

O MODELO DE ANÁLISE FATORIAL

Lucia Silva Kubrusly

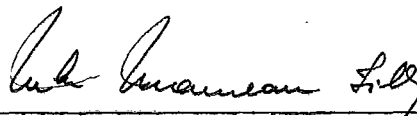
TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS (M.Sc.)

Aprovada por:



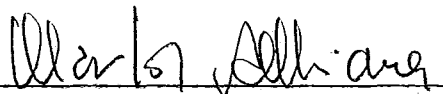
---

Prof. Claudio T. Bornstein



---

Prof. Nelson Maculan Filho



---

Prof. Marlos A. G. Viana

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 1981

SUMÁRIO

Esse trabalho apresenta um enfoque didático do modelo de Análise Fatorial e alguns métodos de solução. Para a solução ortogonal do modelo, são apresentados os métodos do Resíduo Mínimo e dos Fatores Principais. São apresentados ainda dois métodos de rotação ortogonal: o método Quartimax e o método Varimax. Outros métodos de solução ortogonal assim como métodos de rotação oblíqua são apenas mencionados no final do trabalho, a título de informação.

ABSTRACT

This thesis presents a didactic approach for the Factor Analysis Model, and some solutions methods. The methods of Minimum Residuals and Principal Factors are discussed in order to get the model orthogonal solution. Moreover, it is also presented another two methods related to the orthogonal rotation, namely, the Quartimax and Varimax methods. Others orthogonal solution methods, as well as oblique rotation methods, are also slightly mentioned at the end of the present work for sake of completeness.

ÍNDICE

	<u>Pág.</u>
<u>CAPÍTULO 1 - APRESENTAÇÃO</u> .....	1
<u>CAPÍTULO 2 - O MODELO DE ANÁLISE FATORIAL</u> .....	7
2.1 - Apresentação do Modelo.....	11
2.2 - Conceitos Preliminares.....	13
2.3 - Padrão Fatorial.....	19
2.4 - Notação Matricial.....	22
2.5 - Indeterminação do Modelo.....	28
- Síntese do Capítulo 2.....	29
<u>CAPÍTULO 3 - SOLUÇÃO ORTOGONAL</u> .....	31
3.1 - Método Resíduo Mínimo.....	31
3.2 - Método Fatores Principais.....	43
3.3 - Estimativa das Comunalidades e Número de Fatores... 74	74
3.4 - Exemplo.....	87
3.5 - Rotação Ortogonal.....	98
<u>CAPÍTULO 4 - OUTROS MÉTODOS DE ANÁLISE FATORIAL</u> .....	128
4.1 - Outros Métodos Ortogonais.....	128
4.2 - Rotação Oblíqua.....	135
- Considerações Finais.....	143

APÊNDICE I..... 145

APÊNDICE II..... 149

REFERÊNCIAS..... 155

BIBLIOGRAFIA ADICIONAL..... 160

## CAPÍTULO 1

### ANÁLISE MULTIVARIADA: APRESENTAÇÃO

Um estudo estatístico envolve geralmente um conjunto de elementos do mesmo tipo (ex. pessoas, regiões, entidades), sobre os quais se tomam medidas. Esse conjunto de elementos é chamado população.

Seja uma população de  $N$  elementos e suponha que em cada um deles sejam feitas  $n$  medidas. A Análise Multivariada é formada por uma coleção de métodos que têm como objetivo analisar problemas estatísticos através de diversas ( $n$ ) medidas em uma população finita ( $N$ ).

- Alguns exemplos de aplicação da Análise Multivariada em diferentes áreas.

Agricultura - Um número  $N$  de regiões, cada uma produzindo  $n$  produtos agrícolas. Queremos comparar a produtividade entre as regiões, ou determinar quais produtos são os melhores indicadores dessa produtividade.

Sociologia - São aplicados questionários a membros de uma população e espera-se que as respostas variem de acordo com a classe social do indivíduo. São colhidas informações de aluguel, posse de telefone, tipo de educação. Como poderemos construir um índice de classes sociais a partir desses dados ?

Antropologia - De um conjunto de tribos de Índios, uma série de itens são investigados: se possuem um deus da chuva, se possuem alguma agricultura, se a forma de união é monogâmica. O objetivo do estudo é estabelecer um critério a partir desses dados, que informe se uma determinada tribo pertence ou não a um certo grupo étnico, ou indique a qual grupo pertence.

Medicina - Um determinado medicamento é injetado em cobaias que são posteriormente mortas e examinadas. Vários órgãos são afetados. Quantos desses efeitos são significativos? Quais os órgãos mais afetados?

Educação - São aplicados questionários sobre o procedimento do professor, a um grupo de alunos. A partir desses dados, apontar qual o procedimento que o professor deve adotar para ser melhor compreendido pelos alunos.

Esses exemplos ilustram o vasto campo de aplicação da Análise Multivariada. Os métodos utilizados na solução de problemas como esses, podem ser divididos em dois grupos:

1º grupo - métodos que estudam as relações entre os elementos (conjunto de medidas). Ex.: Correlação Canônica, Análise de Variância, Análise de Correlação, etc.

2º grupo - métodos que estudam as relações entre as medidas feitas em cada elemento. Exemplo: Análise das Componentes principais, Análise Fatorial, etc.

Este trabalho considera o 2º grupo de métodos acima mencionados, mais precisamente a Análise Fatorial. Será feita uma apresentação detalhada desse método, e sempre que possível, serão consideradas comparações com a Análise das Componentes Principais. As idéias básicas dos dois métodos serão discutidas abaixo.

- Análise das Componentes Principais (A.C.P.) [14]

A partir das  $n$  variáveis originais ( $x_1, x_2, \dots, x_n$ ) (i.ê., das  $n$  medidas feitas em cada elemento), obtenho  $n$  componentes principais ( $y_1, y_2, \dots, y_n$ ) tais que as componentes principais  $y_j$  sejam combinações lineares das variáveis originais  $x_i$ ,  $i = 1, 2, \dots, n$ , isto é,

$$y_j = a_{j1} x_1 + a_{j2} x_2 + \dots + a_{jn} x_n$$

$$j = 1, 2, \dots, n.$$

Essas novas componentes  $y_j$ , são calculadas de forma que a 1ª componente principal  $y_1$  terá variância máxima. A 2ª componente principal  $y_2$  terá variância máxima e não será correlacionada com a 1ª. A 3ª componente principal terá variância máxima e não será correlacionada nem com  $y_1$  nem com  $y_2$ . E assim sucessivamente até  $y_n$ , que terá variância máxima e não será correlacionada com  $y_1, y_2, \dots, y_{n-1}$ . Dessa forma, a variância das componentes  $y_j$  decresce com  $j$  (desde que as restrições de não-correlação impostas a  $y_j$  aumentam com  $j$ ) e frequentemente as 1ªs  $m \ll n$  componentes principais contribuem com a parte mais significativa na variância total. Nesse caso é pos-



sível desprezar (comentando-se um erro denominado erro sistemático) as últimas componentes principais, reduzindo a dimensão do problema original.

#### - Análise Fatorial (A.F.)

Ao contrário do modelo de Análise de Componentes principais, o modelo de Análise Fatorial procura reproduzir da "melhor forma possível" a correlação entre as variáveis originais.

Parte-se da hipótese de que cada uma das  $n$  variáveis originais é função de um certo número  $m \ll n$  de fatores comuns a todas elas, e de um fator específico, que caracteriza a variável em questão. Ou seja:

$$x_i = a_{i1} f_1 + \dots + a_{im} f_m + d_i u_i, \quad i = 1, 2, \dots, n.$$

onde  $f_p$  são os fatores comuns e  $u_i$  é o fator específico da variável  $x_i$ .

Veremos mais tarde que no modelo de Análise Fatorial descrito acima, a variância das variáveis originais fica dividida em duas partes: parte comum (comunalidades) e parte específica (variância única).

O objetivo da Análise Fatorial é o seguinte: das  $n$  variáveis  $(x_1, x_2, \dots, x_n)$  e a matriz das correlações  $R$ .

$$R = \begin{vmatrix} 1 & r_{12} & \dots & r_{1n} \\ r_{21} & 1 & \dots & r_{2n} \\ \vdots & & & \\ r_{n1} & \dots & \dots & 1 \end{vmatrix}$$

determinar os coeficientes dos fatores ( $a_{ip}$ ,  $d_i$ ) que "melhor" reproduzam as correlações entre as variáveis.

Nos métodos de solução do modelo de Análise Fatorial geralmente é preciso fazer uma estimativa a priori ou das communalidades, ou do número  $m$  de fatores comuns utilizados na descrição de cada variável.

- Comparações dos Métodos A.C.P. e A.F.:

A.F.

- Reproduz as correlações entre as variáveis
- Reduz a dimensão do problema sem ocorrência do erro sistemático.
- Fatores comuns não são obrigatoriamente ortogonais (não correlacionados); pode-se obter uma solução oblíqua.

A.C.P.

- Reproduz a variância total.
- Não reduz a dimensão do problema, a não ser comentando erro sistemático.
- Fatores não correlacionados (solução ortogonal)

A.F.

- Obtenção dos fatores como função das variáveis através de regressão múltipla.
- Modelo indeterminado; permite escolher uma solução tal que facilite a interpretação dos resultados.

A.C.P.

- Obtenção dos fatores como função das variáveis diretamente da solução.
- Solução única; interpretação dos fatores raramente satisfatória.

## CAPÍTULO 2

### O MODELO DE ANÁLISE FATORIAL

Quando estudamos um fenômeno qualquer que envolve um grande número de variáveis, gostaríamos de reduzir a dimensão do problema, ou seja diminuir o número de variáveis envolvidas, perdendo o mínimo de informação possível. A Análise Fatorial fornece um método analítico para se conseguir isso através do estudo das correlações das variáveis. Na verdade do ponto de vista da Análise Fatorial, o que importa em cada variável medida, é a sua contribuição para o problema geral. A característica própria de cada variável, não interessa porque a Análise Fatorial é uma técnica que estuda as correlações, procurando agrupar as variáveis altamente correlacionadas. Por isso, a Análise Fatorial é muitas vezes usada como uma técnica de classificação das variáveis. Um problema onde a análise dos dados é feita com esse objetivo será apresentado no exemplo 2.1.

De um modo geral no entanto, o problema da Análise Fatorial é, a partir de  $n$  variáveis originais, determinar um número  $m$  de novas variáveis, que represente o antigo conjunto de dados. No caso unidimensional, uma variável aleatória é caracterizada por sua média e variância. No caso  $n$ -dimensional, o que caracteriza um certo conjunto de variáveis é a média e variância de cada uma, e a correlação entre elas. Como veremos adiante, o modelo de Análise Fatorial supõe variáveis aleatórias padronizadas (média = 0, variância = 1), então o

conjunto fica caracterizado apenas pelas correlações entre as variáveis. Daí no caso da A.F., quando procuramos um novo conjunto de  $m \ll n$  variáveis para caracterização do problema, reproduzimos as correlações originais, garantindo assim que o novo conjunto de variáveis é equivalente ao original.

O primeiro modelo de Análise Fatorial, apareceu em 1904 quando Charles Spearman [29] fez um estudo estatístico a partir de graus obtidos em  $n$  disciplinas, por vários alunos. Nesse trabalho, Spearman supôs que o grau obtido em cada disciplina podia ser descrito como sendo função de um fator comum, nesse caso interpretado como "inteligência" de um modo geral, mais um fator específico que poderia ser entendido como a aptidão do aluno numa dada disciplina. Ou seja, denotando

$x_i$ : grau obtido na disciplina  $i$ ,  $i = 1, 2, \dots, n$

$f$ : fator comum ("inteligência")

$u_i$ : fator específico (aptidão do aluno na disciplina  $i$ )

temos:

$$x_i = a_i f + u_i$$

Verificou-se no entanto, que tal modelo não reproduzia de forma conveniente as correlações entre as variáveis observadas e partiu-se daí para sua generalização, ou seja o modelo de múltiplos fatores comuns.

Antes de apresentar o modelo geral, vamos através de um exemplo, tentar mostrar a idéia básica da Análise Fatorial.

Suponha que são feitas  $n$  medidas em uma amostra de  $N$  pessoas. Se algumas dessas medidas se correlacionam, podemos grupá-las e definir uma variável fundamental que seria formada pela contribuição das variáveis originais que pertencem a esse grupo. De um modo geral, são obtidos diversos grupos, cada um formado por variáveis que se correlacionam entre si. Podemos ver como isso ocorre no exemplo abaixo.

Exemplo 2.1 - Tirado de uma pesquisa feita em 1941 por Burt e Banks com medidas de nove ( $n = 9$ ) dimensões do corpo humano. A amostra é formada por 2400 homens e as correlações entre as medidas são fornecidas na tabela 2.1.:

Tabela 2.1 - Matriz das Correlações

Variáveis	1	2	3	4	5	6	7	8	9
1. Altura	1.								
2. Altura tronco	.81	1.							
3. Comp.braço	.59	.67	1.						
4. Comp.perna	.58	.29	.19	1.					
5. Comp.coxa	.53	.25	.16	.67	1.				
6. Cintura	.17	.13	.08	.23	.29	1.			
7. Quadril	.33	.39	.28	.17	.22	.70	1.		
8. Ombro	.22	.29	.21	.08	.12	.52	.59	1.	
9. Peso	.40	.41	.29	.28	.33	.77	.83	.62	1.

Assim, em termos das medidas escolhidas, uma descrição das dimensões físicas de qualquer indivíduo, requer o conhecimento dos valores das nove variáveis. Mas se prestarmos atenção nas altas correlações entre as medidas 1 e 2, 2 e 3,

1.e 3, temos a impressão que as medidas 1, 2, 3 estão "juntas", parecem formar um grupo de variáveis razoavelmente relacionadas e suspeitamos que talvez não seja necessário tomar os valores para todas elas, já que podem representar diferentes aspectos de uma variável fundamental, digamos "comprimento do corpo".

Fazendo o mesmo tipo de análise sobre as outras correlações, teremos um 2º grupo destacado: variáveis 4 e 5 que poderia ser visto como "comprimento abaixo da cintura". E da mesma forma, um 3º grupo formado pelas variáveis 6, 7, 8 e 9 fornecendo "massa do corpo".

Inspeções desse tipo não são obviamente satisfatórias mas podem nos dar uma idéia inicial e intuitiva do que é a Análise Fatorial. Grosseiramente falando essas variáveis fundamentais são os fatores comuns nos quais estamos interessados, e a Análise Fatorial fornece um método formal de especificar quantos são esses fatores, e o quanto cada variável original é de fato uma medida para cada um deles.

Mostramos a seguir a solução obtida pela Análise Fatorial. Os valores que aparecem na Tabela 2.2 são os coeficientes de cada fator na descrição de cada variável.

Tabela 2.2 - Padrão Fatorial para o exemplo 2.1

Variável	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
1. Altura	-0,18	0,54	<u>0,71</u>
2. Alt.tronco	-0,20	0,21	<u>0,90</u>
3. Comp.braço	-0,15	0,10	<u>0,78</u>
4. Comp.perna	-0,04	<u>0,81</u>	0,21
5. Comp.coxa	-0,18	<u>0,78</u>	0,10
6. Cintura	<u>-0,79</u>	0,28	-0,13
7. Quadril	<u>-0,88</u>	0,06	0,16
8. Ombro	<u>-0,70</u>	-0,03	0,16
9. Peso	<u>-0,87</u>	0,22	0,19

Confirmando a análise prévia das correlações dos dados, três fatores foram extraídos, e considerando os coeficientes mais altos de cada fator, podemos ver os três grupos distintos de variáveis. F<sub>1</sub> corresponde ao grupo formado pelas variáveis 6, 7, 8 e 9; F<sub>2</sub> corresponde ao grupo 4 e 5; F<sub>3</sub> corresponde ao grupo 1, 2 e 3.

## 2.1 - APRESENTAÇÃO DO MODELO

Em 1931 Thurstone [31] generalizou o modelo de Spearman com um fator comum, sugerindo o modelo de fatores múltiplos que será estudado agora. Esse modelo de Análise Fatorial faz uma descrição teórica de cada uma das  $n$  variáveis  $x_i$  (a linha  $i$  é usada para distinguir a variável reproduzida pelo



modelo, do valor observado  $x_i$ ) como função linear (\*) de um certo número  $m \ll n$  de fatores  $f$ , mais um fator específico  $u_i$  de tal forma que a correlação entre as variáveis seja reproduzida da "melhor maneira possível". Isto é, uma vez encontrada a solução do modelo, as correlações entre as variáveis obtidas devem coincidir com as correlações entre as variáveis observadas. Então, para verificar se a solução é aceitável, não comparamos a variável obtida com a observada, mas sim as respectivas correlações. O referido modelo é matematicamente descrito da seguinte forma:

$$x_i' = \sum_{p=1}^m a_{ip} f_p + d_i u_i \quad (i = 1, 2, \dots, n)$$

onde a primeira e a segunda parcela do lado direito da equação são denominadas de parte comum e parte específica respectivamente, e as variáveis envolvidas são:

$x_i'$ :  $i$ -ésima variável

$f_p$ :  $p$ -ésimo fator comum

$u_i$ : fator específico que caracteriza a variável  $x_i'$

$a_{ip}$ : contribuição do fator  $f_p$  na descrição da variável  $x_i'$ .

À primeira vista pode parecer que o número de variáveis do modelo de A.F. aumentou, já que inicialmente ti-

\* trataremos aqui do modelo linear de Análise Fatorial. Modelos não-lineares foram também estudados. Ver [21] e [22].

nhamos  $n$  variáveis, e agora temos  $m$  fatores comuns, mais  $n$  fatores específicos. Mas o que realmente ocorre é que na solução do modelo sō estamos interessados na parte comum (que descreve a correlação das variáveis) ou seja, nos  $m \ll n$  fatores comuns. Os fatores específicos sō aparecem para separar a variância da variável em duas partes (comum e específica) como ficará claro mais adiante.

O modelo como foi apresentado acima, não levou em consideração o fato de termos vários ( $N$ ) elementos nos quais sō feitas as  $n$  medidas, isto é teremos para cada elemento  $j$  ( $j = 1, 2, \dots, N$ ) a variável  $x_{ij}$  (medida da  $i$ -ésima variável sobre o elemento  $j$ ) que será descrita da seguinte forma:

$$x_{ij} = \sum_{p=1}^m a_{ip} f_{pj} + d_i u_{ij}$$

onde:

$f_{pj}$ : valor do fator  $p$  para o elemento  $j$ .

$u_{ij}$ : valor do fator específico que caracteriza a variável  $x_{ij}$  medida no elemento  $j$ .

## 2.2 - CONCEITOS PRELIMINARES

Apresentaremos, a seguir, alguns conceitos básicos de estatística que serão utilizados na descrição do modelo de A.F.

- Média de uma variável  $i$  medida em  $N$  elementos

$$\bar{x}_i = \frac{1}{N} \sum_{j=1}^N x_{ij} \quad i = 1, 2, \dots, n.$$

- Desvio de uma variável  $x_{ij}$  (mede quanto a variável  $x_{ij}$  se afasta da média)

$$X_{ij} = x_{ij} - \bar{x}_i$$

- Variância de  $x_{ij}$

$$S_i^2 = \frac{1}{N} \sum_{j=1}^N X_{ij}^2$$

- Desvio-padrão

$$S_i = + \sqrt{S_i^2}$$

- Variável padronizada

$$z_{ij} = \frac{X_{ij}}{S_i}$$

Variável padronizada  $z_i = (z_{ij}, j = 1, 2, \dots, N)$ .

. Média da variável padronizada:

$$\bar{z}_i = \frac{1}{N} \sum_{j=1}^N z_{ij} = \frac{1}{N} \sum_{j=1}^N \frac{x_{ij} - \bar{x}_i}{S_i}$$

$$\bar{z}_i = \frac{1}{NS_i} \left| \sum_{j=1}^N x_{ij} - \sum_{j=1}^N \bar{x}_i \right| =$$

$$\bar{z}_i = \frac{1}{NS_i} |N \bar{x}_i - N \bar{x}_i| = 0.$$

. Variância de uma variável padronizada.

$$S_{z_i}^2 = \frac{1}{N} \sum_{j=1}^N (z_{ij} - \bar{z}_i)^2$$

como  $\bar{z}_i = 0$ :

$$S_{z_i}^2 = \frac{1}{N} \sum_{j=1}^N z_{ij}^2, \text{ onde } z_{ij} = \frac{X_{ij}}{S_i}$$

$$S_{z_i}^2 = \frac{1}{N} \sum_{j=1}^N \frac{X_{ij}^2}{S_i^2} = \frac{1}{S_i^2} \frac{1}{N} \sum_{j=1}^N X_{ij}^2 = \frac{1}{S_i^2} S_i^2$$

$$S_{z_i}^2 = 1$$

Portanto concluímos que a variável padronizada  $\bar{z}$  é definida de tal forma que sua média seja nula e sua variância seja igual a um.

