotação dos fatores. A interpretação dos fatores, como nos referimos no Capítulo IV, é subjetiva e pode variar de um pesquisador a outro numa mesma análise. Em alguns casos, a interpretação não é tão primordial, como quando se utiliza Análise Fatorial para a construção de índices, por exemplo.

Os valores de cada fator para cada propriedade são a seguir estimados. Há vários métodos de estimação dos escores fatoriais (ver Capítulo III), mas o mais utilizado é o método de regressão. Na tabela 5.1.4 apresentaremos os escores fatoriais das 83 propriedades para o Fator 1 apenas, a título de ilustração.

TABELA 5.1.4 - ESCORES FATORIAIS DO FATOR 1 PARA AS 83 OBSERVAÇÕES

|         |         |          |         |         |          |         |         |         |         |
|---------|---------|----------|---------|---------|----------|---------|---------|---------|---------|
| -0,0597 | -0,6950 | -6,7874  | 0,8895  | -0,1287 | -1,5661  | 4,1863  | -0,4563 | 3,3900  | 0,8171  |
| -5,1116 | -1,5391 | -0,7609  | -1,6603 | -0,4007 | -16,3219 | -7,9832 | 0,2484  | -8,9387 | 2,3767  |
| -0,5322 | 0,6305  | -12,3811 | 0,5786  | 15,1267 | 1,4196   | -4,3321 | -5,9575 | -2,2414 | -3,2630 |
| -1,7011 | -5,3560 | -0,4969  | -0,7063 | 1,6121  | 2,5174   | 16,2278 | 0,1367  | 3,4138  | -0,1406 |
| 0,6282  | 2,4414  | 7,5729   | 27,4995 | 26,8313 | 16,2278  | -7,2890 | -3,1987 | 1,5077  | 3,4165  |
| 0,0465  | 0,0094  | 0,9786   | 5,1515  | 6,9244  | -2,7433  | -7,2291 | -9,9838 | -6,2940 | 1,7202  |
| -7,8683 | 3,1651  | -1,4837  | 3,1118  | 1,5975  | -6,2492  | 0,1176  | -1,3788 | -1,9819 | 4,9680  |
| -3,4247 | 1,6225  | 0,5023   | -8,5983 | 2,1951  | -7,6247  | -9,3089 | -0,8972 | -1,5005 | -0,7798 |
| 2,2592  | 1,1270  | 2,4982   |         |         |          |         |         |         |         |

Na pesquisa original, à qual nos referimos no início desta secção, a aplicação da técnica de Análise Fatorial foi apenas uma etapa do trabalho. Após terem sido obtidos os escores fatoriais de cada fator para cada proprieda-

de, o autor utilizou o programa "ITERIM" de Análise Discriminante, com o objetivo de obter grupos de fumicultores, segundo os diferentes valores dos fatores, principalmente em relação aos fatores Investimento e Rendimento.

Como comentamos inicialmente, apresentamos aqui apenas uma parte da pesquisa, com o intuito de ilustrar a técnica de Análise Fatorial, outras considerações deveriam ter sido feitas caso o objetivo fosse apresentar os resultados da pesquisa.

Outro comentário a fazer é sobre o uso do método do Fator Principal, tanto neste exemplo como no próximo. Este fato deve-se às facilidades computacionais do método em relação ao de Máxima Verossimilhança, cujo programa não é tão acessível.

## 5.2 - MODELO DE REGIONALIZAÇÃO

Apresentaremos aqui outro exemplo do uso da técnica de Análise Fatorial, onde um conjunto de 16 variáveis sócio-econômicas, observadas em 98 municípios do Estado de Minas Gerais, foi estudado com o objetivo de obter uma hierarquização das cidades, através da construção de um índice que representasse a situação dos municípios e seus moradores.

A finalidade desta hierarquização era estabelecer regiões prioritárias para investimentos na área de saneamento básico em Minas Gerais. Este estudo é de autoria de Nicolas Botto Correa, da Serete Engenharia de São Paulo, que gentilmente nos forneceu as informações para este exemplo.

A técnica de Análise Fatorial, segundo o autor, foi utilizada devido às características do problema, que continha um grande número de variáveis, todas elas, a priori, com a mesma importância, tendo como objetivo imediato obter um

número mínimo de elementos que caracterizassem a importância sócio-econômica do município.