- Covariância entre duas variáveis  $\underline{i}$  e  $\underline{k}$

$$S_{ik} = \frac{1}{N} \sum_{j=1}^N X_{ij} X_{kj}$$

- Coeficiente de correlação

$$r_{ik} = S_{ik}/S_i S_k = \frac{1}{N} \sum_{j=1}^N (X_{ij} X_{kj})/S_i S_k = \frac{1}{N} \sum_{j=1}^N z_{ij} z_{kj}$$

Na descrição e solução do modelo de A.F., todas as variáveis, assim como todos os fatores comuns e específicos, são supostos padronizados. O modelo supõe ainda, que todos os fatores específicos não são correlacionados entre si, nem com nenhum dos fatores comuns, isto é

$$r_{u_i u_k} = \begin{cases} 0 & \text{se } i \neq k \\ 1 & \text{se } i = k \end{cases}$$

$$r_{u_i f_p} = 0 \quad \forall i, p.$$

Reescrevendo o modelo para variáveis padronizadas:

$$z'_{ij} = a_{i1} f_{1j} + \dots + a_{im} f_{mj} + d_i u_{ij}, \quad \begin{matrix} i = 1, 2, \dots, m \\ j = 1, 2, \dots, N \end{matrix} \quad (2.1)$$

- Variância de  $z'_i$ :

Pela definição de variância temos:

$$S^2_{z'_i} = \sum_{j=1}^N (z'_{ij} - \bar{z}'_i)^2 / N$$

como  $z'_i$  é uma variável padronizada:  $\bar{z}'_i = 0$  e  $S^2_{z'_i} = 1$ . Então

$$S^2_{z'_i} = \frac{1}{N} \sum_{j=1}^N z'^2_{ij} = 1$$

Substituindo  $z'_{ij}$  por (2.1) temos:

$$S^2_{z'_i} = \sum_{p=1}^m a_{ip}^2 \left( \sum_{j=1}^N f_{pj}^2 \right) / N + d_i^2 \sum_{j=1}^N u_{ij}^2 / N +$$

$$+ 2 \sum_{p=1}^{m-1} \sum_{q=p+1}^m a_{ip} a_{iq} \left( \sum_{j=1}^N f_{pj} f_{qj} \right) / N + 2d_i \sum_{p=1}^m a_{ip} \left( \sum_{j=1}^N f_{pj} u_{ij} \right) / N = 1$$

Pelas definições de variância e coeficiente de correlação sabemos que:

$$\frac{1}{N} \left( \sum_{j=1}^N f_{pj}^2 \right) = S_{f_p}^2 = 1 \text{ (porque } f_p \text{ é suposto padronizado)}$$

$$\frac{1}{N} \left( \sum_{j=1}^N u_{ij}^2 \right) = S_{u_i}^2 = 1 \text{ (pelo mesmo motivo)}$$

$$\frac{1}{N} \left( \sum_{j=1}^N f_{pj} f_{qj} \right) = r_{f_p f_q}$$

$$\frac{1}{N} \left( \sum_{j=1}^N f_{pj} u_{ij} \right) = r_{f_p u_i} = 0 \text{ (por imposição do modelo)}$$

Então:

$$S^2_{z'_i} = \sum_{p=1}^m a_{ip}^2 + d_i^2 + 2 \sum_{p=1}^{m-1} \sum_{q=p+1}^m a_{ip} a_{iq} r_{f_p f_q} = 1$$

Os fatores comuns podem ou não ser correlacionados entre si. Por enquanto trataremos do caso em que os fato-

res comuns não são correlacionados. Assim temos:

$$r_{f_p f_q} = \begin{cases} 0 & \text{se } p \neq q \\ 1 & \text{se } p = q \end{cases} \text{ e podemos escrever:}$$

$$S_{z_i}^2 = \sum_{p=1}^m a_{ip}^2 + d_i^2 = 1$$

Ou seja, a variância fica dividida em duas partes: uma parte devida aos fatores comuns chamada comunalidade ( $h_i^2 = \sum_{p=1}^m a_{ip}^2$ ), e a outra devida aos fatores específicos chamada variância única ( $d_i^2$ ). Uma vez definidas a comunalidade e a variância única de uma variável, sua variância total é dada da seguinte forma:

$$S_{z_i}^2 = h_i^2 + d_i^2 = 1 \quad (2.2)$$

A partir da expressão acima, tendo-se determinada a comunalidade de uma variável  $z_i$ , a sua variância única fica determinada, bem como o coeficiente do fator específico relativo a essa variável:

$$d_i = \sqrt{1 - h_i^2}$$

Podemos concluir portanto que uma vez determinados os coeficientes dos fatores comuns de uma variável  $z_i$ , é possível calcular sua comunalidade, sua variância única e o coeficiente do fator específico correspondente.

### 2.3 - PADRÃO FATORIAL

No modelo de A.F. para  $n$  variáveis,

$$z'_1 = a_{11}f_1 + a_{12}f_2 + \dots + a_{1m}f_m + d_1u_1$$

$$z'_2 = a_{21}f_1 + a_{22}f_2 + \dots + a_{2m}f_m + 0 + d_2u_2$$

$$\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots$$

$$z'_n = a_{n1}f_1 + a_{n2}f_2 + \dots + a_{nm}f_m + 0 + \dots + d_nu_n$$

chama-se Padrão Fatorial ao conjunto de coeficientes  $a_{ip}$  ( $i = 1, 2, \dots, n$ ;  $p = 1, 2, \dots, m$ ) dos fatores comuns.

A partir da descrição das  $n$  variáveis teóricas em função dos fatores comuns e específicos, podemos calcular a correlação entre quaisquer duas variáveis  $z'_i$  e  $z'_k$ .

- Coeficiente de correlação entre as variáveis  $z'_i$  e  $z'_k$  descritas pelo modelo (coeficiente de correlação reproduzido,  $r'_{ik}$ ):

$$r'_{ik} = \frac{\sum_{j=1}^N z'_{ij} z'_{kj}}{N}$$

onde:

$$z'_i = a_{i1}f_1 + \dots + a_{im}f_m + d_iu_i$$

$$z'_k = a_{k1}f_1 + \dots + a_{km}f_m + d_ku_k$$



$$r'_{ik} = \frac{1}{N} \left[ \sum_{j=1}^N (a_{i1} a_{k1} f_{1j}^2 + \dots + a_{im} a_{km} f_{mj}^2 + d_i d_k u_{ij} u_{kj} + \right. \\ \left. + a_{i1} f_{1j} a_{k2} f_{2j} + a_{i1} f_{1j} a_{k3} f_{3j} + \dots + a_{i1} f_{1j} a_{km} f_{mj} + a_{i1} f_{1j} d_k u_{kj} + \right. \\ \left. + \dots + d_i u_{ij} a_{k1} f_{1j} + \dots + d_i u_{ij} a_{km} f_{mj}) \right]$$

$$r'_{ik} = a_{i1} a_{k1} \frac{1}{N} \sum_{j=1}^N f_{1j}^2 + \dots + a_{im} a_{km} \frac{1}{N} \sum_{j=1}^N f_{mj}^2 + d_i d_k \frac{1}{N} \sum_{j=1}^N u_{ij} u_{kj} \\ + a_{i1} a_{k2} \frac{1}{N} \sum_{j=1}^N f_{1j} f_{2j} + \dots + a_{i1} d_k \frac{1}{N} \sum_{j=1}^N f_{1j} u_{kj} + \dots \\ + d_i a_{k1} \frac{1}{N} \sum_{j=1}^N u_{ij} f_{1j} + \dots + d_i a_{km} \frac{1}{N} \sum_{j=1}^N u_{ij} f_{mj}$$

Pelas definições de variância e coeficiente de correlação podemos escrever:

$$r'_{ik} = a_{i1} a_{k1} S_{f_1}^2 + \dots + a_{im} a_{km} S_{f_m}^2 + d_i d_k r_{u_i u_k} + a_{i1} a_{k2} r_{f_1 f_2} + \\ \dots + a_{i1} d_k r_{f_1 u_k} + \dots + d_i a_{k1} r_{u_i f_1} + \dots + d_i a_{km} r_{u_i f_m} .$$

Mais uma vez, como estamos trabalhando com variáveis padronizadas:

$$S_{f_p}^2 = 1, p = 1, 2, \dots, m$$

É ainda, por imposição do modelo:

$$r_{u_i u_k} = 0 \text{ se } i \neq k$$

$$r_{u_i f_p} = 0 \quad \forall i, p.$$

Além disso, supondo os fatores comuns não correlacionados:

$$r_{f_p f_q} = \begin{cases} 0 & \text{se } p \neq q \\ 1 & \text{se } p = q \end{cases}$$

Podemos escrever:

$$r_{ik}^1 = a_{i1} a_{k1} + \dots + a_{im} a_{km} . \quad (2.3)$$

O coeficiente de correlação reproduzido, no caso dos fatores comuns não correlacionados, é função apenas dos coeficientes dos fatores comuns.

Uma maneira de verificar a validade da solução obtida, é calcular as correlações entre as variáveis observadas (correlação observada  $r_{ik}$ ) e as correlações reproduzidas  $r_{ik}^1$  e comparar os resultados. A correlação residual ( $\bar{r}_{ik} = r_{ik} - r_{ik}^1$ ) deve ser próxima de zero para a solução ser aceitável.

Concluimos então que o objetivo da Análise Fatorial é encontrada um Padrão Fatorial  $a_{ip}$  ( $i = 1, 2, \dots, n$ ;  $p = 1, 2, \dots, m$ ) que "melhor" reproduza as correlações entre as variáveis.

Como uma vez obtido o padrão fatorial, pela Eq. (2.2) podemos obter a descrição completa (fator comum + fator específico), é usual representar-se apenas a parte comum do modelo, ou seja:

$$z_i'' = \sum_{p=1}^m a_{ip} f_p \quad (\text{parte comum})$$

A modificação decorrente de se utilizar apenas a parte comum aparece no cálculo de  $r_{kk}'$ , cujo resultado, ao invés de dar a variância total da variável ( $= 1$ ), dará apenas a parte comum dessa variância, ou seja, a comunalidade da variável  $z_k'$ .

#### 2.4 - NOTAÇÃO MATRICIAL

O modelo de A.F. apresentado descreve a seguinte situação: um conjunto de  $N$  elementos sobre os quais são feitas  $n$  medidas. Podemos considerar as  $n$  variáveis reproduzidas pelo modelo como um vetor-coluna de dimensão  $n$ :

$$z' = \begin{pmatrix} z_1' \\ \vdots \\ z_n' \end{pmatrix}$$

A cada um dos  $N$  elementos corresponde um vetor-coluna de dimensão  $n$ , de tal forma que ao conjunto de todos os dados (todas as variáveis para todos os elementos) corresponde uma matriz  $Z'$  ( $n \times N$ ) chamada matriz dos dados reproduzidos:

$$Z' = \begin{pmatrix} z'_{11} & z'_{12} & \dots & z'_{1N} \\ z'_{21} & z'_{22} & \dots & z'_{2N} \\ \vdots & \vdots & & \vdots \\ z'_{n1} & z'_{n2} & \dots & z'_{nN} \end{pmatrix}$$

Da mesma forma, para os fatores comuns e específicos temos:

Para um elemento:

$$f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix} \quad e \quad u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

Para N elementos

$$F = \begin{vmatrix} f_{11} & f_{12} & \dots & f_{1N} \\ \vdots & \vdots & & \vdots \\ f_{m1} & f_{m2} & \dots & f_{mN} \end{vmatrix} \quad \text{e} \quad U = \begin{vmatrix} u_{11} & u_{12} & \dots & u_{1N} \\ \vdots & \vdots & & \vdots \\ u_{n1} & u_{n2} & \dots & u_{nN} \end{vmatrix}$$

Os coeficientes dos fatores comuns e específicos na forma matricial:

$$A = \begin{vmatrix} a_{11} & \dots & a_{1m} \\ a_{21} & \dots & a_{2m} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nm} \end{vmatrix}$$

A: matriz padrão (fornece o padrão fatorial)

$$D = \begin{vmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & d_n \end{vmatrix}$$

O modelo, utilizando a notação matricial, fica:

$$z' = \begin{bmatrix} A & D \end{bmatrix} \begin{bmatrix} f \\ u \end{bmatrix} = Af + Du$$

Levando em conta somente a parte comum do modelo:

$$z'' = Af \quad (\text{parte comum})$$

ou ainda, para os N elementos:

$$Z'' = AF.$$

Se  $r_{ik}$  ( $i, k = 1, 2, \dots, n$ ) são as correlações entre todas as variáveis observadas, R é a matriz das correlações observadas. Como

$$r_{ik} = \frac{1}{N} \sum_{j=1}^N z_{ij} z_{kj} \quad (2.4)$$

e  $r_{kk} = 1$  (variância total da variável padronizada  $z_k$ ) temos:

$$R = \begin{vmatrix} 1 & \dots & r_{1n} \\ r_{21} & 1 \dots & r_{2n} \\ \vdots & & \vdots \\ r_{n1} & \dots & 1 \end{vmatrix}$$

Substituindo  $r_{ik}$  pela equação (2.4) podemos escrever:

$$R = \frac{1}{N} \begin{vmatrix} 1 & \sum_{j=1}^N z_{1j} z_{2j} & \dots & \sum_{j=1}^N z_{1j} z_{nj} \\ \sum_{j=1}^N z_{2j} z_{1j} & 1 & \dots & \sum_{j=1}^N z_{2j} z_{nj} \\ \vdots & & & \\ \sum_{j=1}^N z_{nj} z_{1j} & \dots & & 1 \end{vmatrix} = \frac{1}{N} ZZ^t$$

onde  $Z$  é a matriz das variáveis observadas, e  $Z^t$  a sua transposta.

Para calcular as correlações das variáveis reproduzidas pelo método temos:

$$R' = \frac{1}{N} Z'Z'^t$$

Se levarmos em conta apenas a parte comum do modelo,  $Z'' = AF$ , e substituirmos na equação acima, obteremos a Matriz das Correlações Reduzida.

$$R^* = AF(AF)^t/N = AFF^tA^t/N$$

Pela definição de coeficiente de correlação, a matriz  $FF^t/N$  é a matriz das correlações entre os fatores comuns. Supondo que tais fatores não são correlacionados, isto é:

$$r_{f_p f_q} = \begin{cases} 0 & \text{se } p \neq q \\ 1 & \text{se } p = q \end{cases} \quad p, q = 1, 2, \dots, m$$

A matriz das correlações  $FF^t/N = I$  (matriz identidade). Nesse caso, chegamos à relação entre a matriz das correlações reproduzidas e a matriz padrão fatorial, supondo fatores ortogonais, apresentada por Thurstone [32]:

$$R^* = AA^t .$$

Cabe lembrar mais uma vez que a matriz  $R^*$  foi obtida a partir da parte comum do modelo, portanto os elementos da sua diagonal  $r'_{kk}$  que forneceria(m) (caso fosse usado o modelo completo) a variância total da variável  $z'_k$ , nesse caso forne-



cem a parte comum da variância, ou seja a comunalidade da variável  $z'_k$ .

$$R^* = \begin{vmatrix} h_1^2 & r'_{12} & \dots & r'_{1n} \\ r'_{21} & h_2^2 & \dots & r'_{2n} \\ \vdots & \vdots & & \vdots \\ r'_{n1} & r'_{n2} & \dots & h_n^2 \end{vmatrix}$$

O objetivo da Análise Fatorial é encontrar a matriz  $A$  tal que a matriz  $R^*$  reproduza da "melhor maneira possível" os elementos da matriz  $R$  fora da diagonal principal.

## 2.5 - INDETERMINAÇÃO DO MODELO

Suponha que se tenha resolvido o modelo de A.F. e que tenha sido encontrada a matriz  $A$  (matriz padrão tal que  $AA^t = R^*$ ).

Seja  $T$  uma matriz ortogonal, isto é:

$$TT^t = I.$$

Então dizemos que  $B = AT$  é uma transformação ortogonal da matriz  $A$ . Neste caso se  $A$  é solução do modelo de Análise Fatorial,  $B$  também o é, porque:

$$BB^t = A^t(AT)^t = A^tT^tA^t = AA^t = R^*$$

Dada uma matriz padrão  $A$  que seja solução ortogonal do modelo de A.F., podemos obter uma infinidade de soluções que igualmente satisfazem as condições impostas pelo modelo, bastando fazer uma transformação ortogonal na solução original. Através do modelo de Análise Fatorial obtemos o espaço de solução (dimensão do espaço que é igual ao número de fatores comuns) mas não fica determinada qual base desse espaço melhor descreve o problema. Tal indeterminação será mais tarde explorada, para se obter uma nova solução, em que a interpretação dos fatores comuns seja mais razoável.

No próximo Capítulo trataremos da solução desse modelo. Chamamos atenção porém, que os problemas de inferência estatística, tais como estimar a matriz de correlação a partir de uma amostra de dados, testar a hipótese sobre o número de fatores comuns, e a própria estimativa desses fatores  $f_p$  não serão abordados e podem ser vistos em [23]. Trataremos apenas da formulação dos métodos e sua solução algébrica.

## - Síntese do Capítulo 2

Foi apresentada nesse capítulo a modelagem geral do problema que a Análise Fatorial pretende resolver. Diversos métodos para solução do problema foram propostos. Existem, no entanto alguns pontos básicos comuns a todos esses métodos. De um modo geral, o modelo fica da seguinte forma:

## - Modelo Geral

Otimizar  $f(A) \rightarrow$  a função  $f$  varia com cada método

s.a:

$$1) z' = \sum_{p=1}^m a_{ip} f_p + d_i u_i \quad i = 1, 2, \dots, n$$

$$2) \sum_{p=1}^m a_{ip}^2 = h_i^2 \leq 1.0, \quad i = 1, 2, \dots, n$$

3)  $f_p, z'_i, u_i$ , são variáveis padronizadas.

4) ajuste as correlações reproduzidas com as correlações observadas.

Ou seja, essas condições são comuns a todos os métodos de Análise Fatorial. São condições que caracterizam o problema que a Análise Fatorial pretende resolver. No próximo capítulo, apresentaremos dois métodos de resolução direta: o método do Resíduo Mínimo, e o método de Fatores Principais. Esses dois métodos supõem soluções ortogonais, ou seja, o conjunto de fatores  $f_1, f_2, \dots, f_m$  forma uma base ortogonal, ou do ponto de vista estatístico, esses fatores não são correlacionados entre si.

### CAPÍTULO 3

#### SOLUÇÃO ORTOGONAL DO MODELO DE ANÁLISE FATORIAL

Até aqui demos uma idéia da Análise Fatorial como sendo uma técnica desenvolvida no sentido de obter a partir de um conjunto de  $n$  variáveis,  $m$  ( $m < n$ ) fatores comuns de tal maneira que as correlações entre as variáveis reproduzidas pelo modelo coincidam com as correlações entre as variáveis observadas. Existem diversos métodos de solução para o modelo, cada um interpretando de maneira diferente essa "reprodução das correlações observadas". Serão apresentados dois métodos para a obtenção de uma solução inicial ortogonal: o método do Resíduo Mínimo e o método de Fatores Principais.

#### 3.1 - MÉTODO DO RESÍDUO MÍNIMO

Através desse método busca-se ajustar as correlações reproduzidas com as observadas minimizando a diferença entre elas. Essa diferença

$$\bar{r}_{ik} = r_{ik} - r'_{ik}$$

é chamada correlação residual. O desenvolvimento seguido foi apresentado por Harman e Jones [11].

Nesse método comparamos os valores observados com os reproduzidos, e como não temos valores observados para as comunalidades, comparamos apenas os elementos fora da diagonal principal da matriz das correlações. Ou seja, busca-se minimizar

$$\bar{r}_{ik} = r_{ik} - r'_{ik} \quad \text{para } i \neq k.$$

Na verdade o objetivo desse método é tornar os termos  $\bar{r}_{ik}$  da matriz das correlações residuais próximo de zero. Portanto a minimização será sobre o valor absoluto das correlações residuais.

$$\min ||\bar{r}_{ik}|| = ||r_{ik} - r'_{ik}||$$

Os termos  $r_{ik}$  são tirados diretamente dos dados, enquanto que  $r'_{ik}$  é dada pela expressão (2.3):

$$r'_{ik} = \sum_{p=1}^m a_{ip} a_{kp},$$

ou seja depende do padrão fatorial encontrado. Nesse ponto é bom lembrar que a expressão (2.3) para calcular as correlações reproduzidas, só é válida se os fatores extraídos forem ortogonais. Então ao definirmos as correlações observadas pela expressão (2.3) estamos impondo a condição de ortogonalidade na solução. Usando essa definição para a correlação reproduzida, o objetivo do método do Resíduo Mínimo é encontrar um padrão

fatorial  $a_{ip}$  ( $i = 1, 2, \dots, n$ ;  $p = 1, 2, \dots, m$ ) tal que:

$$\min ||\bar{r}_{ik}|| = \min ||r_{ik} - \sum_{p=1}^m a_{ip} a_{kp}|| \quad \forall i \neq k.$$

Mas  $\min ||r_{ik} - \sum_{p=1}^m a_{ip} a_{kp}|| = \min \sqrt{(r_{ik} - \sum_{p=1}^m a_{ip} a_{kp})^2}$  que equivale a

$$\min \bar{r}_{ik}^2 = (r_{ik} - \sum_{p=1}^m a_{ip} a_{kp})^2$$

como vamos minimizar todos os elementos  $||\bar{r}_{ik}||^2$ , passamos a minimizar a soma de todos os elementos  $||\bar{r}_{ik}||^2$  da matriz dos resíduos. Então podemos escrever a função objetivo do método do Resíduo Mínimo

$$\min f(a) = \sum_{i=1}^{n-1} \sum_{k=i+1}^n (r_{ik} - \sum_{p=1}^m a_{ip} a_{kp})^2 .$$

Vemos que para se implementar esse método, precisamos do número  $m$  de fatores comuns. Ou seja, precisamos fazer uma estimativa a priori do número de fatores necessários. A medida que o número  $m$  cresce, o problema torna-se muito indeterminado, e provavelmente encontra-se uma solução com resíduo zero. Mas como o objetivo da Análise Fatorial é diminuir a dimensão do problema, normalmente se escolhe  $m \ll n$ . A maneira de se fazer essa estimativa será descrita na seção 3.3. Uma vez obtido o padrão fatorial  $a_{ip}$  ( $i = 1, 2, \dots, n$ ;  $p = 1, 2, \dots, m$ ) podemos calcular as comunalidades de cada variável pela expres

são

$$h_i^2 = \sum_{p=1}^m a_{ip}^2$$

Mas como estamos trabalhando com variáveis padronizadas, o valor das comunalidades não pode exceder a unidade. Portanto essa restrição será imposta na solução do problema.

Podemos então escrever o modelo matemático segundo as hipóteses do método do Resíduo Mínimo:

Dado  $m$ , determinar

$a_{ip}$  ( $i = 1, 2, \dots, n$ ;  $p = 1, 2, \dots, m$ ) tal que

$$\min f(a) = \sum_{i=1}^{n-1} \sum_{k=i+1}^n (r_{ik} - \sum_{p=1}^m a_{ip} a_{kp})^2$$

onde: 1)  $a_{ip}$  ( $i = 1, 2, \dots, n$ ;  $p = 1, 2, \dots, m$ ) é o padrão factorial supondo fatores ortogonais, isto é:

$$r_{ik}^1 = \sum_{p=1}^m a_{ip} a_{kp}$$

$$2) z_i^1 = \sum_{p=1}^m a_{ip} f_p + d_i u_i$$

$$3) 0 \leq h_i^2 = \sum_{p=1}^m a_{ip}^2 \leq 1.0$$

4)  $f_p, z_i^1, u_i$  são variáveis padronizadas

Daí concluímos que as características básicas desse método são:

- estimativas a priori do número de fatores comuns ( $m$ ).
- as comunalidades aparecem como um resultado do método
- os fatores comuns são supostos ortogonais.

Vários métodos foram desenvolvidos para resolver esse problema, e o mais eficiente deles, o método iterativo de Gauss-Seidel será mostrado a seguir. Nesse processo iterativo, partimos de uma matriz padrão  $A$ , e alteramos a 1.<sup>a</sup> linha da matriz obtendo uma outra matriz  $B^1$ . Essa alteração é feita de modo que o valor da função objetivo melhore. A seguir tomando a matriz  $B^1$ , alteramos a 2.<sup>a</sup> linha da matriz e obtemos  $B^2$ . Assim sucessivamente até alterarmos todas as linhas da matriz inicial, obtendo  $B^n$ , e completando uma iteração (ou ciclo). Esses ciclos se repetem até que o valor da função objetivo permaneça constante (a menos de uma tolerância dada) de uma iteração para outra. Passemos a aplicação do método. Inicialmente antes de qualquer alteração temos:

$$z_i^1 = a_{i1}f_1 + a_{i2}f_2 + \dots + a_{im}f_m$$

Uma vez alterada a  $i$ -ésima linha ( $i = 1, 2, \dots, n$ ), temos:

$$z_i^1 = b_{i1}f_1 + b_{i2}f_2 + \dots + b_{im}f_m$$

onde  $b_{ip} = a_{ip} + \epsilon_{ip}$  para  $p = 1, 2, \dots, m$ .



Precisamos então determinar o vetor  $\epsilon_i(\epsilon_{i1}, \epsilon_{i2}, \dots, \epsilon_{in})$  tal que melhore ao máximo a função objetivo. Uma vez alterada a  $i$ -ésima linha da matriz inicial, as correlações entre as variáveis reproduzidas também se alteram:

$$r'_{ik} = \sum_{p=1}^m b_{ip} a_{kp}$$

Essa alteração naturalmente se refletirá na função objetivo  $f(a)$ .

$$f(a) = \sum_{k=1}^{n-1} \sum_{j=k+1}^n (r_{kj} - \sum_{p=1}^m a_{kp} a_{jp})^2 \quad (\text{antes da alteração})$$

A função objetivo  $f(a)$  terá parcelas que incluem os coeficientes  $a_{ip}$  ( $i$  fixo) e parcelas que não os incluem. Uma vez feita a alteração na linha  $i$ , apenas as parcelas que dependem de  $a_{ip}$  ( $i$  fixo,  $p = 1, 2, \dots, m$ ) serão alteradas. Chamando essas parcelas de  $f_i$ , passamos então a minimizar a função  $f_i$  relativa a alteração na  $i$ -ésima linha. Essa função será composta de todos os termos de  $f(a)$  que incluam  $a_{ip}$  ( $i = \text{fixo}$ ,  $p = 1, 2, \dots, m$ ). Assim  $f_i$  fica determinada da seguinte forma. Seja:

$$f(a) = \sum_{k=1}^{n-1} \sum_{j=k+1}^n (r_{kj} - \sum_{p=1}^m a_{kp} a_{jp})^2$$

Vamos expandir essa expressão separando os termos onde apare-

cem  $a_{ip}$  ( $i$  fixo,  $p = 1, 2, \dots, m$ ). Primeiro fazemos  $k \neq i$  e somamos a parcela correspondente a  $k = i$  separadamente:

$$f(a) = \sum_{\substack{k=1 \\ k \neq i}}^{n-1} \sum_{j=k+1}^n (r_{kj} - \sum_{p=1}^m a_{kp} a_{jp})^2 + \sum_{j=i+1}^n (r_{ij} - \sum_{p=1}^m a_{ip} a_{jp})^2$$

Fazendo o mesmo para  $j \neq i$ :

$$f(a) = \sum_{\substack{k=1 \\ k \neq i}}^{n-1} \sum_{\substack{j=k+1 \\ j \neq i}}^n (r_{kj} - \sum_{p=1}^m a_{kp} a_{jp})^2 + \sum_{k=1}^{i-1} (r_{ki} - \sum_{p=1}^m a_{kp} a_{ip})^2 + \\ + \sum_{j=i+1}^n (r_{ij} - \sum_{p=1}^m a_{ip} a_{jp})^2$$

Assim, a 2.<sup>a</sup> e 3.<sup>a</sup> parcela de  $f(a)$  contêm todos termos que dependem de  $a_{ip}$  ( $i$  fixo,  $p = 1, 2, \dots, m$ ) formando a função  $f_i$ .

$$f_i = \sum_{k=1}^{i-1} (r_{ki} - \sum_{p=1}^m a_{kp} a_{ip})^2 + \sum_{j=i+1}^n (r_{ij} - \sum_{p=1}^m a_{ip} a_{jp})^2.$$

Podemos juntar esses dois somatórios fazendo  $j = k$  e lembrando que  $r_{ik} = r_{ki}$ :

$$f_i = \sum_{\substack{k=1 \\ k \neq i}}^n (r_{ik} - \sum_{p=1}^m a_{kp} a_{ip})^2 \quad i \text{ fixo}$$

Como definimos  $f_i$  como função objetivo a ser minimizada ao alteramos a  $i$ -ésima linha da matriz padrão inicial, substituiremos o valor de  $a_{ip}$  pelo seu valor alterado  $b_{ip}$ . Então

$$f_i = \sum_{\substack{k=1 \\ k \neq i}}^n (r_{ik} - \sum_{p=1}^m a_{kp} b_{ip})^2 \quad i \text{ fixo}$$

Substituindo

$$b_{ip} = a_{ip} + \varepsilon_{ip}$$

$$f_i = \sum_{\substack{k=1 \\ k \neq i}}^n (r_{ik} - \sum_{p=1}^m a_{kp} (a_{ip} + \varepsilon_{ip}))^2 = \sum_{\substack{k=1 \\ k \neq i}}^n (r_{ik} - \sum_{p=1}^m a_{kp} a_{ip} - \sum_{p=1}^m a_{kp} \varepsilon_{ip})^2$$

Chamando  $\bar{r}_{ik} = r_{ik} - \sum_{p=1}^m a_{kp} a_{ip}$  (correlação residual antes da alteração)

$$f_i = \sum_{\substack{k=1 \\ k \neq i}}^n (\bar{r}_{ik} - \sum_{p=1}^m a_{kp} \varepsilon_{ip})^2$$

Para minimizar  $f_i$  derivamos em relação a  $\varepsilon_{ip}$  ( $p = 1, 2, \dots, m$ ) e igualamos a zero. Tomando um determinado valor de  $p = q$  temos:

$$\frac{\partial f_i}{\partial \varepsilon_{iq}} = 2 \sum_{\substack{k=1 \\ k \neq i}}^n (\bar{r}_{ik} - \sum_{p=1}^m a_{kp} \varepsilon_{ip}) (-a_{kq}) = 0 \quad (q = 1, 2, \dots, m)$$

$$\frac{\partial f_i}{\partial \varepsilon_{iq}} = \sum_{\substack{k=1 \\ k \neq i}}^n (-\bar{r}_{ik} a_{kq} + \sum_{p=1}^m a_{kp} a_{kq} \varepsilon_{ip}) = 0 \quad (q = 1, 2, \dots, m)$$

$$\sum_{\substack{k=1 \\ k \neq i}}^n \sum_{p=1}^m a_{kp} a_{kq} \varepsilon_{ip} = \sum_{\substack{k=1 \\ k \neq i}}^n \bar{r}_{ik} a_{kq} \quad (q = 1, 2, \dots, m)$$

Separando  $\epsilon_{ip}$

$$\sum_{p=1}^m \epsilon_{ip} \left( \sum_{\substack{k=1 \\ k \neq i}}^n a_{kp} a_{kq} \right) = \sum_{\substack{k=1 \\ k \neq i}}^n \bar{r}_{ik} a_{kq} \quad (q = 1, 2, \dots, m)$$

Se definirmos

$\epsilon_i = (\epsilon_{i1}, \epsilon_{i2}, \dots, \epsilon_{im}) \rightarrow$  vetor de acréscimo para a linha  $i$ .

$A_i =$  matriz  $A$  cujos elementos da  $i$ -ésima linha são substituídos por zeros.

$\bar{r}_i^0 = (\bar{r}_{i1}, \bar{r}_{i2}, \dots, \bar{r}_{ii-1}, 0, \bar{r}_{ii+1}, \dots, \bar{r}_{in}) \rightarrow$  vetor das correlações residuais da variável  $z_i$  com todas as demais variáveis, substituindo  $\bar{r}_{ii}$  por zero).

Podemos escrever a expressão acima como:

$$\epsilon_i A_i^t A_i = \bar{r}_i^0 A \text{ ou ainda}$$

$$\epsilon_i = \bar{r}_i^0 A (A_i^t A_i)^{-1}$$

Obtendo uma expressão para o vetor acréscimo relativo a  $i$ -ésima linha, que minimiza  $f_i$ .

Esse processo se repete para todas as linhas até formar um ciclo (ou iteração). Serão feitas tantas iterações quantas forem necessárias até que o valor da função objetivo total  $f(a)$  não se altere (a menos de uma tolerância). A

última matriz obtida é o padrão fatorial procurado. Isto é, aquele que minimiza  $f(a)$ . Mas o método de Gauss-Seidel não garante que as comunalidades das variáveis sejam valores entre zero e um. Portanto, a solução obtida só será viável se ao calcularmos todas as comunalidades, nenhuma delas exceder a unidade. Caso algum desses valores for maior que um, diz-se que ocorreu o caso de Heywood. O procedimento então é o seguinte: toma-se a linha (ou linhas) correspondente a variável cuja comunalidade excedeu um, e torna-se a resolver o problema, levando em consideração a restrição sobre as comunalidades. Isto é, seja uma variável  $z_i$  cuja comunalidade  $h_i^2 > 1$ . Então temos o seguinte problema:

$$\min f_i = \sum_{\substack{k=1 \\ k \neq i}}^n (r_{ik} - \sum_{p=1}^m a_{kp} b_{ip})^2$$

s.a.:

$$\sum_{p=1}^m b_{ip}^2 \leq 1.0 \quad (3.1)$$

Para resolver esse problema, o primeiro passo é transformar a restrição de desigualdade em uma igualdade. Para isso utiliza-se o resultado do seguinte teorema:

"Se o mínimo de  $f_i$  é atendido num ponto fora da região definida por (3.1), então o mínimo de  $f_i$  sob a restrição (3.1) será atendido num ponto de contorno dessa região, e pode-se substituir essa restrição por:

$$\sum_{p=1}^m b_{ip}^2 = 1.0$$

A prova desse teorema, bem como toda a solução para o caso de Heywood esta em [12]. Uma vez provado o teorema o problema a ser resolvido passa a ser;

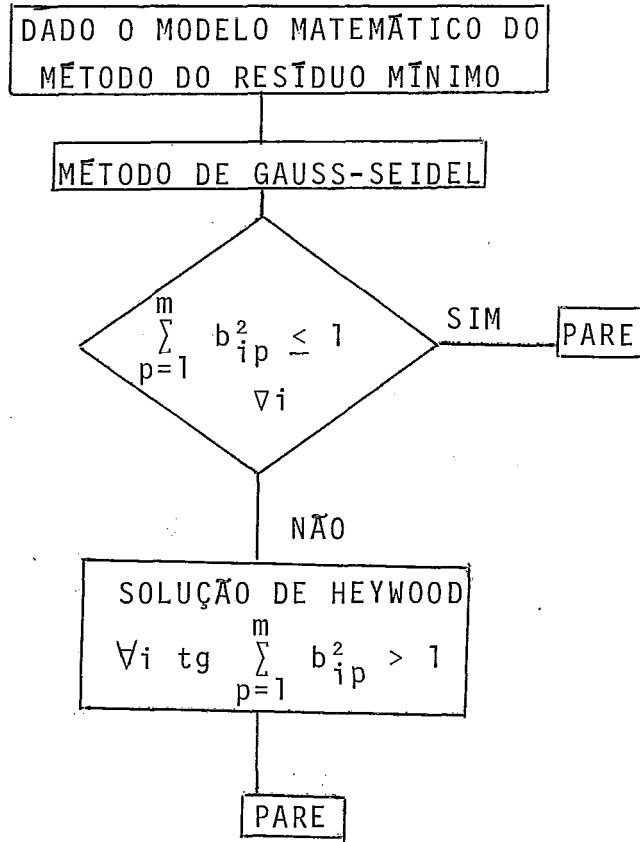
$$\min f_i = \sum_{\substack{k=1 \\ k \neq i}}^n (r_{ik} - \sum_{p=1}^m a_{kp} b_{ip})^2$$

s.a.

$$\sum_{p=1}^m b_{ip}^2 = 1.0$$

A solução é obtida utilizando-se multiplicadores de Lagrange e pode ser visto em [12].

Afim de sintetizar as idéias aqui apresentadas mostramos abaixo um fluxograma, com os principais passos para a solução do método de Resíduo Mínimo:



### 3.2 - O MÉTODO DOS FATORES PRINCIPAIS

Em 1934 Thomson [30] foi o primeiro a tentar adaptar o método de Análise de Componentes Principais no modelo de Análise Fatorial. O modelo de A.C.P. busca maximizar a variância total de cada componente de tal forma que as componentes sejam ortogonais, e a solução seja obtida a partir da fatoração da matriz das correlações dos dados [14]. No método de Fatores Principais, ao invés da variância total, cada fator será determinado de tal forma que sua contribuição na comunalidade total seja máxima sujeita a condição da correlação reproduzida ser igual a correlação observada (resíduo zero). Como vamos trabalhar com as comunalidades das variáveis, utilizaremos apenas a parte comum do modelo, e conseqüentemente na matriz das correlações observadas consideraremos as comunalidades na diagonal principal e não a variância total de cada variável (que seria igual a um).

Como vimos na seção 3.1, para solução do método do Resíduo Mínimo era necessária uma estimativa a priori do número de fatores comuns. No método de Fatores Principais, como vamos trabalhar com a matriz das correlações reduzidas, é necessária a estimativa a priori das comunalidades de todas as variáveis. Na verdade esses dois problemas estão ligados, e geralmente os métodos analíticos de Análise Fatorial precisam de uma das duas estimativas a priori. O problema das comunalidades e do número de fatores será descrito na seção 3.3.

Tomando a parte comum do modelo de A.F. temos:



$$z_1'' = a_{11} f_1 + a_{12} f_2 + \dots + a_{1m} f_m$$

$$z_2'' = a_{21} f_1 + a_{22} f_2 + \dots + a_{2m} f_m$$

⋮

$$z_n'' = a_{n1} f_1 + a_{n2} f_2 + \dots + a_{nm} f_m$$

- Comunalidade de uma variável:

$$h_i^2 = \sum_{p=1}^m a_{ip}^2$$

- Contribuição do p-ésimo fator para essa comunalidade:  $a_{ip}^2$

- Comunalidade total:  $\sum_{i=1}^n h_i^2 = \sum_{i=1}^n \sum_{p=1}^m a_{ip}^2$

- Contribuição do p-ésimo fator para comunalidade total

$$V_p = \sum_{i=1}^n a_{ip}^2$$

A restrição de que a correlação reproduzida se ajuste perfeitamente com a correlação observada fica da seguinte forma:

$$r_{ik} = r'_{ik} = \sum_{p=1}^m a_{ip} a_{kp} \quad i, k = 1, 2, \dots, n$$

onde  $r_{ik} = r_{ki}$  e  $r_{ii} = r_{ii} = h_i^2$  (comunalidade da  $i$ -ésima variável)

Nesse ponto lembramos mais uma vez que a definição das correlações reproduzidas

$$r'_{ik} = \sum_{p=1}^m a_{ip} a_{kp} \quad i, k = 1, 2, \dots, n$$

são válidas para fatores comuns ortogonais. Assim, a solução obtida para esse método também será ortogonal. Podemos então descrever o modelo matemático segundo as hipóteses do método dos Fatores Principais como a solução de  $m$  problemas análogos. São eles:

Problema 1: Determinar  $a_{i1}$   $i = 1, 2, \dots, n$ , tal que:

$$\max V_1 = \sum_{i=1}^n a_{i1}^2$$

s.a.

$$r_{ik} = \sum_{p=1}^m a_{ip} a_{kp} \quad i, k = 1, 2, \dots, n \quad i \neq k$$

Problema 2: Determinar  $a_{i2}$   $i = 1, 2, \dots, n$ , tal que:

$$\max V_2 = \sum_{i=1}^n a_{i2}^2$$

s.a.

$$r_{ik} = \sum_{p=1}^m a_{ip} a_{kp} \quad i, k = 1, 2, \dots, n$$

onde  $r_{ij}$  = comunalidade da  $i$ -ésima variável, estimada a priori.  
e  $a_{i1}$  ( $i = 1, 2, \dots, n$ ) são conhecidos através da solução do problema 1.

Problema m: Determinar  $a_{im}$   $i = 1, 2, \dots, n$ , tal que:

$$\max V_m = \sum_{i=1}^n a_{im}^2$$

s.a.

$$r_{ik} = \sum_{p=1}^m a_{ip} a_{kp} \quad i, k = 1, 2, \dots, n$$

onde  $r_{ij}$  = comunalidade da  $i$ -ésima variável estimada a priori  
e são conhecidos:

$a_{i1}$  ( $i = 1, 2, \dots, n$ ) através da solução do problema 1.

$a_{i2}$  ( $i = 1, 2, \dots, n$ ) através da solução do problema 2.

.....  
.....

$a_{im-1}$  ( $i = 1, 2, \dots, n$ ) através da solução do problema  $m-1$ .

São válidas ainda as seguintes hipótese para os  
m problemas:

1)  $a_{ip}$  ( $i = 1, 2, \dots, n$ ;  $p = 1, 2, \dots, m$ ) padrão fatorial su  
pondo fatores ortogonais.

$$2) z' = \sum_{p=1}^m a_{ip} f_p + d_i u_j$$

$$3) 0 \leq h_i^2 \leq 1.0 \quad \forall i \text{ (estimada a priori)}$$

4)  $f_p, z'_i, u_j$ , são variáveis padronizadas.

Portanto, as características básicas desse método são:

- estimativa das comunalidades das  $n$  variáveis
- os fatores comuns são supostos ortogonais.

### Solução do Método de Fatores Principais

A solução desse método consiste, como vimos até agora, na solução de  $m$  problemas de maximização com restrições. Passemos à solução do primeiro problema:

#### Problema 1 (p=1)

$$\max V_1 = a_{11}^2 + a_{21}^2 + a_{31}^2 + \dots + a_{n1}^2$$

s.a.

$$\sum_{p=1}^m a_{ip} a_{kp} = r_{ik} \quad i_1, k = 1, 2, \dots, n \quad (3.2)$$

Resolvendo este primeiro problema, estamos determinando o vetor coluna  $a_1$ , correspondente a primeira coluna da matriz A. Para resolver esse problema de maximização com res-

trição utilizaremos o método dos multiplicadores de Lagrange (veja apêndice |A.I|).

Seja a função a ser maximizada:

$$T = V_1 - \sum_{i=1}^n \sum_{k=1}^n \sum_{p=1}^m \mu_{ik} a_{ip} a_{kp} + \sum_{i=1}^n \sum_{k=1}^n \mu_{ik} r_{ik}$$

onde  $\mu_{ik} = \mu_{ki}$  são os multiplicadores de Lagrange

Derivando em relação a cada variável  $a_{ip}$  e igualando a zero.

Para  $p = 1$

$$\frac{\partial T}{\partial a_{i1}} = \frac{\partial V_1}{\partial a_{i1}} - \frac{\partial}{\partial a_{i1}} \sum_{j=1}^n \sum_{k=1}^n \sum_{p=1}^m \mu_{jk} a_{jp} a_{kp} +$$

$$+ \frac{\partial}{\partial a_{i1}} \sum_{j=1}^n \sum_{k=1}^n \mu_{jk} r_{jk} = 0$$

$$\text{mas } \frac{\partial}{\partial a_{i1}} \sum_{j=1}^n \sum_{k=1}^n \mu_{jk} r_{jk} = 0, \text{ então}$$

$$\frac{\partial T}{\partial a_{i1}} = 2a_{i1} - \frac{\partial}{\partial a_{i1}} \sum_{j=1}^n \sum_{k=1}^n \sum_{p=1}^m \mu_{jk} a_{jp} a_{kp} = 0$$

Mas

$$\frac{\partial}{\partial a_{i1}} \sum_{j=1}^n \sum_{k=1}^n \sum_{p=1}^m \mu_{jk} a_{jp} a_{kp} = \frac{\partial}{\partial a_{i1}} \left( \sum_{k=1}^n \mu_{ik} a_{i1} a_{k1} + \right.$$

$$\left. + \sum_{k=1}^n \sum_{p=2}^m \mu_{ik} a_{ip} a_{kp} + \sum_{\substack{j=1 \\ j \neq i}}^n \mu_{ji} a_{j1} a_{i1} + \right.$$

$$+ \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{p=2}^m \mu_{ji} a_{jp} a_{ip} + \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{\substack{k=1 \\ k \neq i}}^n \sum_{p=1}^m \mu_{jk} a_{jp} a_{kp}$$

Dessa forma vemos que a derivada em relação a  $a_{i1}$  será zero para a 2.<sup>a</sup>, 4.<sup>a</sup> e 5.<sup>a</sup> parcelas. Então

$$\begin{aligned} \frac{\partial}{\partial a_{i1}} \sum_{j=1}^n \sum_{k=1}^n \sum_{p=1}^m \mu_{jk} a_{jp} a_{kp} &= \sum_{\substack{k=1 \\ k \neq i}}^n \mu_{ik} a_{k1} + 2a_{i1} \mu_{ii} + \\ &+ \sum_{\substack{j=1 \\ j \neq i}}^n \mu_{ji} a_{j1} \end{aligned}$$

Os dois somatórios podem ser escritos como um só, porque  $\mu_{ij} = \mu_{ji}$ , e a parcela intermediária incluída:

$$\frac{\partial}{\partial a_{i1}} \sum_{j=1}^n \sum_{k=1}^n \sum_{p=1}^m \mu_{jk} a_{jp} a_{kp} = 2 \sum_{k=1}^n \mu_{ik} a_{k1}$$

Assim:

$$\frac{\partial T}{\partial a_{i1}} = 2a_{i1} - 2 \sum_{k=1}^n \mu_{ik} a_{k1} = 0 \quad (3.3)$$

$$\sum_{k=1}^n \mu_{ik} a_{k1} = a_{i1} \quad (3.4)$$

Para  $p = 2, 3, \dots, m$ :

$$\frac{\partial T}{\partial a_{ip}} = \frac{\partial V_1}{\partial a_{ip}} - \frac{\partial}{\partial a_{ip}} \sum_{i=1}^n \sum_{k=1}^n \sum_{p=1}^m \mu_{ik} a_{ip} a_{kp} +$$

$$+ \frac{\partial}{\partial a_{ip}} \sum_{i=1}^n \sum_{k=1}^n \mu_{ik} r_{ik} = 0$$

Mas:

$$\frac{\partial V_1}{\partial a_{ip}} = 0 \text{ para } p \neq 1, \text{ al\em disso:}$$

$$\frac{\partial}{\partial a_{ip}} \sum_{i=1}^n \sum_{k=1}^n \mu_{ik} r_{ik} = 0$$

Ent\~ao utilizando racioc\~inio semelhante ao desenvolvido acima:

$$\frac{\partial T}{\partial a_{ip}} = - \sum_{k=1}^n \mu_{ik} a_{kp} = 0, \text{ para } p = 2, 3, \dots, m. \quad (3.5)$$

Derivando T em rela\~cao a  $\mu_{ik}$  e igualando a zero temos:

$$\frac{\partial}{\partial \mu_{ik}} (V_1 - \sum_{j=1}^n \sum_{\ell=1}^n \sum_{p=1}^m \mu_{j\ell} a_{jp} a_{\ell p} + \sum_{j=1}^n \sum_{\ell=1}^n \mu_{j\ell} r_{j\ell}) = 0$$

Desenvolvendo essa derivada da mesma forma que no caso anterior, temos:

$$\frac{\partial T}{\partial \mu_{ik}} = - \sum_{p=1}^m a_{ip} a_{kp} + r_{ik} = 0$$

o que nos leva a eq. (3.2)

$$\sum_{p=1}^m a_{ip} a_{kp} = r_{ik}$$

Podemos juntar as eqs. (3.3) e (3.5) numa expressão sãõ sãõ:

$$\frac{\partial T}{\partial a_{ip}} = \delta_{1p} a_{i1} - \sum_{k=1}^n \mu_{ik} a_{kp} = 0 \quad (p = 1, 2, \dots, m) \quad (3.6)$$

$$\text{onde } \delta_{1p} = \begin{cases} 1 & \text{se } p = 1 \\ 0 & \text{se } p \neq 1 \end{cases}$$

Multiplicando a eq. (3.6) por  $a_{i1}$  e somando para todo  $i$ :

$$\delta_{1p} \sum_{i=1}^n a_{i1}^2 - \sum_{i=1}^n \sum_{k=1}^n \mu_{ik} a_{i1} a_{kp} = 0$$

Mas pela eq. (3.4)  $\sum_{i=1}^n \mu_{ik} a_{i1} = a_{k1}$ , portanto

$$\delta_{1p} \sum_{i=1}^n a_{i1}^2 - \sum_{k=1}^n a_{k1} a_{kp} = 0$$

Chamando



$$\sum_{i=1}^n a_{i1}^2 = \lambda^1, \quad (3.7)$$

a equação acima fica da seguinte forma:

$$\delta_{1p} \lambda^1 - \sum_{k=1}^n a_{k1} a_{kp} = 0 \quad (3.8)$$

Multiplicando por  $a_{ip}$  e somando em  $p$ :

$$\sum_{p=1}^m a_{ip} \delta_{1p} \lambda^1 - \sum_{p=1}^m \sum_{k=1}^n a_{k1} a_{ip} a_{kp} = 0$$

Pela definição de  $\delta_{1p}$ , o primeiro termo da equação fica:

$$\sum_{p=1}^m a_{ip} \delta_{1p} \lambda^1 = a_{i1} \lambda^1$$

$$\text{Então } a_{i1} \lambda^1 - \sum_{k=1}^n a_{k1} \sum_{p=1}^m a_{ip} a_{kp} = 0$$

mas  $\sum_{p=1}^m a_{ip} a_{kp} = r_{ik}$ , então:

$$\sum_{k=1}^n r_{ik} a_{k1} - \lambda^1 a_{i1} = 0 \quad (3.9)$$

Temos  $n$  equações do tipo (3.9), uma para cada  $i$  ( $i = 1, 2, \dots, n$ ) que desenvolvidas formam o seguinte sistema homogêneo:

$$(r_{11} - \lambda^1) a_{11} + r_{12} a_{21} + \dots + r_{1n} a_{n1} = 0$$

$$r_{21} a_{11} + (r_{22} - \lambda^1) a_{21} + \dots + r_{2n} a_{n1} = 0$$

$$\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots$$

$$r_{n1} a_{11} + r_{n2} a_{21} + \dots + (r_{nn} - \lambda^1) a_{n1} = 0$$

Mas  $r_{ij} = h_i^2$  (comunalidade da  $i$ -ésima variável). Então podemos escrever:

$$(h_1^2 - \lambda^1) a_{11} + r_{12} a_{21} + \dots + r_{1n} a_{n1} = 0$$

$$r_{21} a_{11} + (h_2^2 - \lambda^1) a_{21} + \dots + r_{2n} a_{n1} = 0$$

$$\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots$$

$$r_{n1} a_{11} + r_{n2} a_{21} + \dots + (h_n^2 - \lambda^1) a_{n1} = 0$$

Nesse sistema de equações, queremos determinar  $\lambda^1$  e o vetor  $a_1$ . Os coeficientes  $r_{jk}$  são calculados a partir dos dados (correlação observada, que nesse modelo é exatamente igual a correlação reproduzida). Mas  $r_{ii} = h_i^2$  (porque estamos trabalhando com a parte comum do modelo) não pode ser calculada através dos dados e deve ser estimada a priori para se obter a solução do problema.