As 16 variáveis consideradas foram as seguintes:

1. SINS - (População sem instrução)/(População + de 5 anos)
2. LIGA - (Nº de ligações d'água)/(Nº de domicílios)
3. NDEN - (Nº de dentistas)/(População total)
4. NMED - (Nº de médicos)/(População total)
5. LIGE - (Nº de ligações elétricas)/(População total)
6. DEPO - (Depósitos em Bancos)/(População total)
7. RECE - (Receita Federal)/(População total)
8. EMIN - (Nº de empregos na indústria)/(População total)
9. ENSE - (Nº de empregos em serviço)/(População total)
10. EMAG - (Nº de empregos na agricultura)/(População total)
11. ICM - (Receita ICM)/(População total)
12. EFX1 - (Nº Pessoas com 1-5 anos estudo)/(Popul. + 5 anos)
13. EFX2 - (Nº Pessoas com 6-9 anos estudo)/(Popul. + 5 anos)
14. EFX3 - (Nº Pessoas com 10-12 anos estudo)/(Popul. + 5 anos)
15. EFX4 - (Nº Pessoas com 13-17 anos estudo)/(Popul. + 5 anos)
16. NVEI - (Nº de veículos)/(População total).

A Análise Fatorial feita a partir da matriz de correlação, cuja diagonal principal foi alterada, introduzindo-se as comunalidades estimadas de cada variável, como elementos diagonais. O estimador inicial de cada comunalidade é dado usualmente pelo quadrado do coeficiente de correlação múltipla. Várias iterações podem ser feitas para obter a comunalidade estimada.

A Análise Fatorial foi feita pela Solução dos Fatores Principais. Este método produz fatores ortogonais, como vimos. O programa utilizado é o "FACTOR" do SPSS.

Feito o processamento dos dados, foram extraídos dois fatores, associados às raízes características  $\geq 1$  da ma

triz de correlação alterado. O traço desta matriz, corresponde à comunalidade total das variáveis e é dado por 13,0366.

As raízes características e as percentagens da variância comum (comunalidade total) explicada pelos dois fatores são dados na Tabela 5.2.1.

TABELA 5.2.1

| FATOR | RAÍZ CARACTERÍSTICA | % DA VARIÂNCIA COMUM | % ACUMULADA |
|-------|---------------------|----------------------|-------------|
| 1     | 9,8708              | 75,7157              | 75,7157     |
| 2     | 1,8990              | 14,5672              | 90,2829     |

A tabela 5.2.2 mostra a matriz das cargas fatoriais e as comunalidades de cada variável

TABELA 5.2.2 - MATRIZ FATORIAL E COMUNALIDADES

| VARIÁVEL | FATOR 1 | FATOR 2 | COMUNALIDADES |
|----------|---------|---------|---------------|
| SINS     | 0,8911  | -0,1273 | 0,8103        |
| LIGA     | -0,7541 | 0,2901  | 0,6528        |
| NDEN     | -0,7746 | -0,3535 | 0,7249        |
| NMED     | -0,7829 | -0,3068 | 0,7071        |
| LIGE     | -0,9112 | 0,0497  | 0,8329        |
| DEPO     | -0,7605 | -0,3543 | 0,7039        |
| RECE     | -0,6060 | 0,5315  | 0,6498        |
| EMIN     | -0,5721 | 0,5679  | 0,6498        |
| ENSE     | -0,8794 | -0,2448 | 0,8333        |
| EMAG     | 0,8845  | -0,2408 | 0,8403        |
| ICM      | -0,3775 | 0,6446  | 0,5581        |
| EFX1     | -0,7035 | 0,2816  | 0,5742        |
| EFX2     | -0,9304 | -0,0166 | 0,8659        |
| EFX3     | -0,8865 | -0,2408 | 0,8439        |
| EFX4     | -0,7730 | -0,3749 | 0,7381        |
| NVEI     | -0,8654 | -0,1867 | 0,7839        |

As cargas dos fatores rotados pelo método VARIMAX são apresentadas na tabela 5.2.3.

TABELA 5.2.3 - MATRIZ FATORIAL ROTADA VARIMAX

| VARIÁVEL | FATOR 1 | FATOR 2 |
|----------|---------|---------|
| SINS     | 0,6693  | -0,6019 |
| LIGA     | -0,4649 | 0,6608  |
| NDEN     | -0,8402 | 0,1376  |
| NMED     | -0,8211 | 0,1811  |
| LIGE     | -0,7292 | 0,5487  |
| DEPO     | -0,8289 | 0,1291  |
| RECE     | -0,2074 | 0,7789  |
| EMIN     | -0,1590 | 0,7903  |
| ENSE     | -0,8668 | 0,2862  |
| EMAG     | 0,6006  | -0,6925 |
| ICM      | 0,0453  | 0,7457  |
| EFX1     | -0,4275 | 0,6256  |
| EFX2     | -0,7821 | 0,5042  |
| EFX3     | -0,8705 | 0,2935  |
| EFX4     | -0,8508 | 0,1190  |
| NVEI     | -0,8228 | 0,3267  |

A seguir o programa fornece os coeficientes de regressão dos fatores sobre as variáveis, denominados coeficientes dos escores fatoriais. Estes coeficientes são utilizados para determinar os escores fatoriais para cada observação. A tabela 5.2.4 mostra os coeficientes dos escores fatoriais.