A maneira de se fazer essa estimativa será discutida na seção 3.3. Supondo portanto que são conhecidos os coe-

ficientes  $r_{ik}$  e  $h_i^2$ , o sistema homogêneo de  $n$  equações pode ser escrito utilizando a seguinte forma matricial:

$$R_1 = \begin{vmatrix} h_1^2 & r_{12} & \cdots & r_{1n} \\ r_{21} & h_2^2 & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ r_{n1} & r_{n2} & \cdots & h_n^2 \end{vmatrix} \quad \text{e} \quad a_1 = \begin{vmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{vmatrix}$$

Onde:

$R_1 \rightarrow$  matriz das correlações observadas reduzida, isto é, com as comunalidades das variáveis na diagonal principal e  $r_{ik}$  = correlação observada se  $i \neq k$ .

$a_1 \rightarrow$  vetor coluna a ser calculado

Então:

$$|R_1 - \lambda^1 I| a_1 = 0$$

$$R_1 a_1 = \lambda^1 a_1$$

Essa é uma equação de auto-valores e auto-vetores da matriz  $R_1$  (ver [A.II]). Para que tenha uma solução não trivial ( $a_1 \neq 0$ ), a condição necessária e suficiente é que:

$$\det |R_1 - \lambda^1 I| = 0$$

Se expandirmos esse determinante cairemos numa equação em  $\lambda^1$  de ordem  $n$ . Ou seja, obteremos  $n$  valores de  $\lambda^1$  (raízes da equação). Essas raízes são os auto-valores da matriz  $R_1$  cada um deles, uma vez substituídos no sistema de equações, dando uma solução  $a_1$  diferente ( $n$  auto-vetores). Mas pela equação (3.7),  $\lambda^1 = \sum_{i=1}^n a_{i1}^2 = V_1$  é a função objetivo do nosso problema. Portanto  $V_1$  é o maior auto-valor obtido na fatoração da matriz  $R_1$ .

Uma vez determinado  $\lambda^1$  temos, pela equação de auto-valores:

$$R_1 a_1 = \lambda^1 a_1$$

Para determinar  $a_1$ , auto-vetor associado ao auto-valor  $\lambda^1$ , resolvemos o seguinte sistema homogêneo:

$$(R_1 - \lambda^1 I) a_1 = 0$$

Na solução desse sistema obtemos uma família de vetores paralelos, todos eles solução da equação característica |A.II|. Para determinar qual deles é o vetor  $a_1$ , 1.<sup>a</sup> coluna da matriz A, adicionamos a restrição imposta pelo problema na eq. (3.7):

$$\lambda^1 = \sum_{i=1}^n a_{i1}^2 = ||a_1||^2$$

Assim fica determinado o vetor  $a_1$  que é a 1.<sup>a</sup> coluna da matriz

padrão A. Passemos portanto para o segundo problema.

Problema 2 ( $p = 2$ )

$$\max V_2 = a_{12}^2 + a_{22}^2 + a_{32}^2 + \dots + a_{n2}^2$$

s.a.

$$\sum_{p=1}^m a_{ip} a_{kp} = r_{ik} \quad i, k = 1, 2, \dots, n$$

Mas como  $a_{i1}$  já foi determinado, a restrição quanto a correlação  $r_{ik}$  fica:

$$r_{ik} = a_{i1} a_{k1} + \sum_{p=2}^m a_{ip} a_{kp}$$

Então:

$${}_2r_{ik} = r_{ik} - a_{i1} a_{k1} = \sum_{p=2}^m a_{ip} a_{kp} \quad (3.10)$$

onde  ${}_2r_{ik}$  é a correlação residual, descontada a correlação devida ao primeiro fator, isto é, a contribuição desse fator na correlação reproduzida.

A idéia é calcular o 1º fator e as correlações reproduzidas devido a ele. Calcula-se então a correlação residual. Se o resíduo for diferente de zero, passamos ao cálculo do 2º fator.

Reescrevendo o problema 2:

$$\max V_2 = \sum_{i=1}^n a_{i2}^2$$

s.a.

$$2^{r_{ik}} = \sum_{p=2}^m a_{ip} a_{kp} \quad i, k = 1, 2, \dots, n.$$

O problema é análogo ao anterior, apenas substituindo  $r_{ik}$  por  $2^{r_{ik}}$ . Resolvendo pelo mesmo método

$$T = V_2 - \sum_{i=1}^n \sum_{k=1}^n \mu_{ik} \sum_{p=2}^m a_{ip} a_{kp} + \sum_{i=1}^n \sum_{k=1}^n \mu_{ik} 2^{r_{ik}}$$

Derivando em relação a cada  $a_{ip}$  e  $\mu_{ik}$  e igualando a zero temos (ver problema 1)

$$\text{Para } p = 2 \quad \frac{\partial T}{\partial a_{i2}} = 2a_{i2} - 2 \sum_{k=1}^n \mu_{ik} a_{k2} = 0 \quad (3.11)$$

Para  $p = 3, 4, \dots, m$

$$\frac{\partial T}{\partial a_{ip}} = - \sum_{k=1}^n \mu_{ik} a_{kp} = 0$$

$$\frac{\partial T}{\partial \mu_{ik}} = - \sum_{p=2}^m a_{ip} a_{kp} + 2^{r_{ik}} = 0 \quad (3.10)$$

Assim, podemos escrever;

$$\delta_{2p} a_{i2} - \sum_{k=1}^n \mu_{ik} a_{kp} = 0 \quad (p = 2, 3, \dots, m)$$

$$\text{onde } \delta_{2p} = \begin{cases} 1 & \text{se } p = 2 \\ 0 & \text{se } p \neq 2 \end{cases}$$

Multiplicando por  $a_{i2}$  e somando para todo  $i$ :

$$\delta_{2p} \sum_{i=1}^n a_{i2}^2 - \sum_{i=1}^n \sum_{k=1}^n \mu_{ik} a_{i2} a_{kp} = 0$$

Mas pela eq. (3.11)

$$\sum_{i=1}^n \mu_{ik} a_{i2} = a_{k2} \quad \text{então:}$$

$$\delta_{2p} \sum_{i=1}^n a_{i2}^2 - \sum_{k=1}^n a_{k2} a_{kp} = 0$$

$$\text{chamando } \lambda^2 = \sum_{i=1}^n a_{i2}^2 \quad (3.12)$$

$$\delta_{2p} \lambda^2 - \sum_{k=1}^n a_{k2} a_{kp} = 0 \quad (3.13)$$

Multiplicando por  $a_{ip}$  e somando em  $p$

$$\sum_{p=2}^m a_{ip} \delta_{2p} \lambda^2 - \sum_{p=2}^m \sum_{k=1}^n a_{k2} a_{ip} a_{kp} = 0$$

ou ainda:

$$a_{i2} \lambda^2 - \sum_{k=1}^n a_{k2} \sum_{p=2}^m a_{ip} a_{kp} = 0$$

Mas por (3.10)  $\sum_{p=2}^m a_{ip} a_{kp} = 2^{r_{ik}}$

Então:

$$\sum_{k=1}^n 2^{r_{ik}} a_{k2} - \lambda^2 a_{i2} = 0 \quad i = 1, 2, \dots, n.$$

O sistema, para as  $n$  equações fica:

$$(2^{r_{11}} - \lambda^2) a_{12} + 2^{r_{12}} a_{22} + \dots + 2^{r_{1n}} a_{n2} = 0$$

$$2^{r_{21}} a_{12} + (2^{r_{22}} - \lambda^2) a_{22} + \dots + 2^{r_{2n}} a_{n2} = 0$$

$$\vdots \quad \quad \quad \vdots$$

$$2^{r_{n1}} a_{12} + 2^{r_{n2}} a_{22} + \dots + (2^{r_{nn}} - \lambda^2) a_{n2} = 0$$

Mas pela eq. (3.10)

$2^{r_{ii}} = r_{ii} - a_{i1}^2 = h_i^2 - a_{i1}^2 = 2h_i^2$  (comunalidade da  $i$ -ésima variável, descontada a contribuição do primeiro fator). Podemos reescrever o sistema:



$$(2h_1^2 - \lambda^2) a_{12} + 2r_{12} a_{22} + \dots + 2r_{1n} a_{n2} = 0$$

$$\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots$$

$$2r_{n1} a_{12} + 2r_{n2} a_{22} + \dots + (2h_n^2 - \lambda^2) a_{n2} = 0$$

Chegamos portanto num sistema de equações do mesmo tipo do primeiro problema apenas substituindo

$$r_{ik} \text{ por } 2r_{ik}$$

$$\lambda^1 \text{ por } \lambda^2$$

$$a_1 \text{ por } a_2$$

$$h_i^2 \text{ por } 2h_i^2$$

Pela eq. (3.12),  $\lambda^2$  é a quantidade a ser maximizada, portanto o problema agora é calcular o maior auto-valor da matriz  $R_2$  formada por todos os  $2r_{ik}$  ( $i, k = 1, \dots, n$ ).

Mas como  $2r_{ik} = r_{ik} - a_{i1} a_{k1}$ , a matriz  $R_2$  pode ser escrita através da seguinte relação:

$$R_2 = R_1 - a_1 a_1^t$$

$$\text{onde } a_1 a_1^t = \begin{vmatrix} a_{11} \\ \vdots \\ a_{n1} \end{vmatrix} |a_{11} \dots a_{n1}| =$$

$$= \begin{vmatrix} a_{11}^2 & a_{11} \cdot a_{21} & \dots & a_{11} \cdot a_{n1} \\ \vdots & & & \vdots \\ a_{n1} \cdot a_{11} & & \dots & a_{n1}^2 \end{vmatrix}$$

Provaremos a seguir que os auto-vetores associados a matriz  $R_1$ , também são auto-vetores da matriz  $R_2$ .

Seja  $\alpha_p$  um auto-vetor da matriz  $R_1$ . Então pela relação acima, podemos escrever:

$$R_2 \alpha_p = (R_1 - a_1 a_1^t) \alpha_p$$

$$R_2 \alpha_p = R_1 \alpha_p - a_1 a_1^t \alpha_p = \lambda_p^1 \alpha_p - a_1 a_1^t \alpha_p$$

onde  $\lambda_p^1$  é o  $p$ -ésimo auto-valor da matriz  $R_1$ .

Se  $p = 1$  teremos:

$$\alpha_1 = a_1$$

$$R_2 a_1 = \lambda_1^1 a_1 - a_1 a_1^t a_1$$

$$\text{mas } a_1^t a_1 = (a_{11} \ a_{21} \ \dots \ a_{n1}) \begin{vmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{vmatrix} = \sum_{i=1}^n a_{i1}^2 = \lambda_1^1$$

$$\text{Então: } R_2 a_1 = \lambda_1^1 a_1 - a_1 \lambda_1^1 = (\lambda_1^1 - \lambda_1^1) a_1 = 0 \cdot a_1$$

Ou seja, o auto-vetor  $a_1$  corresponde ao maior auto valor  $\lambda_1^1$  da matriz  $R_1$  também é auto-vetor de  $R_2$  mas o auto-valor  $\lambda_1^2$  associado é zero.

Se  $p \neq 1$ :

$a_1^t \alpha_p = 0$  porque os auto-vetores de  $R_1$  são ortogonais.

Então:

$$R_2 \alpha_p = R_1 \alpha_p - a_1 a_1^t \alpha_p$$

$$R_2 \alpha_p = \lambda_p^1 \alpha_p, \text{ para } p = 2, 3, \dots, m.$$

Concluimos portanto que:

- 1) Os auto-vetores de  $R_2$  são iguais aos auto-vetores de  $R_1$ .
- 2) Os auto-valores de  $R_2$  são iguais aos auto-valores de  $R_1$ , exceto para  $\lambda_1^1$ , porque  $\lambda_1^2 = 0$ . Conseqüentemente o maior auto-valor de  $R_2$  será o segundo maior auto-valor de  $R_1$  ( $\lambda_2^1$ ). Este será o valor da função objetivo do problema 2, e terá como auto-vetor associado  $a_2$ , que será a 2.<sup>a</sup> coluna da matriz A. Então para obtermos a solução do problema 2, basta calcular o segundo maior auto-valor de  $R_1$  e seu auto-vetor associado. Assim agrupamos os resultados obtidos para os dois problemas.

Problema 1 ( $p = 1$ )  $R_1 \rightarrow$  cálculo do maior auto-valor de  $R_1$  e correspondente auto-vetor  $a_1$

Solução:  $V_1 = \lambda_1^1 \rightarrow$  valor da função objetivo

$a_1 \rightarrow$  1.<sup>a</sup> coluna da matriz A.

Problema 2 ( $p = 2$ )  $R_2 \rightarrow$  cálculo do maior auto-valor de  $R_2$  e correspondente auto-vetor  $a_2$ . Nesse ponto foi provado que o maior auto-valor de  $R_2$  é o 2.<sup>o</sup> maior auto-valor de  $R_1$  e o auto-vetor correspondente é o mesmo para as duas matrizes.

Solução:  $V_2 = \lambda_2^1 \rightarrow$  valor da função objetivo

$a_2 \rightarrow$  2.<sup>a</sup> coluna da matriz A.

Problema m ( $p = m$ )

O mesmo procedimento se repete para os  $m$  problemas análogos ou seja, quando  $p = m$ , o problema será determinar  $V_m$  que será o maior auto-valor da matriz  $R_m$  formada por:

${}_m r_{ik} = r_{ik} - \sum_{p=1}^{m-1} a_{ip} a_{kp} \rightarrow$  correlação residual descontada a correlação devida aos  $m-1$  primeiros fatores.

${}_m r_{ii} = {}_m h_i^2 = h_i^2 - \sum_{p=1}^{m-1} a_{ip}^2 \rightarrow$  comunalidade da  $i$ -ésima variável descontada a contribuição dos  $m-1$  primeiros fatores.

Repetindo o mesmo procedimento dos dois problemas anteriores chegamos ao seguinte sistema de equações a ser resolvido:

$$({}_m h_1^2 - \lambda^m) a_{1m} + {}_m r_{12} a_{2m} + \dots + {}_m r_{1n} a_{nm} = 0$$

$$\vdots$$

$$\vdots$$

$${}_m r_{ni} a_{1m} + \dots + ({}_m h_n^2 - \lambda^m) a_{nm} = 0$$

Vamos agora provar que os auto-vetores associados a matriz  $R_1$  também são auto-vetores da matriz  $R_m$ .

A partir da definição de  ${}_m r_{ik}$  podemos escrever

$$R_m = R_1 - {}_{m-1}A {}_{m-1}A^t$$

onde

$${}_{m-1}A = \begin{vmatrix} a_{11} & \cdots & a_{1m-1} \\ \vdots & & \vdots \\ a_{1n} & \cdots & a_{nm-1} \end{vmatrix}$$

é a matriz formada pelos  $m-1$  primeiros vetores coluna de  $A$  calculados nos  $m-1$  problemas anteriores.

$$\text{Então } {}_{m-1}A {}_{m-1}A^t = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1m-1} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm-1} \end{vmatrix} \begin{vmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ \vdots & & & \vdots \\ a_{1m-1} & a_{2m-1} & \cdots & a_{nm-1} \end{vmatrix}$$

$$= \begin{vmatrix} \sum_{p=1}^{m-1} a_{1p}^2 & \sum_{p=1}^{m-1} a_{1p} a_{2p} & \cdots & \sum_{p=1}^{m-1} a_{1p} a_{np} \\ \vdots & & & \vdots \\ \sum_{p=1}^{m-1} a_{np} a_{1p} & \sum_{p=1}^{m-1} a_{np} a_{2p} & \cdots & \sum_{p=1}^{m-1} a_{np}^2 \end{vmatrix}$$

Seja  $\alpha_p$  um auto-vetor da matriz  $R_1$ . Podemos escrever para  $p = 1, 2, \dots, m-1$ :

$$\alpha_p = a_p \\ R_m a_p = (R_1 - {}_{m-1}A {}_{m-1}A^t) a_p = R_1 a_p - {}_{m-1}A {}_{m-1}A^t a_p$$

Mas

$$\begin{aligned}
 {}_{m-1}A^t a_p &= \begin{vmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & & & \vdots \\ a_{1m-1} & a_{2m-1} & \cdots & a_{nm-1} \end{vmatrix} \begin{vmatrix} a_{1p} \\ a_{2p} \\ \vdots \\ a_{np} \end{vmatrix} = \\
 &= \begin{vmatrix} \sum_{i=1}^n a_{i1} a_{ip} \\ \sum_{i=1}^n a_{i2} a_{ip} \\ \vdots \\ \sum_{i=1}^n a_{im-1} a_{ip} \end{vmatrix}
 \end{aligned}$$

E pela eq. (3.8)

$$\sum_{i=1}^m a_{i1} a_{ip} = \begin{cases} 0 & \text{se } p \neq 1 \\ \lambda^1 & \text{se } p = 1 \end{cases}$$

Pelo desenvolvimento do problema 2, eq. (3.13)

$$\sum_{i=1}^n a_{i2} a_{ip} = \begin{cases} 0 & \text{se } p \neq 2 \\ \lambda^2 & \text{se } p = 2 \end{cases}$$

Pelo desenvolvimento do problema (m-1):

$$\sum_{i=1}^n a_{i(m-1)} a_{ip} = \begin{cases} 0 & \text{se } p \neq m-1 \\ \lambda^{m-1} & \text{se } p = m-1 \end{cases}$$

$$\text{Então para } p = 1 \rightarrow {}_{m-1}A^t a_1 = \begin{vmatrix} \lambda^1 \\ 0 \\ \vdots \\ 0 \end{vmatrix}$$

$$p = 2 \rightarrow {}_{m-1}A^t a_2 = \begin{vmatrix} 0 \\ \lambda^2 \\ \vdots \\ 0 \end{vmatrix}$$

.....

$$p = m - 1 \rightarrow {}_{m-1}A^t a_{m-1} = \begin{vmatrix} 0 \\ \vdots \\ \lambda^{m-1} \end{vmatrix}$$



Então se  $p = 1$ :

$${}_{m-1}A \ {}_{m-1}A^t \ a_1 = \left| \begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1m-1} & \lambda^1 \\ \vdots & & & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm-1} & 0 \end{array} \right| = \left| \begin{array}{cc|c} a_{11} & \lambda^1 & \\ a_{21} & \lambda^1 & \\ \vdots & & \\ a_{n-1} & \lambda^1 & \end{array} \right| = \lambda^1 \ a_1$$

Para  $p = 2$ :

$${}_{m-1}A \ {}_{m-1}A^t \ a_2 = \left| \begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1m-1} & 0 \\ \vdots & & & \vdots & \lambda^2 \\ \vdots & & & \vdots & \vdots \\ a_{n-1} & a_{n2} & \dots & a_{nm-1} & 0 \end{array} \right| = \lambda^2 \ a_2$$

Para  $p = m-1$ :

$${}_{m-1}A \ {}_{m-1}A^t \ a_{m-1} = \left| \begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1m-1} & 0 \\ \vdots & & & \vdots & 0 \\ \vdots & & & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm-1} & \lambda^{m-1} \end{array} \right| = \lambda^{m-1} \ a_{m-1}$$

Então para  $p = 1, 2, \dots, m-1$ :

$$R_m a_p = R_1 a_p - {}_{m-1}A {}_{m-1}A^t a_p$$

$$R_m a_p = \lambda_p^1 a_p - \lambda^p a_p$$

Mas como  $\lambda^2 = \lambda_2^1$  (desenvolvimento do problema 2)

$\lambda^3 = \lambda_2^1$  (desenvolvimento do problema 3)

$\vdots$       $\vdots$

$\lambda^{m-1} = \lambda_{m-1}^1$  (desenvolvimento do problema (m-1))

Então:

$$R_m a_p = \lambda_p^1 a_p - \lambda_p^1 a_p = 0$$

Ou seja, para  $p = 1, 2, \dots, m-1$ ,  $a_p$  é auto-valor da matriz  $R_m$  com auto-valores

$$\lambda_1^m = \lambda_2^m = \dots = \lambda_{m-1}^m = 0$$

Para  $p = m$ :

$$R_m \alpha_p = (R_1 - {}_{m-1}A {}_{m-1}A^t) \alpha_p = R_1 \alpha_p - {}_{m-1}A {}_{m-1}A^t \alpha_p$$

$${}_{m-1}A^t \alpha_m = \begin{vmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m-1} & a_{2m-1} & \cdots & a_{nm-1} \end{vmatrix} \begin{vmatrix} \alpha_{1m} \\ \alpha_{2m} \\ \vdots \\ \alpha_{nm} \end{vmatrix} = \begin{vmatrix} a_{1m-1}^t \alpha_m \\ a_{2m-1}^t \alpha_m \\ \vdots \\ a_{nm-1}^t \alpha_m \end{vmatrix} = 0$$

porque todos os auto-vetores de  $R_1$  são ortogonais.