TABELA 5.2.4 - COEFICIENTES DOS  
ESCORES FATORIAIS

| VARIÁVEL | 1       | 2       |
|----------|---------|---------|
| SINS     | -0,0192 | -0,1533 |
| LIGA     | 0,0300  | 0,1334  |
| NDEN     | -0,1364 | -0,1090 |
| NMED     | -0,1167 | -0,0829 |
| LIGE     | -0,0579 | 0,1206  |
| DEPO     | -0,1262 | -0,1023 |
| RECE     | 0,0870  | 0,2078  |
| EMIN     | 0,0961  | 0,2187  |
| EMSE     | -0,1861 | -0,0925 |
| EMAG     | -0,0304 | -0,2662 |
| ICM      | 0,0954  | 0,1870  |
| EFX1     | 0,0247  | 0,1046  |
| EFX2     | -0,1109 | 0,0936  |
| EFX3     | -0,1975 | -0,0950 |
| EFX4     | -0,1492 | -0,1243 |
| NVEI     | -0,1227 | -0,0403 |

Finalmente, os escores fatoriais são obtidos e temos, então, os valores dos fatores para cada município. A partir destes valores foi possível calcular um índice de classificação para os municípios.

A metodologia utilizada para a construção do índice baseou-se no fato de que os fatores resultantes são ortogonais e, portanto, linearmente independentes. Calculou-se o módulo da resultante da soma vetorial de todos os fatores de uma mesma observação, ponderados cada um deles, por sua percentagem da variância explicada. Posteriormente estes índices ( $I_i$ ) foram normalizados dividindo-se cada um deles pela diferença entre o maior e o menor, de tal forma a obter índices entre 0 e 1.

Seja  $f_{ij}$ ,  $i=1, \dots, N$ ,  $j=1, \dots, m$  o valor do escore

fatorial para a  $i$ -ésima observação e o  $j$ -ésimo fator e seja  $P_j$  a percentagem da variância comum (comunalidade total) explicada pelo fator  $j$ . O índice é dado por:

$$I_i = \sqrt{\sum_{j=1}^m P_j f_{ij}^2}, \quad i=1, \dots, N$$

onde  $N$  é o número de observações e  $m$  é o número de fatores.

Calculou-se, então, estes índices para as 98 cidades mineiras, normalizando-os entre 0 e 1. Classificou-se então as cidades e os resultados para as primeiras 5 e últimas 5 cidades, que são os seguintes:

|                                |       |
|--------------------------------|-------|
| 1. Belo Horizonte . . . . .    | 1,000 |
| 2. Juiz de Fora . . . . .      | 0,820 |
| 3. Uberaba. . . . .            | 0,767 |
| 4. Poços de Caldas. . . . .    | 0,698 |
| 5. Itajubá. . . . .            | 0,688 |
| ⋮                              | ⋮     |
| 94. Águas Formosas . . . . .   | 0,135 |
| 95. Salinas. . . . .           | 0,023 |
| 96. Brasília de Minas. . . . . | 0,020 |
| 97. Monte Azul . . . . .       | 0,003 |
| 98. São Francisco. . . . .     | 0,000 |

Como comenta o autor nas conclusões, o objetivo de hierarquizar regiões foi plenamente atendido pelo modelo.

Os índices, acima citados, poderiam ter sido obtidos sem a interpretação dos fatores, mas a análise relativa à interpretação dos dois fatores, neste trabalho é interessante.

Observando-se a matriz de cargas fatoriais rotados vemos que o primeiro fator dá maior importância (negativa) às variáveis:

-- Número de dentistas

- Número de médicos
- Depósitos em Bancos
- Empregos em serviços
- Pessoas com 10 - 12 anos de estudo
- Pessoas com 13 - 17 anos de estudo

e dá uma importância positiva às variáveis:

- Pessoas sem instrução
- Empregos na agricultura.

Conclui-se, então, que o primeiro fator representa adequadamente a *situação sócio-econômica-cultural da população* do município.

O segundo fator dá uma importância (positiva) às variáveis:

- Receita Federal
- I.C.M.
- Empregos na indústria
- ligações d'agua.

Assim, o segundo fator representa adequadamente a *situação econômica e industrial do município*.