Então para  $p = m$ , temos:

$$R_m \alpha_m = R_1 \alpha_m - {}_{m-1}A {}_{m-1}A^t \alpha_m$$

Como:

$${}_{m-1}A^t \alpha_m = 0$$

$$R_m \alpha_m = R_1 \alpha_m = \lambda_m^1 \alpha_m$$

Então:

$$R_m \alpha_m = \lambda_m^1 \alpha_m$$

Isto é, o auto-vetor  $\alpha_m$  da matriz  $R_1$  é auto-vetor da matriz  $R_m$  com auto-valor  $\lambda_m^m = \lambda_m^1$ , que será o único auto-valor da matriz  $R_m$  diferente de zero. Logo calcular o maior auto-valor de  $R_m$  é calcular o auto-valor  $\lambda_m^1$  da matriz  $R_1$  e seu auto-vetor as

sociado será  $a_m$ .

- Resumo da resolução do modelo de Fatores Principais

A resolução é obtida através da solução de  $m$  problemas análogos. Em cada um deles estamos interessados no maior auto-valor de uma matriz e seu respectivo auto-vetor.

Problema 1: Matriz  $R_1$

Auto-valores  $\lambda_1^1 \geq \lambda_2^1 \geq \lambda_3^1 \geq \dots \geq \lambda_n^1$

Auto-vetores  $a_1, a_2, a_3, \dots, a_n$

Solução:  $V_1 = \lambda_1^1$  valor da função objetivo

$a_1 \rightarrow 1^a$  coluna da matriz  $A$ .

Problema 2: Matriz  $R_2$

Auto-valores  $\lambda_1^2, \lambda_2^2, \lambda_3^2, \dots, \lambda_n^2$

Auto-vetores  $a_1, a_2, \dots, a_n$  (iguais aos da matriz  $R_1$ )

Nesse ponto provamos que os auto-vetores de  $R_2$  são os mesmos que os de  $R_1$ , e que os auto-valores são iguais a menos de  $\lambda_1^2 = 0$ . Então achar o maior auto-valor de  $R_2$  é achar o

2º maior auto-valor da matriz  $R_1$ .

Solução:  $V_2 = \lambda_2^1 \rightarrow$  valor da função objetivo.

$a_2 \rightarrow$  2º vetor coluna de A.

Problema m: Matriz  $R_m$

Auto-valores  $\lambda_1^m, \lambda_2^m, \dots, \lambda_m^m$

Auto-vetores  $a_1, a_2, \dots, a_m$  (iguais aos da matriz  $R_1$ )

Nesse ponto provamos que:

$$\lambda_1^m = \lambda_2^m = \dots = \lambda_{m-1}^m = 0 \text{ e que}$$

$$\lambda_m^m = \lambda_m^1$$

Solução:  $V_m = \lambda_m^1 \rightarrow$  valor da função objetivo

$a_m =$  m-ésima coluna da matriz A.

Podemos concluir portanto que para resolver os  $m$  problemas propostos, basta calcular os  $m$  auto-valores e auto-vetores associados a matriz  $R_1$ . O número  $m$  fica determinado porque para cada fator extraído calculamos as correlações reproduzidas e comparamos com as observadas. Teremos extraído todos os fatores comuns, quando as correlações residuais for-

rem nulas. Nesse momento teremos:

$$R_1 = \underset{m}{A} \underset{m}{A}^t = AA^t$$

e teremos calculado todos os  $\underline{m}$  auto-valores diferentes de zero da matriz de correlação reduzida  $R_1$ . O que equivale a dizer que o rank de  $R_1$  é  $\underline{m}$ . Essa relação entre o número de fatores comuns e o rank da matriz de correlação reduzida será mostrada formalmente na seção 3.3. Então, ao contrário do método do Resíduo mínimo onde se estimava o número de fatores, mas as comunalidades eram fornecidas como resultado do método, aqui é exatamente o oposto: estima-se as comunalidades mas obtém-se através do método o número de fatores comuns.

Comparando com o método de Análise das Componentes Principais, a diferença do ponto de vista operacional, é que nesse método, fatoramos a matriz  $R$  (matriz das correlações dos dados), enquanto que no método de Fatores Principais fatoramos  $R_1$  (matriz das correlações reduzida, com estimativa das comunalidades na diagonal principal). A discussão dessa estimativa, bem como da estimativa do número de fatores será feita a seguir.

### 3.3 - ESTIMATIVA DAS COMUNALIDADES E DO NÚMERO DE FATORES COMUNS

Como vimos nas seções 3.1 e 3.2, os métodos de A.F. apresentados, supõe estimativas a priori, ou das comunalidades ou do número de fatores comuns. Na verdade, para qualquer método de A.F. usado, pelo menos uma dessas duas estimativas é necessária. Faremos nessa seção, uma análise desses dois problemas e apresentaremos, para ambos os casos, regras práticas usadas pelos estatísticos quando utilizam os métodos de A.F.

Existe uma ligação entre esses dois problemas (comunalidades e número de fatores) já que o número de fatores  $m$  é igual ao rank da matriz das correlações reduzida  $R^*$  (isto é, com as comunalidades na diagonal principal). Ou seja, ao se fazer uma A.F., procura-se valores  $h_i^2$  ( $i = 1, 2, \dots, n$ ) para substituir a diagonal unitária da matriz das correlações  $R$ , tal que a nova matriz assim obtida  $R^*$  tenha rank igual a  $m < n$  determinando a dimensão do espaço comum do problema isto é, o número de fatores comuns necessários para descrever a parte comum das variáveis. Será mostrado a seguir esse resultado.

Teorema: Se o rank de  $R^*(n \times n) = m$ , então são necessários no mínimo  $m$  fatores comuns para a descrição das variáveis.

Prova: Seja rank  $R^*(n \times n) = m$ .

Seja  $A(n \times r)$  matriz padrão tal que

$$R^* = AA^t.$$

A partir do resultado abaixo, demonstrado em [9]

$$\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$$

no nosso caso podemos escrever:

$$m = \text{rank}(AA^t) \leq \min\{\text{rank}(A), \text{rank}(A^t)\}$$

Ou seja:

$$\text{rank}(A) \geq m$$

Como  $A(n \times r)$ , onde  $r$  é o número de fatores comuns e pela hipótese do modelo de Análise Fatorial  $r < n$ ,  $r$  é a menor dimensão da matriz  $A$ . Então:

$$\text{rank}(A) \leq r$$

Mas como

$$\text{rank}(A) \geq m$$

Temos:

$$m \leq r$$

Ou seja o número mínimo de fatores comuns necessários para descrever as variáveis segundo o modelo de Análise Fatorial é  $m$ .



Através desse resultado concluímos que ao substituir as variâncias das variáveis pelas suas comunalidades na matriz das correlações, isto é, levando em conta apenas a parte comum a todas as variáveis, diminuimos a dimensão do espaço vetorial que descreve as correlações entre elas. Podemos entender isso como se fosse necessária uma dimensão maior ( $n$ ) para descrever as correlações das variáveis completas (parte comum + parte específica). Entendemos que uma base para esse espaço teria  $n$  eixos de referência, e não seria possível descrever todas as correlações com um número menor de eixos. Jogando fora a parte específica da variância das variáveis, um certo número de eixos ( $n - m$ ) torna-se desnecessário. Esses eixos serviam para descrever a parte específica das variáveis. Ou seja reduzir a variância ao valor da variância comum, corresponde a jogar fora a variância específica, e junto com ela, os eixos de referência que são eram necessários para descrever essa parte específica. Os eixos restantes formam a base do espaço vetorial que descreve a parte comum das variáveis. É o espaço comum das variáveis descritas pelo modelo de Análise Fatorial, e obter os fatores comuns é obter uma base para esse espaço.

Outra maneira de ver a relação entre o valor das comunalidades e o número de fatores comuns, é através da contribuição de cada fator para a variância total. Quando fazemos uma A.C.P. de um conjunto de  $n$  variáveis, extraímos  $n$  componentes principais para reproduzir a variância total dos dados. Se ao invés da variância total, procuramos reproduzir apenas a parte comum da variância, que corresponde a uma certa percentagem da variância total, será suficiente um número menor de fa-

tores. A Figura 3.1 mostra a variância acumulada para os  $n$  fatores. Vemos que chamando  $H^2$  a percentagem da variância total correspondente a sua parte comum, precisamos de apenas  $m < n$  fatores para atingi-la.

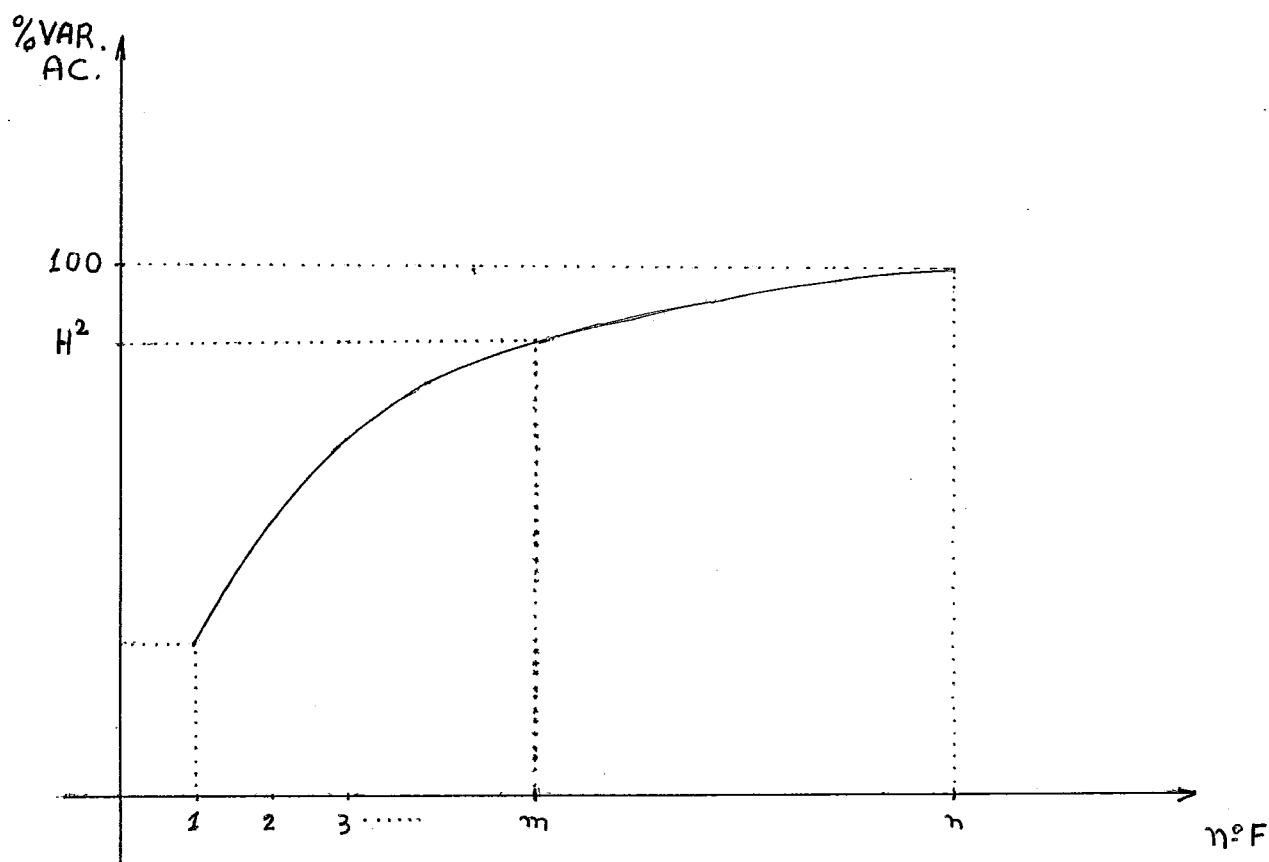


Figura 3.1 - Variância acumulada X número fatores

#### - Estimativa das Comunalidades

São basicamente três os processos de se estimar as comunalidades de um conjunto de  $n$  variáveis.

- 1.- Solução analítica utilizando condições necessárias para se obter um rank m.
- 2 - Através de aproximações empíricas.
- 3 - Através de um processo iterativo.

### 1 - Solução Analítica

Supondo sabido o rank da matriz de correlação reduzida, rank  $R^*(n \times n) = m$ , temos o seguinte problema a ser resolvido:

Determinar  $h_1^2, h_2^2, \dots, h_n^2$

onde: 1)  $R^* = \begin{vmatrix} h_1^2 & r_{12} & \dots & r_{1n} \\ r_{21} & h_2^2 & \dots & r_{2n} \\ \vdots & \vdots & & \vdots \\ r_{n1} & r_{n2} & \dots & h_n^2 \end{vmatrix}$

2) rank  $R^* = m$

3)  $r_{ik} =$  correlação observada  $\forall i, k, i \neq k$ .

4)  $0 \leq h_i^2 \leq 1$

A condição rank  $R^* = m$ , impõe que todos os determinantes menores da matriz  $R^*$ , de ordem  $> m$  serão nulos. Na solução desses determinantes as comunalidades aparecerão como in

cōgnitas a serem determinadas por:

$$\det|\nabla \text{ sub-matriz quadrada de } R^* \text{ de ordem } < m| = 0$$

$$\text{e } 0 \leq h_i^2 \leq 1 \quad (1)$$

A resolução de cada um desses determinantes de ordem igual a  $m + 1, m + 2, \dots, n$  corresponderá a um polinômio de ordem  $m + 1, m + 2, \dots, n$ , respectivamente. O sistema formado por esses polinômios deve ser resolvido levando-se ainda em conta a restrição para o valor das comunalidades entre zero e um. A medida que crescem o número de variáveis  $n$  e o número de fatores  $m$ , a solução analítica desse problema torna-se tão trabalhosa que é abandonada em favor de outros métodos de aproximação que veremos a seguir. Um exemplo para 4 variáveis e  $\text{rank} = 2$  está resolvido em [10].

## 2.- Aproximações Empíricas

Diante da dificuldade de se obter uma solução analítica para o problema geral das comunalidades, surgiram aproximações que foram testadas na prática e algumas delas mostraram-se eficientes. Por exemplo:

### a) S. M. C. (Square Multiple Correlation)

O conceito de correlação múltipla vem da análise de regressão múltipla onde, em um conjunto de  $n$  variáveis desejamos escrever uma (por exemplo  $x_n$ ) em função das  $n-1$  variáveis

veis restantes ( $x_1, x_2, \dots, x_{n-1}$ ). O problema a ser resolvido na análise de regressão múltipla é determinar  $\beta_1, \beta_2, \dots, \beta_{n-1}$  tal que:

$$y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{n-1} x_{n-1}$$

e  $y$  tenha máxima correlação com  $x_n$ . Essa correlação máxima é chamada de coeficiente de correlação múltipla de  $x_n$  com todas as  $(n-1)$  variáveis restantes. O quadrado da correlação múltipla é igual a porção da variância de  $x_n$  que pode ser atribuída ao conjunto  $x_1, x_2, \dots, x_{n-1}$ , segundo a regressão feita. Esse valor é dado por:

$$r_i^2 = 1 - \frac{1}{S_i^2 r^{ii}}$$

onde:

- $r_i^2$  é o quadrado da correlação múltipla da  $i$ -ésima variável com todas as demais (S.M.C.).
- $r^{ii}$  é o  $i$ -ésimo elemento da diagonal da matriz  $R^{-1}$ .
- $S_i^2$  é a variância da  $i$ -ésima variável.

Como estamos trabalhando com variáveis padronizadas, podemos escrever:

$$r_i^2 = 1 - \frac{1}{r^{ii}}$$

Overall and Klett em [25] afirmam que o S.M.C. usado como estimativa das comunalidades tem uma forte justificativa do ponto de vista estatístico, porque pode ser interpretado como a porção da variância de cada variável que é dividida com as demais. Além disso, em [4] foi mostrado que o S.M.C. é um limite inferior para o valor das comunalidades, isto é:

$$r_i^2 \leq h_i^2$$

Dessa maneira, ao fazermos

$$r_i^2 = h_i^2$$

estamos subestimando o valor das comunalidades e portanto a comunalidade total será também subestimada. Consequentemente podemos tomar um número insuficiente de fatores para descrever a parte comum das variáveis (ver Figura 3.1). Temos então o controle do possível erro que cometemos na análise dos dados, o que não ocorre para outras estimativas. Mas apesar das vantagens de se usar o S.M.C. como aproximação para as comunalidades, outras estimativas foram criadas sem maiores justificativas teóricas ou estatísticas, apenas com a vantagem de serem menos trabalhosas, não requerendo o cálculo da inversa da matriz de correlação  $R^{-1}$ . Exemplos:

$$b) h_i^2 = \max_j r_{ij}$$

$$c) h_i^2 = \max_{j, k} \left( \frac{r_{ij} r_{ik}}{r_{jk}} \right) \text{ para } j \neq i, j \neq k.$$

### 3 - Processo Iterativo

Supõe-se um número  $m$  de fatores comuns e usa-se uma estimativa qualquer para as comunalidades, construindo-se a matriz de correlação reduzida inicial  $R^0$ . Faz-se uma Análise das Componentes Principais de  $R^0$  ( $n$  componentes) e obtem-se o valor das novas comunalidades, somando-se o quadrado dos coeficientes das  $m$  primeiras componentes para cada variável ( $\sum_{p=1}^m a_{ip}^2$ ). Substitui-se essas novas comunalidades na diagonal principal de  $R^0$  obtendo-se  $R^1$ . Fatora-se  $R^1$  e calcula-se outro conjunto de comunalidades, e assim sucessivamente até que as comunalidades calculadas numa dada iteração sejam iguais (a menos de uma tolerância) às obtidas na iteração anterior. Apresentamos a seguir esse processo sob a forma de algoritmo:

#### - Algoritmo para Estimativa das Comunalidades:

Dados:  $R$ ,  $m$ ,  $h_i^0$  ( $i = 1, 2, \dots, n$ ),  $tol$

$k = 0$

Passo 1: Construa  $R^k$  substituindo  $h_i^k$  na diagonal principal de  $R$ .

Passo 2: Faça uma A.C.P. de  $R^k$  e calcule as novas comunalidades  $h_i^{k+1} = \sum_{p=1}^m a_{ip}^2$

Se  $|h_i^k - h_i^{k+1}| \leq tol \forall i$ , pare

Senão, faça  $k = k + 1$  e vá para 1.

Esse método iterativo é o que fornece as melhores estimativas para as comunalidades mas só pode ser empregado com auxílio de computador. Na verdade embora no início do desenvolvimento da Análise Fatorial isso fosse uma questão que preocupava, atualmente não se pensa em fazer uma Análise Fatorial sem o auxílio de computadores. Além disso existem "pacotes" com programas que resolvem o problema de A.F. por diversos métodos diferentes, e geralmente quando são requeridas as estimativas das comunalidades, o S.M.C. é usado como valor inicial. Caso se queira, pode-se a partir daí, utilizar-se o processo iterativo. Portanto concluímos que o S.M.C. é a melhor das estimativas empíricas e que caso queiramos com um esforço computacional maior, obter resultados mais precisos, usamos o processo iterativo como um refinamento da solução.

#### - Estimativa do Número de Fatores Comuns

Até aqui, tratando do problema das comunalidades supúnhamos conhecido o número de fatores ou seja o rank da matriz de correlação reduzida. Passaremos agora a tratar do problema de estimar o número de fatores necessários para a descrição das variáveis segundo o modelo de A.F. Os métodos para estimar o número de fatores são de dois tipos:

- 1 - Inferência estatística
- 2 - Regras práticas.



## 1 - Inferência Estatística

Testes estatísticos foram desenvolvidos para estimar o número de fatores, partindo da hipótese da multinormalidade das variáveis [23]. Destaca-se entre muitos o trabalho de Rippe que propõe um teste de significância dos fatores que em relação a outros trabalhos do mesmo tipo, tem a vantagem de independer do método utilizado para solução do modelo. Ver [27].

## 2 - Regras Práticas

Não tendo sido resolvido analiticamente de um modo geral o problema do número de fatores, surgiram regras práticas que não são provadas matematicamente, mas são muito utilizadas por serem facilmente observáveis a partir de uma Análise das Componentes Principais.

### a) Teste da variação dos auto valores

Esse teste foi apresentado em 1966 por Cattell [3], e resulta da observação prática do nível da variância devido a cada fator, dada pelo auto valor correspondente. Ao fazermos uma A.C.P., extraímos  $n$  fatores, dos quais  $m$  são fatores comuns, cada um correspondendo a um auto valor que fornece a sua contribuição na variância total. Os  $(n - m)$  fatores restantes, podem ser considerados como medidas de erro aleatório e por isso, a contribuição desses fatores na variância total será aproximadamente a mesma. Então os auto valores correspon-

dentes serão também aproximadamente iguais. A Figura 3.2 mostra essa característica através do gráfico % variância total X ordem do fator

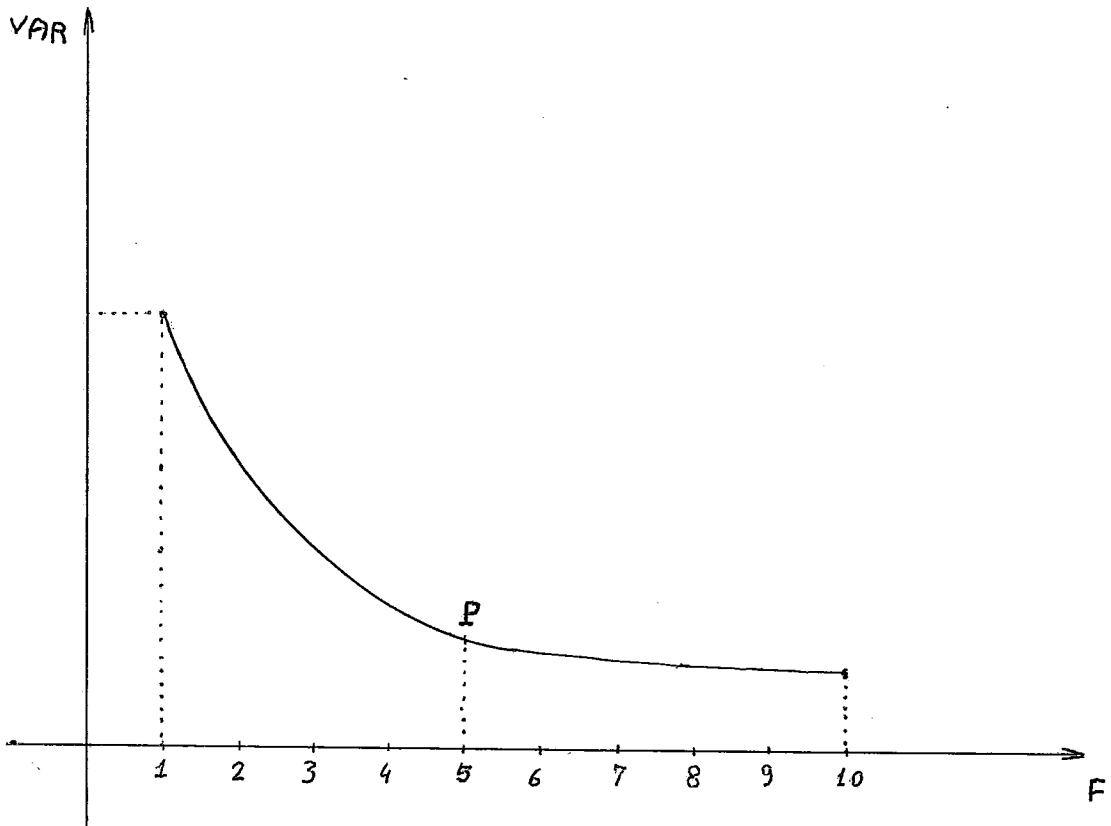


Figura 3.2 - Teste da variação dos auto valores.