Portanto, o índice obtido, representa uma resultante soma ponderada das situações do município e de sua população.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Afifi, A.A. and Azen, S.P. (1972) - Statistical Analysis, A Computer Oriented Approach, Academic Press, New York.
- Anderson, T.W. (1958) - An Introduction to Multivariate Statistical Analysis, Wiley, New York.
- Anderson, T.W. and Rubin, H. (1956) - "Statistical Inference in Factor Analysis", Proc. Third Berkeley Symp. Math. Statist. Probab., 5, 111-150.
- Bartlett, M.S. (1962) - "Factor Analysis in Psychology as a Statistician sees it", in Essays on Probability and Statistics, Wiley, New York.
- Burt, C. (1941) - The Factor of the Mind; An Introduction to Factor Analysis in Psychology, MacMillan, New York.
- Dubois, P.H. (1957) - Multivariate Correlational Analysis, Harper, New York.
- Girshick, M.A. (1936) - "Principal Componentes", J. Amer. Statist. Assoc., 31, 519-528.
- Girshick, M.A. (1939) - "On the Sampling Theory of Roots of Determinantal Equations". Annals Math. Statist., 10, 203-224.
- Guilford, J.P. (1956) - Fundamental Statistics in Psychology and Education, McGraw-Hill, New York.
- Harman, H.H. (1960) - Modern Factor Analysis, Univ. Chicago Press, Chicago.

- Hotelling, H. (1935) - "Analysis of a Complex of Statistical Variables into Principal Components", J. Educ. Psychol., 24, 417-441.
- Howe, W.G. (1955) - Some Contributions to Factor Analysis, USAEC, Rep. ORNL - 1919 (Ph.D. Thesis, Univ. of North Carolina).
- Jöreskog, K.G. (1966) - "UMFLA - A Computer Program for Unrestricted Maximum Likelihood Factor Analysis", Research Memorandum, 66, 20, Princeton, N.J., Educational Testing Service.
- Jöreskog, K.G. and Gruvaeus, G. (1967) - "RMLFA - A Computer Program for Restricted Maximum Likelihood Factor Analysis", Research Memorandum, 67, 21, Princeton, N.J., Educational Testing Service.
- Kaiser, H.F. (1958) - "The Varimax Criterion for Analytic Rotation in Factor Analysis", Psychometrika, 23, 187-200.
- Kendall, M.G. (1950) - "Factor Analysis as a Statistical Technique", J.R. Statist. Soc., B, 12, 60-73.
- Kendall, M.G. and Lawley, D.N. (1956) - "The Principles of Factor Analysis", J.R. Statist. Soc., A, 119, 83-84.
- Lawley, D.N. (1940) - "The Estimation of Factor Loadings by the Method of Maximum Likelihood", Proc. R. Soc. Edinb., A, 60, 64-82.
- Lawley, D.N. and Maxwell, A.E. (1971) - Factor Analysis as a Statistical Method, Butterworths, London.
- Lawley, D.N. and Maxwell, A.E. (1973) - "Regression and Factor Analysis", Biometrika, 60, 331-338.
- Maxwell, A.E. - "Recent Trends in Factor Analysis", J.R. Statist. Soc., A, 124, 49-59. (1961)
- Maxwell, A.E. and Humphreys, L.G. (1974) - Analysis Factorial,

- Enciclop. Internacional de las Ciencias Sociais, Vol.1,  
Ed. Aguilar, Madri.
- Morettin, P. (1975)-Análise Multivariada, Notas de Aula,IME,  
USP, São Paulo.
- Morrison, D.F. (1967) - Multivariate Statistical Methods,  
McGraw-Hill, New York.
- Nenhaus, J. and Wrigley, C. (1954) - "The Quartimax Method:  
An Analytical Approach to Ortogonal Simple Structure",  
British J. Statist. Psychol., 7, 81-91.
- Pearson, K. (1901)-"On Lines and Planes of Closest Fit to  
Systems of Points in Space", Phil. Mag., 6, 559-572.
- Press, J. (1972)-Applied Multivariate Analysis, Holt, Rine-  
hart and Winston, New York.
- Rao, C.R. (1955) - "Estimation and Tests of Significance in  
Factor Analysis", Psychol., 20, 93-111.
- Rao, C.R. (1973) - Linear Statistical Inference and its Ap-  
plications, Wiley, New York.
- Spearman, C. (1904) - "General Intelligence Objectively De-  
termined and Measured", Amer.J.Psychol., 15, 201-293.
- Tatsuoka,M.M. (1971)-Multivariate Analysis: Techniques for  
Educational and Psychological Research, Wiley, N. York.
- Thomson, G.H. (1951) - The Factorial Analysis of Human Abi-  
lity, London University Press.
- Thurstone, L.L. (1935) - The Vectors of Mind, University of  
Chicago Press, Chicago.
- Thurstone, L.L. (1947) - Multiple Factor Analysis, Universi-  
ty of Chicago Press, Chicago.
- Whittle,P. (1953)- "On Principal Component and Least Squares  
Method of Factor Analysis" Skand.Aktuarietrdsknift, 35,  
223-239.