O ponto P no gráfico indica que o último fator considerado é o quinto. Os demais, segundo esse teste, medem apenas erro aleatório.

#### b) Correlação entre fatores e variáveis

Esse critério depende da grandeza do coeficiente da variável para cada fator ( $a_{ip}$ ), extraídos na A.C.P. Como

$$S_i^2 = \sum_{p=1}^n a_{ip}^2$$

é a variância da  $i$ -ésima variável,  $a_{ip}^2$  é a contribuição para essa variância fornecida pelo  $p$ -ésimo fator. Segundo esse critério, serão excluídos os fatores que não contribuam para no mínimo 10% da variância de ao menos uma variável. Além disso, como estamos interessados em fatores comuns, devem ser excluídos aqueles que tenham apenas um coeficiente alto, pois esses fatores podem ser vistos mais como específicos para uma variável que propriamente fatores comuns.

### c) Auto valor um

Critério baseado na Análise de Componentes Principais que fornece  $n$  fatores com variância igual ao auto valor correspondente. Como a variância de cada variável padronizada é igual a um, cortar os fatores com auto valores menores que um significa excluir aqueles que contribuem na variância total com uma parcela menor que a variância de uma variável. Kaiser em 1960 [18] afirmou que essa era a melhor maneira de se encontrar o número de fatores necessários para a A.F. e esse é na verdade o método mais usado na prática.

Após a apresentação dessas regras, alertamos para o fato de que seu uso deve ser feito com certo cuidado, seguindo sempre de uma interpretação dos resultados. De nada adianta seguir uma dessas regras mecanicamente e obter uma solução que não seja interpretável.

### 3.4 - APLICAÇÕES DOS MÉTODOS DE FATORES PRINCIPAIS E RESÍDUO MÍNIMO

#### - Exemplo 3.1

Trata-se de um exemplo clássico de Análise Fatorial que foi primeiramente apresentado em 1937 por Holzinger e Swineford [15]. Esse exemplo foi escolhido por ter sido utilizado na aplicação dos métodos apresentados aqui, de tal forma que facilita a comparação entre eles.

O estudo em questão, consta de uma coleção de 24 testes psicológicos aplicados em 145 alunos das sétima e oitava série. Cada um desses testes mede diferentes capacidades do indivíduo. Apresentamos a seguir as principais características de cada um deles.

- 1) Percepção visual - capacidade de concentração
- 2) Cubos - agilidade e coordenação motora
- 3) Quadros de papel - agilidade mental
- 4) Bandeiras - capacidade dedutiva
- 5) Informações gerais - grau de relação com o mundo exterior
- 6) Compreensão de parágrafos - disponibilidade, ansiedade, agilidade mental
- 7) Completar sentença - expressão verbal, deteta patologias
- 8) Classificação de palavras - capacidade verbal
- 9) Significado de palavras - agilidade mental e nível de educação

- 10) Adição - capacidade l $\bar{o}$ gica simb $\bar{o}$ lica
- 11) C $\bar{o}$ digo - capacidade l $\bar{o}$ gica simb $\bar{o}$ lica e ansiedade
- 12) Contagem de pontos - ansiedade, obsess $\bar{a}$ o, defensas neur $\bar{o}$ ticas
- 13) Mai $\bar{u}$ sculas retas e curvas - agilidade mental
- 14) Reconhecimento de palavras - percep $\bar{c}$ o $\bar{a}$ o visual
- 15) Reconhecimento de n $\bar{u}$ meros - percep $\bar{c}$ o $\bar{a}$ o visual e mem $\bar{o}$ ria
- 16) Reconhecimento de figuras - percep $\bar{c}$ o $\bar{a}$ o visual, capacidade de abstra $\bar{c}$ o $\bar{a}$ o
- 17) N $\bar{u}$ mero-objeto - percep $\bar{c}$ o $\bar{a}$ o visual
- 18) Figura-n $\bar{u}$ mero - percep $\bar{c}$ o $\bar{a}$ o visual
- 19) Figura-palavra - percep $\bar{c}$ o $\bar{a}$ o visual
- 20) Dedu $\bar{c}$ o $\bar{a}$ o - capacidade dedutiva
- 21) Quebra cabe $\bar{c}$ a num $\bar{e}$ rico - capacidade dedutiva
- 22) Racioc $\bar{i}$ nios em problemas - informa $\bar{c}$ o $\bar{a}$ o geral, dedu $\bar{c}$ o $\bar{a}$ o
- 23) Completar s $\bar{e}$ rie - mostra estrutura do inconsciente
- 24) Problemas aritm $\bar{e}$ ticos - problemas ou defici $\bar{e}$ ncias educacio $\bar{n}$ ais.

A tabela 3.1 apresenta uma an $\bar{a}$ lise estat $\bar{i}$ stica b $\bar{a}$ sica (m $\bar{e}$ dia e desvio padr $\bar{a}$ o) para cada vari $\bar{a}$ vel. As tabelas 3.2 e 3.3 apresentam respectivamente a matriz das correla $\bar{c}$ o $\bar{e}$ s dos dados, e a solu $\bar{c}$ o $\bar{a}$ o para o problema segundo o m $\bar{e}$ todo de An $\bar{a}$ lise dos Fatores Principais.

Tabela 3.1

TESTE $x_i$	MÉDIA ( $\bar{x}_i$ )	DESVIO PADRÃO $S_i$
1- Percepção visual	29.6	6.9
2- Cubos	24.8	4.5
3- Quadros de papel	15.6	3.1
4- Bandeiras	36.3	8.4
5- Informações gerais	44.9	11.7
6- Compressão de parágrafo	9.9	3.4
7- Completar sentença	18.8	4.6
8- Classificação de palavras	28.2	5.3
9- Significado de palavras	17.2	7.9
10- Adição	90.2	23.6
11- Código	68.4	16.8
12- Contagem de pontos	109.8	21.0
13- Maiúsculas retas e curvas	191.8	37.0
14- Reconhecimento de palavras	176.1	10.7
15- Reconhecimento de números	89.4	7.6
16- Reconhecimento de figuras	103.4	6.7
17- Número-objeto	7.1	4.6
18- Figura-número	9.4	4.5
19- Figura-palavra	15.2	3.6
20- Dedução	30.4	19.8
21- Quebra cabeça numérico	14.5	4.8
22- Raciocínios em problemas	27.7	9.8
23- Completar série	18.8	9.3
24- Problemas aritméticos	25.8	4.7

TABELA 3.2 - Matriz das Correlações para as 24 Variáveis

$x_j$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
1																									
2	.3																								
3	.4	.3																							
4	.5	.2	.3																						
5	.3	.3	.2	.3																					
6	.3	.2	.3	.3	.6																				
7	.3	.2	.2	.3	.7	.7																			
8	.3	.2	.4	.4	.6	.5	.6																		
9	.3	.2	.2	.3	.7	.7	.7	.5																	
10	.1	.1	.0	.1	.3	.2	.2	.3	.1																
11	.3	.1	.1	.1	.3	.4	.2	.3	.3	.5															
12	.3	.1	.1	.2	.2	.1	.2	.3	.1	.6	.4														
13	.5	.2	.3	.3	.3	.3	.3	.3	.3	.5	.5	.5													
14	.1	.1	.2	.1	.3	.3	.2	.3	.3	.2	.3	.1	.2												
15	.2	.1	.6	.1	.2	.3	.2	.2	.2	.2	.2	.2	.1	.4											

continua....

TABELA 3.2 - Continuação

$x_j$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
16	.4	.3	.3	.3	.2	.3	.2	.3	.2	.1	.3	.1	.3	.4	.3									
17	.2	.0	.2	.2	.2	.3	.2	.3	.3	.3	.4	.3	.2	.3	.3	.3								
18	.4	.3	.2	.3	.3	.2	.2	.2	.2	.3	.3	.3	.3	.2	.3	.3	.4							
19	.3	.1	.3	.1	.2	.3	.2	.3	.3	.2	.3	.1	.3	.2	.2	.2	.2	.3	.4					
20	.4	.3	.3	.3	.4	.4	.5	.4	.4	.2	.2	.2	.2	.3	.3	.4	.3	.3	.2					
21	.4	.3	.2	.3	.3	.3	.3	.4	.3	.4	.4	.4	.4	.2	.2	.3	.2	.4	.3	.4				
22	.4	.2	.2	.4	.4	.4	.4	.4	.5	.2	.3	.2	.3	.2	.2	.3	.3	.3	.3	.3	.5	.4		
23	.5	.3	.4	.3	.4	.4	.4	.5	.5	.3	.3	.3	.4	.2	.3	.4	.3	.3	.3	.3	.5	.5	.5	
24	.3	.2	.2	.2	.4	.4	.4	.4	.4	.5	.4	.4	.4	.3	.1	.3	.3	.4	.4	.4	.4	.4	.4	.4



Tabela 3.3 - Solução do método de Fatores Principais (estimativa das comunalidades: S.M.C.)

TESTE $x_i$	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$	COMUNALIDADE ESTIMADA
1	.595	-.039	.369	-.184	-.073	.511
2	.374	.026	.270	-.147	.121	.300
3	.433	.115	.396	-.112	-.276	.440
4	.501	.108	.290	-.178	.044	.409
5	.701	.312	-.273	-.050	.003	.673
6	.683	.404	-.213	.067	-.102	.677
7	.676	.412	-.284	-.082	-.046	.684
8	.680	.206	-.088	-.115	-.118	.564
9	.690	.446	-.212	.076	.036	.713
10	.450	-.469	-.446	-.128	.105	.579
11	.589	-.372	-.198	.076	-.185	.541
12	.449	-.491	-.154	-.263	.044	.537
13	.590	-.268	.018	-.300	-.255	.539
14	.435	-.063	-.012	.418	-.057	.358
15	.390	-.102	.055	.362	.101	.293
16	.512	-.098	.325	.259	.006	.429
17	.471	-.212	-.036	.388	-.087	.412
18	.521	-.331	.118	.145	.028	.443
19	.450	-.115	.110	.167	-.178	.367
20	.623	.135	.142	.049	.252	.464
21	.596	-.220	.076	-.140	.204	.473
22	.600	.103	.138	.053	.142	.449
23	.685	.063	.160	-.096	.154	.561
24	.635	-.169	-.192	-.009	.079	.527

continua...

Tabela 3.3 - Continuação...

TESTE $x_i$	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$	COMUNALIDADE ESTIMADA
$V_p$	7.665	1.672	1.208	.920	.447	Total:11.943
100 $V_p/11.943$	64.2	14.0	10.1	7.7	3.7	100

As características gerais de uma solução do método de Fatores Principais são observadas na tabela 3.3. Primeiro, as contribuições dos fatores para a comunalidade total das variáveis decresce a cada novo fator calculado. Com isso aparece imediatamente uma regra para o erro cometido ao se parar de fatorar. Se o fator extraído tem, por exemplo, uma contribuição de 5% na variância total, sabemos que o próximo e todos os subsequentes terão contribuições menores. Nesse exemplo, os cinco primeiros fatores contribuem praticamente com toda a comunalidade estimada a priori (a estimativa foi feita pelo cálculo do quadrado da correlação múltipla (S.M.C.) que como vimos na seção 3.3, é o método mais utilizado) por isso foram mantidos esses fatores como significativos.

- Interpretação da solução do método de Fatores Principais

$F_1$  → fator geral que caracteriza a solução de F.P. - difícil interpretação, pois está altamente correlacionada com quase todas as variáveis.

$F_2$  → fator bipolar - caso seja interpretável representará duas grandezas "opostas"

coef. correlação  $> 0.4$  → variáveis 6, 7, 9 - corresponde a testes relacionados com "expressão verbal".

coef. correlação  $< -0,4$  → variável 10, 12.

Os fatores  $F_3$ ,  $F_4$ ,  $F_5$ , são fatores bi-polares que não apresentam coeficientes altos, ou seja, quase todos os coeficientes são menores (em módulo) que 0,4. Portanto, não podemos associá-los a grupos de variáveis, dificultando assim a

interpretação dos mesmos. Para se obter uma interpretação mais precisa da solução, normalmente se faz uma mudança no referencial das variáveis (ver sec. 3.5), identificando cada fator com um grupo de variáveis.

Apresentamos a seguir, a tabela 3.4 com a solução do método de Resíduo Mínimo.

Tabela 3.4 - Solução do método do Resíduo Mínimo (cinco fatores comuns)

TESTE $x_i$	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$	$h_i^2$
1	.598	.024	.381	-.210	-.086	.555
2	.373	-.035	.261	-.137	.072	.232
3	.421	-.125	.367	-.128	-.119	.358
4	.484	-.112	.258	-.182	.060	.350
5	.686	-.293	-.279	-.044	-.027	.637
6	.688	-.402	-.212	.063	-.140	.703
7	.676	-.407	-.310	-.088	-.029	.728
8	.673	-.194	-.094	-.109	-.038	.513
9	.696	-.445	-.235	.066	.000	.741
10	.473	.541	-.476	-.085	.170	.779
11	.560	.376	-.168	-.092	-.271	.565
12	.465	.494	-.125	-.219	.065	.528
13	.614	.306	.028	-.371	-.378	.751
14	.426	.051	.003	.422	-.145	.384
15	.392	.084	.092	.369	.003	.305
16	.512	.075	.346	.251	-.073	.456
17	.465	.195	-.004	.394	-.027	.410
18	.518	.307	.165	.160	.080	.421
19	.443	.093	.102	.137	-.031	.235
20	.621	-.147	.141	.055	.226	.481
21	.596	.210	.080	-.199	.180	.452
22	.613	-.113	.120	.042	.140	.424
23	.692	-.071	.152	-.092	.182	.550
24	.651	.168	-.187	.017	.158	.512
Variância	7.637	1.717	1.233	.948	.498	12.07

Foram escolhidos cinco fatores comuns, porque esse número foi considerado suficiente na solução do método de Fatores Principais. As comunalidades calculadas pelo método do Resíduo Mínimo a partir de um pré-determinado número de fatores são as melhores aproximações do conceito de comunalidade. Como podemos observar, as soluções de F.P. e R.M. são muito parecidas, e isso sempre ocorrerá desde que sejam feitas boas estimativas para comunalidades na solução F.P. Foi feito, para solução do Resíduo Mínimo o teste de hipótese de Rippe da significância do número de fatores, e a solução com cinco fatores é, segundo o teste, satisfatória.

- Interpretação da solução - como a solução é muito parecida com a obtida pelo método de F.P., a interpretação é a mesma.

Examinando essas duas soluções para o problema de A.F., vemos que apresentam uma dificuldade muito grande de interpretação. Na verdade muito pouco ficamos sabendo a respeito dos dados após essa análise. Tudo que conseguimos saber é que existem cinco fatores que contribuem significativamente para a descrição dos dados, mas não conseguimos através da interpretação das soluções, descobrir o significado de cada um desses fatores. Esse problema é muito comum na A.F., e procurando facilitar a interpretação da solução foram desenvolvidos métodos de rotação, com o objetivo principal de identificar cada fator com um grupo de variáveis. Esses métodos são aplicados após a determinação de uma solução inicial, na qual não foi possível interpretar o significado de cada fator. A seção 3.5 trata exatamente desse assunto.

### 3.5 - ROTAÇÃO ORTOGONAL DA SOLUÇÃO INICIAL

Como já foi dito no início desse trabalho um dos objetivos da Análise Fatorial, é obter  $m < n$  fatores comuns relativos aos diversos grupos de variáveis correlacionadas. Ou seja, cada um dos  $m$  fatores deve corresponder a um grupo de variáveis correlacionadas. Mas as soluções propostas na seção 3.2 (Método dos Fatores Principais e Método do Resíduo Mínimo) não fornecem fatores comuns com tais características. Na verdade essas soluções iniciais, geralmente servem apenas para fornecer a nova dimensão do problema, isto é, o número de fatores comuns. Mas na seção 2.5, vimos que o modelo de A.F. é indeterminado, ou seja uma vez obtida uma solução, qualquer transformação ortogonal (rotação) dela, também é solução do problema. Portanto podemos escolher entre essa infinidade de soluções, uma tal que permita a identificação dos grupos de variáveis correlacionadas.

- A idéia da Rotação dos Fatores:

Uma vez obtida uma solução inicial qualquer, podemos escrever cada uma das  $n$  variáveis como combinação linear dos  $m$  fatores. Dessa forma, cada variável  $z_i$  pode ser vista como um ponto no espaço de dimensão  $m$ . Tomemos um exemplo simples de cinco variáveis das quais foram extraídos dois fatores comuns.

- Exemplo 3.2 - A Idéia da Rotação dos Fatores:

Tabela 3.5 - Matriz Padrão para Cinco Variáveis - Solução Inicial

VARIÁVEIS	$F_1$	$F_2$
1	. 6	. 3
2	. 7	. 5
3	. 14	. 4
4	. 5	- . 8
5	. 8	- . 4

A representação geométrica dessa solução será dada por cinco pontos no  $R^2$ :

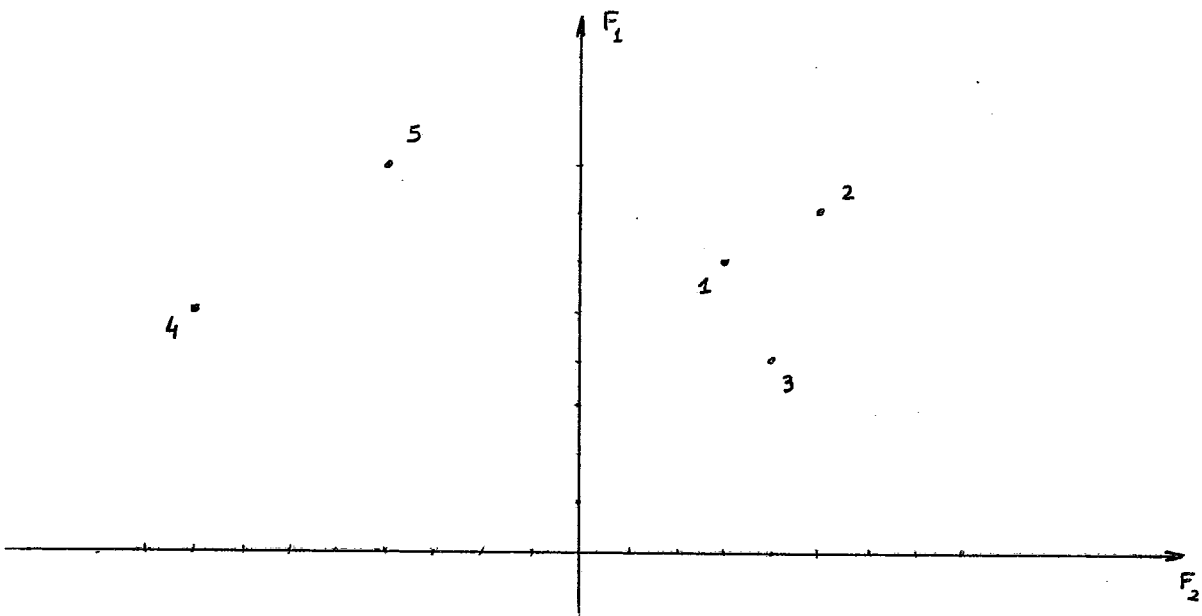


Figura 3.3 - Representação gráfica da solução inicial



Examinando a figura 3.3, percebemos dois grupos de variáveis: variáveis 1, 2, 3 e variáveis 4 e 5, apesar da matriz padrão inicial não fornecer tal idéia. Uma rotação de  $45^\circ$  nos eixos  $F_1$  e  $F_2$  nos daria a seguinte situação:

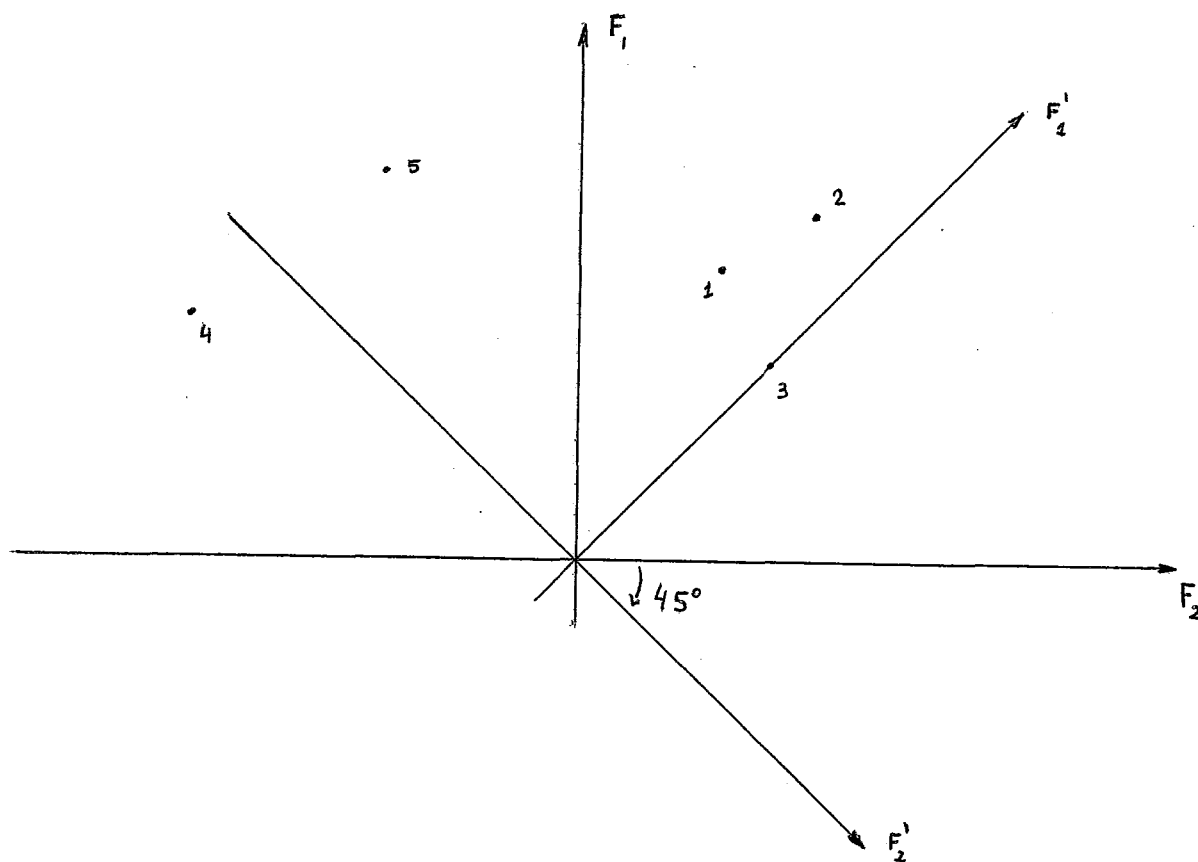


Figura 3.4 - Representação gráfica das soluções inicial e final

Com a seguinte matriz padrão:

Tabela 3.6 - Matriz Padrão para Cinco Variáveis após a Rotação de  $45^{\circ}$  nas Fatores:

VARIÁVEIS	$F_1'$	$F_2'$
1	. 64	- . 21
2	. 85	- . 14
3	. 57	- . 0
4	- . 21	- . 92
5	. 28	- . 85

Observamos que nessa nova matriz padrão os coeficientes das variáveis (coeficientes de correlação entre cada variável e cada fator) nos dão exatamente a idéia dos dois grupos de variáveis. Comparando com a matriz inicial (tabela 3.5), vemos que nessa solução, para cada fator os coeficientes são mais ou menos uniformes (em valor absoluto), ao contrário da solução com rotação na qual podemos separar os coeficientes em "altos" e "baixos". Para o fator  $F_1'$ , as variáveis 1, 2 e 3 tem os maiores coeficientes, isto é, essas variáveis são altamente correlacionadas com o fator  $F_1'$  enquanto que o fator  $F_2'$  tem coeficientes altos para as variáveis 4 e 5 (valor absoluto) fornecendo os dois grupos previstos pela configuração das variáveis.

Esse procedimento gráfico, no entanto, só pode ser seguido para o caso de extrairmos dois fatores comuns, ou no máximo três. Como a A.F. é uma técnica utilizada para diminuir a dimensão do problema, comumente o número de variáveis é

muito grande, e conseqüentemente o número de fatores comuns também é grande. É necessário portanto um critério para rotação que prescindia da visualização do problema. A primeira solução para o problema da rotação dos fatores, foi proposta por Thurstone [33] que utiliza a noção de "estrutura simples" da solução. Ele supõe que uma solução tem estrutura simples, se tem as seguintes características:

- 1 - cada variável deve ter no mínimo um coeficiente baixo (não significativo).
- 2 - Se foram extraídos  $m$  fatores comuns, deve haver no mínimo  $m$  coeficientes não significativos para cada fator.
- 3 - Para cada par de fatores deve haver várias variáveis cujos coeficientes relativos aos dois fatores sejam baixos para um deles e altos para o outro.
- 4 - Para todo par de fatores uma grande parte das variáveis deve ter coeficientes baixos em ambos, quando existem mais de quatro fatores.
- 5 - Para todo par de fatores deve existir apenas um pequeno número de variáveis com coeficientes altos em ambos.

Examinando cada uma delas, as regras de Thurstone podem ser traduzidas dessa forma:

- (1) - minimizar o número de fatores que aparecem na descrição de cada variável.
- (2) - evita o fator geral. Cada fator corresponde a um número não muito grande de variáveis, e em consequên-

cia disso a variância devida ao fator geral da solução inicial é distribuída igualmente entre os fatores rodados.

- (3 e 5) - evita que dois fatores definam o mesmo grupo de variáveis.
- (4) - evita que apenas dois fatores descrevam todas as variáveis, quando são extraídos mais de quatro fatores, isto porque se todas as variáveis fossem descritas por apenas dois fatores, os demais não corresponderiam a nenhum grupo de variáveis e não seriam interpretáveis.

Podemos conferir, na solução apresentada na tabela 3.6, que esta obedece o critério de estrutura simples descrito por Thurstone. Mas esse critério, apesar de ser verificado pela simples observação da matriz padrão rodada, não fornece um método analítico para se chegar a solução.

Em 1954, Ferguson [5] propôs uma solução analítica para o problema, partindo da idéia de que a estrutura simples procurada por Thurstone era baseada no princípio da parcimônia na descrição das variáveis. Do ponto de vista do número de fatores, a idéia de parcimônia fica clara, mas uma vez determinado o espaço de dimensão  $m$ , qual das soluções seria mais parcimoniosa? Obviamente aquela em que cada variável fosse descrita por um número mínimo de fatores, isto é, na descrição da variável  $z_j$  temos:

$$z_i' = a_{i1} f_1 + \dots + a_{im} f_m + d_i u_i$$

A solução será mais parcimoniosa, quando o número de coeficientes altos for pequeno. A esse número, chama-se complexidade da variável. Ou seja, máxima parcimônia corresponde a mínima complexidade das variáveis.

Ferguson procurou então formular matematicamente essa idéia de parcimônia da descrição fornecida pela solução. Partindo de um exemplo simples, um ponto no  $R^2$ , o máximo de parcimônia na descrição desse ponto, ocorre quando um dos eixos passa por ele. Daí, quando o eixo de referência se aproxima desse ponto, o produto das duas coordenadas diminui (já que uma delas se aproxima de zero). Extendendo esse raciocínio, Ferguson sugeriu que uma função da soma dos produtos das coordenadas tomados os eixos dois a dois podia ser usada como medida de parcimônia. Para não levar em conta o sinal das coordenadas, tomou a soma dos quadrados dos produtos destas.

Essa medida, para o caso de  $n$  variáveis e  $m$  fatores ortogonais é:

$$\sum_{i=1}^n \sum_{p=1}^{m-1} \sum_{q=p+1}^m (a_{ip} a_{iq})^2 \quad (3.14)$$

a função que fornece a medida de parcimônia para todas as variáveis  $z_i'$  ( $i = 1, 2, \dots, n$ ), descritas pelo modelo de A.F., segundo Ferguson. A idéia então, é fazer uma rotação ortogonal na solução inicial  $A$ , tal que o novo padrão obtido minimize a função dada em (3.14).

Por outro lado, ao fazermos uma transformação ortogonal numa solução do modelo de A.F. temos:

$$B = AT$$

onde:

$T$  → transformação ortogonal

$A$  → matriz padrão inicial

$B$  → matriz padrão com rotação

E como foi visto anteriormente na seção 2.5:

$BB^t = R^* = AA^t$ , então: os elementos da diagonal de  $R^*$  são:

$$\text{diag } R^* = h_i^2 = \sum_{p=1}^m a_{ip}^2, \text{ ou ainda}$$

$$\text{diag } R^* = \text{diag } BB^t = \sum_{p=1}^m b_{ip}^2$$

Então:

$$h_i^2 = \sum_{p=1}^m a_{ip}^2 = \sum_{p=1}^m b_{ip}^2$$

Assim, podemos concluir que a comunalidade das variáveis não varia quando fazemos uma rotação ortogonal na solução inicial.

Tomando o quadrado das comunalidades:

$$\left( \sum_{p=1}^m b_{ip}^2 \right)^2 = \sum_{p=1}^m b_{ip}^4 + 2 \sum_{p=1}^{m-1} \sum_{q=p+1}^m b_{ip}^2 b_{iq}^2 = \text{cte}$$

Somando para as n variáveis

$$\sum_{i=1}^n \sum_{p=1}^m b_{ip}^4 + 2 \sum_{i=1}^n \sum_{p=1}^{m-1} \sum_{q=p+1}^m b_{ip}^2 b_{iq}^2 = \text{cte} \quad (3.15)$$

Examinando essa expressão vemos que a 2.<sup>a</sup> parcela é o que Ferguson chamou de medida de parcimônia para a descrição de n variáveis. Podemos então proceder de duas maneiras diferentes: a primeira é minimizar diretamente a equação (3.14); a outra é a partir da equação (3.15). Nessa equação, temos a soma de duas parcelas igual a uma constante, então para minimizar a segunda parcela da equação (igual a expressão dada na eq. 3.14), podemos maximizar a primeira.

Ou seja, para obter

$$\min 2 \sum_{i=1}^n \sum_{p=1}^{m-1} \sum_{q=p+1}^m b_{ip}^2 b_{iq}^2 ,$$

basta calcular:

$$\text{Max } Q = \sum_{i=1}^n \sum_{p=1}^m b_{ip}^4 \quad (3.16)$$

Obviamente  $Q$  varia com  $b_{ip}$ , ou seja com a posição dos eixos de referência, já que  $b_{ip}$  é a coordenada da variável  $z'_i$  no eixo de referência  $f_p$ . A idéia então é a seguinte: Maximizando  $Q$ , estamos automaticamente obtendo o novo pa-

drão fatorial  $b_{ip}$  ( $i = 1, 2, \dots, n$ ;  $p = 1, 2, \dots, m$ ) que minimiza a medida de parcimônia das  $n$  variáveis, ou seja um novo padrão que nos forneça uma estrutura simples. O limite teórico acontece quando a complexidade de cada variável é um. Pode-se dizer que foi atingido o máximo grau de estrutura ou organização possível para uma dada configuração.

A seguir veremos o método de rotação Quartimax, cuja idéia é diferente da seguida por Ferguson, mas como veremos a seguir, a função objetivo é a mesma. Trataremos nessa ocasião da solução do problema de maximização correspondente.

#### - O Método Quartimax

Carrol [1], Neuhaus e Wrigley [24] e Saunders [28] desenvolveram o método Quartimax e chegaram a equação (3.16) sem no entanto usar o conceito de medida de parcimônia para descrição das variáveis. O objetivo foi nos três trabalhos, chegar a uma estrutura simples e conseqüentemente diminuir a complexidade de cada variável. O caso ideal ocorre quando cada variável é descrita por apenas um fator, ou seja, todos os coeficientes referentes a  $m-1$  fatores são iguais a zero. Como a variância comum (comunalidade) de cada variável é dada por  $\sum_{p=1}^m a_{ip}^2$ , no caso da variável ser descrita por apenas um fator, sua comunalidade será função apenas do coeficiente desse fator. Então a solução obtida será tão próxima do caso ideal quanto menor for o número de coeficientes altos e maior o número de coeficientes baixos na descrição de uma variável. Para isso, procura-se aumentar os coeficientes altos e dimi-



nuir os coeficientes baixos da matriz padrão inicial. Procedendo dessa forma, estamos incrementando a desigualdade entre os coeficientes dos fatores correspondentes a cada variável  $z_i^1$ . Ou ainda, fazendo com que a variação dos coeficientes dos fatores seja máxima. Como estamos interessados, na grandeza do coeficiente independente do sinal, Neuhaus e Wrigley propuseram que se maximizasse a variação entre os quadrados dos coeficientes dos fatores. Uma forma de medir a variação de uma variável qualquer é calcular a sua variância, então o problema se resume a determinar um novo padrão fatorial  $b_{ip}$  ( $i = 1, 2, \dots, n$ ;  $p = 1, 2, \dots, m$ ) tal que maximize a variância na distribuição dos quadrados dos coeficientes dos fatores. Assim temos, para os  $m$  fatores e  $n$  variáveis, a seguinte função a ser maximizada:

$$S_{b^2}^2 = \frac{1}{mn} \sum_{i=1}^n \sum_{p=1}^m (b_{ip}^2 - \bar{b}^2)^2 \quad (3.17)$$

onde  $b_{ip}$  ( $i = 1, 2, \dots, n$ ;  $p = 1, 2, \dots, m$ ) é o novo padrão fatorial e

$$\bar{b}^2 = \frac{1}{mn} \sum_{i=1}^n \sum_{p=1}^m b_{ip}^2 \quad (3.18)$$

Expandindo a (eq. 3.17) temos:

$$\frac{1}{mn} \sum_{i=1}^n \sum_{p=1}^m (b_{ip}^4 - 2b_{ip}^2 \bar{b}^2 + (\bar{b}^2)^2)$$

Usando o resultado da (eq. 3.18) podemos escrever:

$$\frac{1}{mn} \sum_{i=1}^n \sum_{p=1}^m b_{ip}^4 - \frac{2}{mn} \sum_{i=1}^n \sum_{p=1}^m b_{ip}^2 \frac{1}{mn} \sum_{i=1}^n \sum_{p=1}^m b_{ip}^2 +$$

$$+ \frac{1}{mn} \sum_{i=1}^n \sum_{p=1}^m (\bar{b}^2)^2$$

$$\text{Mas: } \frac{2}{mn} \sum_{i=1}^n \sum_{p=1}^m b_{ip}^2 \frac{1}{mn} \sum_{i=1}^n \sum_{p=1}^m b_{ip}^2 = 2(\bar{b}^2)^2$$

$$\text{e } \frac{1}{mn} \sum_{i=1}^n \sum_{p=1}^m (\bar{b}^2)^2 = \frac{(\bar{b}^2)^2}{mn} \dots nm = (\bar{b}^2)^2$$

Então teremos:

$$\frac{1}{mn} \sum_{i=1}^n \sum_{p=1}^m b_{ip}^4 - 2(\bar{b}^2)^2 + (\bar{b}^2)^2$$

$$M = \frac{1}{mn} \sum_{i=1}^n \sum_{p=1}^m b_{ip}^4 - (\bar{b}^2)^2 \quad (3.19)$$

M definida pela (eq. 3.19),  $\bar{b}$  é a função a ser maximizada. Mas já vimos anteriormente que  $\sum_{p=1}^m b_{ip}^2$  (comunalidade fornecidas pelo padrão fatorial  $b_{ip}$ ) independe da solução desde que a transformação seja ortogonal, portanto:

$(\bar{b}^2)^2 = \left( \frac{1}{mn} \sum_{i=1}^n \sum_{p=1}^m b_{ip}^2 \right)^2$  é constante para qualquer padrão

$b_{ip}$ . Logo:

$$\text{Max } M = \frac{1}{mn} \sum_{i=1}^n \sum_{p=1}^m b_{ip}^4 - (\bar{b}^2)^2$$

equivale a:

$$\text{Max } \frac{1}{mn} \sum_{i=1}^n \sum_{p=1}^m b_{ip}^4 = \text{Max } \frac{Q}{mn}$$

Onde  $Q$  é a função obtida por Ferguson (eq. 3.16). Ou seja maximizar a equação de Ferguson é equivalente à utilização do método do Quartimax para determinar um novo padrão fatorial  $b_{ip}$  ( $i = 1, 2, \dots, n; p = 1, 2, \dots, m$ ).

#### - Solução do Problema

Voltando à idéia da rotação dos fatores, ao procurarmos um novo padrão  $b_{ip}$  ( $i = 1, 2, \dots, n; p = 1, 2, \dots, m$ ) estamos na verdade rodando os eixos  $f_1, f_2, \dots, f_m$  de um determinado ângulo  $\psi$  através de uma transformação ortogonal. O procedimento é o seguinte: Dado um plano formado pelos fatores  $p$  e  $q$ , determinamos o padrão fatorial tal que forneça o máximo valor para  $Q$  considerando apenas a rotação desse plano  $pq$ . Chamemos essa função de  $Q_{pq}$ . Repetimos esse procedimento até que sejam feitas todas as rotações em todos os pares de fatores, fornecendo uma matriz padrão  $B$ . Todo o ciclo de opera-

ções em todos os pares de fatores é repetido, até que o valor da função  $Q$  de um dado ciclo (iteração) comparada com o seu valor no ciclo (iteração) anterior não se altere. Escreveremos a seguir esse procedimento sob a forma de algoritmo.

- Algoritmo para Rotações Quartimax

Passo 0: Dados  $K = 0$ ,  $Q_{pq}^0 = 0$ ,  $m$  fatores comuns, formar todos os pares possíveis de fatores.

Passo 1: Faça  $K = K+1$ ; para cada par  $pq$  ( $p, q = 1, \dots, m, p \neq q$ ), maximizar a função

$$Q_{pq}^k = \sum_{i=1}^n (b_{ip}^4 + b_{iq}^4)$$

Passo 2: Formar o novo padrão fatorial com a solução obtida para todos os pares e calcular  $Q^k = \sum_{i=1}^n \sum_{p=1}^m b_{ip}^4$

Passo 3: Se  $|Q^k - Q^{k-1}| \leq t_0$  pare. Senão vá para 1.

Para obtermos  $Q_{pq}$ , determinamos o ângulo  $\psi_{pq}$  de uma transformação ortogonal do tipo:

$$T_{pq} = \begin{vmatrix} \cos \psi_{pq} & -\text{sen } \psi_{pq} \\ \text{sen } \psi_{pq} & \cos \psi_{pq} \end{vmatrix}$$

Então sendo:

$B_{pq}$  = Matriz formada pelas duas colunas p e q após a rotação

$A_{pq}$  = Matriz original, tomando-se apenas as colunas p e q.

Podemos escrever:

$$B_{pq} = A_{pq} T_{pq}$$

ou ainda:

$$b_{ip} = a_{ip} \cos \psi_{pq} + a_{iq} \sin \psi_{pq} \quad (3.20a)$$

$$b_{iq} = -a_{ip} \sin \psi_{pq} + a_{iq} \cos \psi_{pq} \quad (3.20b)$$

A função

$$Q_{pq}(\psi_{pq}) = \sum_{i=1}^n (b_{ip}^2 + b_{iq}^2) \quad (3.21)$$

passa a ser uma função do parâmetro  $\psi_{pq}$ .

O problema é determinar  $\psi_{pq}$  que maximize a função  $Q_{pq}$ . Para não sobrecarregar a notação usaremos  $Q(\psi)$ , mas lembrando que estamos trabalhando no plano pq.

Substituindo as eq. (3.20a) e (3.20b) em (3.21)

temos:

$$b_{ip}^4 = (a_{ip} \cos \psi + a_{iq} \sin \psi)^4$$

$$b_{iq}^4 = (-a_{ip} \sin \psi + a_{iq} \cos \psi)^4$$

$$Q(\psi) = \sum_{i=1}^n [(a_{ip} \cos \psi + a_{iq} \sin \psi)^4 + (-a_{ip} \sin \psi + a_{iq} \cos \psi)^4]$$

O ângulo  $\psi^*$  que maximiza a função  $Q$ , é tal que:

$$\left. \begin{array}{l} \frac{\partial Q(\psi^*)}{\partial \psi} = 0 \\ \frac{\partial^2 Q(\psi^*)}{\partial \psi^2} < 0 \end{array} \right\} \begin{array}{l} \rightarrow \text{condições necessárias e suficientes para} \\ \psi^* \text{ seja um ponto de máximo da função } Q. \end{array} \quad \begin{array}{l} \text{que} \\ \text{Ver} \\ |6| \end{array}$$

Podemos derivar cada parcela da função  $Q$  e somar para todo  $i$  obtendo  $\frac{\partial Q(\psi)}{\partial \psi}$ :

$$\begin{aligned} b_{ip}^4 &= (a_{ip} \cos \psi + a_{iq} \sin \psi)^4 = a_{ip}^4 \cos^4 \psi + a_{iq}^4 \sin^4 \psi + \\ &+ 4a_{ip}^3 a_{iq} \cos^3 \psi \sin \psi + 4a_{ip} a_{iq}^3 \cos \psi \sin^3 \psi + \\ &+ 6a_{ip}^2 a_{iq}^2 \cos^2 \psi \sin^2 \psi. \end{aligned}$$

$$\begin{aligned} b_{iq}^4 &= (-a_{ip} \sin \psi + a_{iq} \cos \psi)^4 = a_{ip}^4 \sin^4 \psi + a_{iq}^4 \cos^4 \psi - \\ &- 4a_{ip}^3 \sin^3 \psi a_{iq} \cos \psi - 4a_{ip} a_{iq}^3 \sin \psi \cos^3 \psi + \\ &+ 6a_{ip}^2 a_{iq}^2 \cos^2 \psi \sin^2 \psi. \end{aligned}$$

$$\begin{aligned}
b_{ip}^4 + b_{iq}^4 &= (a_{ip}^4 + a_{iq}^4) \operatorname{sen}^4 \psi + (a_{ip}^4 + a_{iq}^4) \operatorname{cos}^4 \psi + \\
&+ a_{ip}^3 a_{iq} (4 \operatorname{cos}^3 \psi \operatorname{sen} \psi - 4 \operatorname{sen}^3 \psi \operatorname{cos} \psi) + \\
&+ a_{iq}^3 a_{ip} (4 \operatorname{cos} \psi \operatorname{sen}^3 \psi - 4 \operatorname{cos}^3 \psi \operatorname{sen} \psi) + 12 a_{ip}^2 a_{iq}^2 \operatorname{cos}^2 \psi \operatorname{sen}^2 \psi
\end{aligned}$$

Mas na 3.<sup>a</sup> e 4.<sup>a</sup> parcelas podemos substituir:

$$\operatorname{sen} 4\psi = 2(\operatorname{sen} 2\psi \operatorname{cos} 2\psi) = 2[(2 \operatorname{sen} \psi \operatorname{cos} \psi)(\operatorname{sen}^2 \psi - \operatorname{cos}^2 \psi)]$$

$$\operatorname{sen} 4\psi = 4 \operatorname{sen}^3 \psi \operatorname{cos} \psi - 4 \operatorname{cos}^3 \psi \operatorname{sen} \psi \quad (3.22)$$

Podemos então reescrever:

$$\begin{aligned}
b_{ip}^4 + b_{iq}^4 &= (a_{ip}^4 + a_{iq}^4) \operatorname{sen}^4 \psi + (a_{ip}^4 + a_{iq}^4) \operatorname{cos}^4 \psi - \\
&- a_{ip}^3 a_{iq} \operatorname{sen} 4\psi + a_{iq}^3 a_{ip} \operatorname{sen} 4\psi + 12 a_{ip}^2 a_{iq}^2 \operatorname{cos}^2 \psi \operatorname{sen}^2 \psi.
\end{aligned}$$

Derivando em relação a  $\psi$ :

$$\begin{aligned}
\frac{\partial (b_{ip}^4 + b_{iq}^4)}{\partial \psi} &= 4(a_{ip}^4 + a_{iq}^4) \operatorname{sen}^3 \psi \operatorname{cos} \psi - 4(a_{ip}^4 + a_{iq}^4) \operatorname{cos}^3 \psi \operatorname{sen} \psi - \\
&- 4a_{ip}^3 a_{iq} \operatorname{cos} 4\psi + 4a_{iq}^3 a_{ip} \operatorname{cos} 4\psi + \\
&+ 24 a_{ip}^2 a_{iq}^2 (-\operatorname{cos} \psi \operatorname{sen}^3 \psi + \operatorname{sen} \psi \operatorname{cos}^3 \psi)
\end{aligned}$$

Mas pela eq. (3.22), a última parcela pode ser escrita:

$$24a_{ip}^2 a_{iq}^2 \left(-\frac{1}{4} \sin 4\psi\right) = -6a_{ip}^2 a_{iq}^2 \sin 4\psi$$

E agrupando as duas primeiras parcelas, e utilizando ainda o resultado da eq. (3.22)

$$(a_{ip}^4 + a_{iq}^4)(4\sin^3\psi \cos\psi - 4\cos^3\psi \sin\psi) = (a_{ip}^4 + a_{iq}^4)\sin 4\psi$$

A expressão fica da seguinte forma:

$$\begin{aligned} \frac{\partial(b_{ip}^4 + b_{iq}^4)}{\partial\psi} &= (a_{ip}^4 + a_{iq}^4)\sin 4\psi + (4a_{iq}^3 a_{ip} - 4a_{ip}^3 a_{iq})\cos 4\psi \\ &\quad - 6a_{ip}^2 a_{iq}^2 \sin 4\psi. \end{aligned}$$

$$\begin{aligned} \frac{\partial(b_{ip}^4 + b_{iq}^4)}{\partial\psi} &= (a_{ip}^4 + a_{iq}^4 - 6a_{ip}^2 a_{iq}^2)\sin 4\psi + \\ &\quad + (4a_{ip}^3 a_{ip} - 4a_{ip}^3 a_{iq})\cos 4\psi. \end{aligned}$$

Somando para todo  $i$ , obtemos a expressão da derivada da função  $Q$  em relação a  $\psi$ :

$$\begin{aligned} \frac{\partial Q(\psi)}{\partial\psi} &= \sum_{i=1}^n (a_{ip}^4 + a_{iq}^4 - 6a_{ip}^2 a_{iq}^2)\sin 4\psi + \\ &\quad + \sum_{i=1}^n (4a_{ip}^3 a_{ip} - 4a_{ip}^3 a_{iq})\cos 4\psi. \end{aligned}$$

Aplicando a primeira condição de máximo:



$$\frac{\partial Q(\psi)}{\partial \psi} = 0$$

$$\sum_{i=1}^n (4a_{iq}^3 a_{ip} - 4a_{ip}^3 a_{iq}) \cos 4\psi = -$$

$$- \sum_{i=1}^n (a_{ip}^4 + a_{iq}^4 - 6a_{ip}^2 a_{iq}^2) \sin 4\psi$$

Chamando:

$$v = \sum_{i=1}^n (4a_{iq}^3 a_{ip} - 4a_{ip}^3 a_{iq})$$

$$\delta = - \sum_{i=1}^n (a_{ip}^4 + a_{iq}^4 - 6a_{ip}^2 a_{iq}^2)$$

temos:

$$v \cos 4\psi = \delta \sin 4\psi \quad (3.23)$$

$$\operatorname{tg} 4\psi = - \frac{\sum_{i=1}^n (4a_{iq}^3 a_{ip} - 4a_{ip}^3 a_{iq})}{\sum_{i=1}^n (a_{ip}^4 + a_{iq}^4 - 6a_{ip}^2 a_{iq}^2)} = \frac{v}{\delta} \quad (3.24)$$

Aplicando a segunda condição de máximo:

$$\frac{\partial^2 Q}{\partial \psi^2} = -4v \sin 4\psi - 4\delta \cos 4\psi < 0$$

Da equação (3.23) temos

$$\cos 4\psi = \frac{\delta \operatorname{sen} 4\psi}{v}$$

Substituindo na segunda condição de máximo

$$-4v \operatorname{sen} 4\psi - \frac{4\delta^2 \operatorname{sen} 4\psi}{v} < 0$$

$$- \frac{(v^2 + \delta^2)}{v} \operatorname{sen} 4\psi < 0$$

Como  $(v^2 + \delta^2) > 0$ , a condição se reduz a:

$$- \frac{\operatorname{sen} 4\psi}{v} < 0$$

$$\frac{\operatorname{sen} 4\psi}{v} > 0$$

Podemos concluir que  $v$  deve ter o mesmo sinal que  $\operatorname{sen} 4\psi$ . O período da função  $\operatorname{sen} 4\psi$  é  $90^\circ$ , portanto ao fazermos um estudo da sua variação de sinal, basta considerar um intervalo qualquer com essa amplitude. Esse estudo é apresentado na tabela 3.7.

Tabela 3.7

SINAL DE $v$ E DE $\sin 4\psi$	SINAL DE $\delta$	SINAL DE $\operatorname{tg} 4\psi$	SINAL DE $\cos 4\psi$	QUADRANTE $4\psi$	LIMITE PARA $\psi$
+	+	+	+	1º quad.	$0^\circ$ a $22,5^\circ$
+	-	-	-	2º quad.	$22,5^\circ$ a $45^\circ$
-	-	+	-	3º quad.	$-45^\circ$ a $-22,5^\circ$
-	+	-	+	4º quad.	$-22,5^\circ$ a $0^\circ$

O ângulo  $\psi^*$  que maximiza  $Q(\psi)$  é obtido através da expressão para  $\operatorname{tg} 4\psi$  eq. (3.24), que só depende dos coeficientes da matriz padrão inicial  $A$ . Ou seja, uma vez obtida a solução inicial  $A$ , podemos calcular a  $\operatorname{tg} 4\psi^*$ . Como:

$$\operatorname{tg} \theta = \operatorname{tg}(180^\circ + \theta),$$

o ângulo  $4\psi^*$  não fica bem determinado. Para isso utilizamos a tabela 3.7 para determinar o quadrante do ângulo  $4\psi^*$  e consequentemente o valor de  $\psi^*$  que será a solução do problema.

#### - O Método Varimax

No método de rotação Quartimax, como vimos, buscamos simplificar a matriz padrão maximizando a variância dos seus elementos, obtendo coeficientes altos e coeficientes baixos. Mas nesse método não houve preocupação sobre a localização desses coeficientes. Portanto podíamos ter uma solução Quartimax que fornecesse uma coluna de coeficientes altos, e as demais com coeficientes baixos. O objetivo do método esta-

ria satisfeito, mas nesse caso teríamos um fator geral, que segundo Thrustone deve ser evitado. Para resolver esse problema, Kaiser em 1958 [16] propôs o método de rotação Varimax, que tem como objetivo minimizar o número de variáveis presentes em cada fator, ou seja, ao invés de simplificar a matriz padrão como um todo, simplifica a estrutura dos fatores, ou ainda, simplifica a matriz padrão por colunas.

A estrutura de um fator será simplificada quando este apresentar um número grande de coeficientes baixos (próximos de zero) e um número pequeno de coeficientes altos. Então, aqui como no método Quartimax, estaremos interessados em maximizar a variação dos coeficientes de um dado fator. Aqui também tomaremos os quadrados dos coeficientes pois estamos interessados no seu valor absoluto. E a maneira de maximizar essa variação entre os coeficientes será ainda através da máxima variância de sua distribuição. Assim, pela definição de variância, temos para um dado fator  $f_p$ :

$$S_p^2 = V_p = \frac{1}{n} \sum_{i=1}^n (b_{ip}^2 - \bar{b}_p^2)^2$$

onde:

$$\bar{b}_p^2 = \frac{1}{n} \sum_{i=1}^n b_{ip}^2$$

Comparando com a função objetivo do método Quartimax, vemos que lá calculávamos a variância de todos os coeficientes da matriz padrão em relação a média total:

$$S_{b^2}^2 = \frac{1}{mn} \sum_{i=1}^n \sum_{p=1}^m (b_{ip}^2 - \bar{b}^2)^2$$

$$\text{onde } \bar{b}^2 = \frac{1}{mn} \sum_{i=1}^n \sum_{p=1}^m b_{ip}^2$$

Agora, a função  $V_p$  fornece a variância dos coeficientes de uma dada coluna  $p$  em relação a média dos coeficientes dessa coluna.

Assim:

$$V_p = \frac{1}{n} \sum_{i=1}^n (b_{ip}^4 - 2b_{ip}^2 \bar{b}_p^2 + (\bar{b}_p^2)^2)$$

$$V_p = \frac{1}{n} \sum_{i=1}^n b_{ip}^4 - \frac{2}{n} \sum_{i=1}^n b_{ip}^2 \frac{1}{n} \sum_{i=1}^n b_{ip}^2 + \left( \frac{1}{n} \sum_{i=1}^n b_{ip}^2 \right)^2$$

$$V_p = \frac{1}{n} \sum_{i=1}^n b_{ip}^4 - 2 \left( \frac{1}{n} \sum_{i=1}^n b_{ip}^2 \right)^2 + \left( \frac{1}{n} \sum_{i=1}^n b_{ip}^2 \right)^2$$

$$V_p = \frac{1}{n} \sum_{i=1}^n b_{ip}^4 - \left( \frac{1}{n} \sum_{i=1}^n b_{ip}^2 \right)^2$$

Levando em conta os  $m$  fatores comuns, a função a ser maximizada será:

$$\max V' = \sum_{p=1}^m \left[ \frac{1}{n} \sum_{i=1}^n b_{ip}^4 - \left( \frac{1}{n} \sum_{i=1}^n b_{ip}^2 \right)^2 \right]$$

A função  $V'$  depende dos quadrados dos coeficientes  $b_{ip}$ . Esses coeficientes são as coordenadas das  $n$  variáveis

veis no espaço de dimensão  $m$  formado pelos fatores comuns. Para que todas as  $n$  variáveis contribuam igualmente para o problema de rotação, Saunders sugeriu que fossem normalizadas. Como a norma do vetor que descreve a variável  $z_i$  no espaço dos  $m$  fatores é:

$$||z_i|| = \sqrt{\sum_{p=1}^m b_{ip}^2} = \sqrt{h_i^2} = h_i,$$

basta dividir todos os coeficientes  $b_{ip}$  dessa variável por esse valor. Segundo esse procedimento, Kaiser apresentou o chamado método Varimax normal com a seguinte função objetivo

$$\text{Max } V = \frac{1}{n} \sum_{i=1}^m \sum_{p=1}^n (b_{ip}/h_i)^4 - \frac{1}{n^2} \sum_{p=1}^m \left( \sum_{i=1}^n b_{ip}^2/h_i^2 \right)^2$$

Chamando

$$v_{ip} = b_{ip}/h_i$$

o coeficiente normalizado da  $i$ -ésima variável para o  $p$ -ésimo fator, podemos reescrever a função objetivo  $V$ .

$$\text{Max } V = \frac{1}{n} \sum_{p=1}^m \sum_{i=1}^n v_{ip}^4 - \frac{1}{n^2} \sum_{p=1}^m \left( \sum_{i=1}^n v_{ip}^2 \right)^2$$

Para maximizar  $V$ , o procedimento é análogo ao usado no método Quartimax. O problema será resolvido para cada par de fatores  $p$  e  $q$  obtendo  $V_{pq}$ :

$$V_{pq} = \frac{1}{n} \left( \sum_{i=1}^n v_{ip}^4 + \sum_{i=1}^n v_{iq}^4 \right) - \frac{1}{n^2} \left[ \left( \sum_{i=1}^n y_{ip}^2 \right)^2 + \left( \sum_{i=1}^n y_{iq}^2 \right)^2 \right]$$

Levando em conta que os coeficientes  $v_{ip}$  serão determinados por uma rotação ortogonal

$$T = \begin{vmatrix} \cos \psi & - \operatorname{sen} \psi \\ \operatorname{sen} \psi & \cos \psi \end{vmatrix}$$

temos:

$$v_{ip} = \frac{b_{ip}}{h_i} = \frac{a_{ip}}{h_i} \cos \psi + \frac{a_{iq}}{h_i} \operatorname{sen} \psi$$

$$v_{iq} = \frac{b_{iq}}{h_i} = - \frac{a_{ip}}{h_i} \operatorname{sen} \psi + \frac{a_{iq}}{h_i} \cos \psi$$

Dessa forma, escrevemos  $V_{pq}$  como função do ângulo  $\psi$  no plano  $pq$ .

Aplicando as condições de máximo na função  $V_{pq}$

$$\frac{\partial V_{pq}(\psi)}{\partial \psi} = 0$$

$$\frac{\partial^2 V_{pq}(\psi)}{\partial \psi^2} < 0$$

cairemos numa equação trigonométrica cuja solução apresentamos abaixo:

$$\operatorname{tg}4\psi = \frac{D - 2AB/n}{C - (A^2 - B^2)/n}$$

onde:

$$A = \sum_{i=1}^n (v_{ip}^2 - v_{iq}^2)$$

$$B = \sum_{i=1}^n 2v_{ip} v_{iq}$$

$$C = \sum_{i=1}^n [(v_{ip}^2 - v_{iq}^2)^2 - (2v_{ip} v_{iq})^2]$$

$$D = 2 \sum_{i=1}^n (v_{ip}^2 - v_{iq}^2) 2v_{ip} v_{iq}$$

A solução completa desse problema pode ser vista em [17]. Aqui novamente vale a utilização da tabela 3.7 para a determinação do ângulo  $\psi^*$ . Assim determinamos a solução para maximizar  $V_{pq}$ . Como no método Quartimax, esse procedimento é repetido para cada par de fatores formando um ciclo, e serão feitos tantos ciclos quantos forem necessários para que o acréscimo na função objetivo seja desprezível (menor que uma tolerância pré-fixado).

### - Exemplo 3.3 - 24 Testes Psicológicos

Mostraremos para efeito de comparação dos dois métodos de rotação, a solução para o problema dos 24 testes



psicológicos apresentado na seção 3.4, depois de efetuadas rotações Quartimax e Varimax.

Tabela 3.8 - Soluções Quartimax e Varimax

$x_i$	QUARTIMAX				VARIMAX			
	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>4</sub>	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>4</sub>
1	.37	.19	.60	.07	.14	.19	.67	.17
2	.24	.07	.38	.04	.10	.07	.43	.10
3	.31	.01	.48	.01	.15	.02	.54	.08
4	.36	.07	.46	-.01	.20	.09	.54	.07
5	.81	.14	-.02	-.04	.75	.21	.22	.13
6	.81	.03	.00	.06	.75	.10	.23	.21
7	.85	.07	-.04	-.10	.82	.16	.21	.08
8	.66	.20	.20	-.04	.54	.26	.38	.12
9	.86	-.06	-.02	.10	.80	.01	.22	.25
10	.23	.70	-.12	.11	.15	.70	-.06	.24
11	.31	.62	.01	.23	.17	.60	.08	.36
12	.16	.69	.19	-.01	.02	.69	.23	.11
13	.35	.57	.32	-.08	.18	.59	.41	.06
14	.32	.19	-.03	.42	.22	.16	.04	.50
15	.25	.11	.10	.45	.12	.07	.14	.50
16	.29	.13	.37	.37	.08	.10	.41	.43
17	.28	.24	.02	.57	.14	.18	.06	.64
18	.22	.32	.30	.47	.00	.26	.32	.54
19	.28	.18	.19	.32	.13	.16	.24	.39

continua...

continuação

$x_i$	QUARTIMAX				VARIMAX			
	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>4</sub>	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>4</sub>
20	.52	.09	.35	.14	.35	.11	.47	.25
21	.35	.38	.35	.14	.15	.38	.42	.26
22	.53	.04	.30	.26	.36	.04	.41	.36
23	.55	.19	.44	.09	.35	.21	.57	.22
24	.49	.43	.10	.20	.34	.44	.22	.34
$V_p$	5.59	2.42	1.96	1.42	3.50	2.44	3.08	2.36

Foram mantidos quatro fatores após a rotação ortogonal. Aplicando os cinco critérios de estrutura simples de Thurstone, vemos que a solução Varimax satisfaz a todos eles, e a Quartimax não satisfaz ao 2º critério, segundo o qual espera-se que cada fator tenha no mínimo  $m$  coeficientes não significativos, onde  $m$  é o número de fatores comuns. Considerando não significativos os coeficientes  $< 0,2$ , o primeiro fator extraído na solução Quartimax, só apresenta um coeficiente com essa característica. Ou seja esse fator se apresenta como um fator geral, tal qual na solução inicial (sem rotação). Comparando essas duas soluções com as apresentadas nas tabelas 3.3 e 3.4 (solução F.P. e Res.Min.) veremos que alguns coeficientes são muito mais altos nas soluções com rotação, o que permite uma melhor interpretação dos fatores. Assim sendo:

Fator 1: alta correlação ( $r \geq .5$ ) com variáveis:

Varimax → 5, 6, 7, 8, 9

Quartimax → 5, 6, 7, 8, 9, 20, 22, 23

Fator 2: alta correlação com as variáveis:

Varimax: 10, 11, 12, 13

Quartimax: 10, 11, 12, 13

Fator 3: alta correlação com as variáveis:

Varimax: 1, 3, 4, 23

Quartimax: 1

Fator 4: alta correlação com as variáveis:

Varimax: 14, 15, 17, 18

Quartimax: 17.

A partir dessa análise da solução concluímos que a solução Varimax fornece grupos de variáveis para cada fator, enquanto que a solução Quartimax, não fornece senão um fator (o fator 2) associado a um grupo de variáveis, considerando que o 1º fator tem características gerais, e portanto não corresponderia a um grupo, no sentido classificatório. Psicólogos, analisando a solução Varimax, de acordo com a natureza dos testes que estão altamente correlacionados com os fatores, sugeriram a seguinte interpretação:

Fator 1: Associado à expressão verbal

Fator 2: Associado a velocidade de raciocínio

Fator 3: Dedução

Fator 4: Memória

Com essa solução conseguimos agrupar as variáveis em 4 fatores comuns e também reduzir o número de dados, uma vez que muitas das variáveis iniciais não aparecem altamente correlacionadas, com nenhum dos fatores. Para escolher quais dessas variáveis podem ser excluídas do estudo, leva-se em conta, não apenas o valor absoluto dos seus coeficientes associados aos fatores, mas também sua confiabilidade, facilidade de medida, abrangência etc. Não existe uma regra para excluir essa ou aquela variável, e cada caso deve ser estudado separadamente de acordo com o objetivo final do estudo. Algumas vezes a redução do problema não é tão importante, preferindo-se manter uma descrição mais exata dos dados iniciais. Outras vezes o que se quer é uma redução drástica no número de variáveis. Obviamente, o critério utilizado na eliminação das variáveis será diferente nos dois casos.

## CAPÍTULO 4

### OUTROS MÉTODOS DE ANÁLISE FATORIAL

O objetivo desse capítulo é principalmente informar a respeito de outros métodos de Análise Fatorial largamente utilizados pelos analistas. As apresentações serão breves, e o leitor interessado em algum desses métodos em particular deve procurar o seu desenvolvimento completo nas referências citadas. Ainda nesse capítulo apresentaremos dois métodos de rotação que diferem dos até aqui mencionados, por relaxarem a condição de ortogonalidade para os fatores comuns. Serão introduzidas então as alterações na solução do modelo para o caso oblíquo.

#### - Seção 4.1 - Outros Métodos Ortogonais

Nessa seção veremos três métodos de solução ortogonal a partir do mesmo modelo básico de Análise Fatorial mas com enfoques diferentes dos até aqui apresentados. O primeiro é baseado na análise de correlação canônica, e os outros dois utilizam o conceito de universo de variáveis que será apresentado na ocasião.

#### - Análise Fatorial Canônica

Antes de apresentarmos esse método, daremos uma idéia a grosso modo do que é uma análise de correlação canôni-

ca. Essa análise é parecida com a análise de regressão múltipla, onde uma variável é dada em função de um conjunto de outras variáveis. O coeficiente de correlação múltipla indica quanto da variância dessa variável é explicada pelo conjunto das outras. Na análise de correlação canônica, um conjunto de variáveis é dado em função de um outro conjunto. A correlação canônica mede o quanto da variância de um conjunto é explicado pela variância do outro.

Na Análise Fatorial Canônica, o modelo utilizado é o mesmo da Análise dos Fatores Principais, mas ao invés de procurar os fatores que maximizam a variância das variáveis, aqui procuramos fatores que podem ser melhor previstos a partir dos dados. Em outras palavras, procura-se estimar os fatores comuns de tal forma que o primeiro fator tenha máxima correlação com os dados. O segundo terá máxima correlação com os dados mas será ortogonal ao primeiro. E assim sucessivamente. A solução desse problema é obtida através da análise de correlação canônica das variáveis hipotéticas  $f_p$  ( $p = 1, 2, \dots, m$ ) e as variáveis medidas  $z_i$  ( $i = 1, 2, \dots, n$ ). Tomando a matriz formada pela matriz dos dados observados  $Z(n \times N)$  e a matriz dos fatores comuns  $F(m \times N)$ , temos:

$$D((n+m) \times N) = \begin{vmatrix} Z \\ F \end{vmatrix}$$

A matriz de correlação total será:

$$R_T = \frac{1}{N} DD^t = \frac{1}{N} \begin{vmatrix} Z \\ F \end{vmatrix} \begin{vmatrix} Z^t & F^t \end{vmatrix}$$

$$R_T = \frac{1}{N} \begin{vmatrix} ZZ^t & ZF^t \\ FZ^t & FF^t \end{vmatrix}$$

Supondo os fatores ortogonais, temos:

$$\frac{1}{N} FF^t = I$$

Então:

$$R_T((n+m) \times (n+m)) = \begin{vmatrix} R & A \\ A^t & I \end{vmatrix}$$

Resolvendo-se esse problema através da análise de correlação canônica chega-se a seguinte equação para o quadrado das correlações canônicas ( $\nu$ )

$$\det |AA^t - \nu R| = 0$$

Ou, usando a variância específica das variáveis:

$$|R - D^2| = R^* = AA^t$$

$$\det |R - D^2| - \nu R| = 0$$

O desenvolvimento completo desse problema pode ser visto em [13].

### - Análise Alfa

Quando nos referimos a estudos estatísticos, geralmente os elementos da amostra são pessoas, entidades, regiões, sobre as quais fazemos medidas. Nesse caso a generalização dos resultados da amostra sobre os outros elementos da população, é o que chamamos inferência estatística. Aqui veremos uma situação inversa, onde os elementos sobre os quais fazemos as medidas formam uma população, mas as  $n$  medidas são uma amostra de um universo de variáveis. Agora, as generalizações feitas sobre os resultados obtidos para essa amostra são referentes a inferências na área das variáveis. Para diferenciar do caso acima descrito, a essa generalização chamou-se inferência psicométrica, já que grande parte do trabalho feito envolvendo o conceito de universo de variáveis é no campo da psicologia.

Em 1965, Kaiser e Caffrey [19], desenvolveram um método chamado Análise Alfa, para solução do modelo de A.F., usando esse conceito de inferência psicométrica, no qual o princípio básico é determinar os fatores comuns da amostra de  $n$  variáveis de tal forma que tenham máxima correlação com os fatores comuns correspondentes no universo de variáveis. O quadrado dessa correlação é chamado coeficiente de generalização ou coeficiente  $\alpha$ . Assim, a determinação desses fatores, im



plica na maximização desse coeficiente.

Os fatores comuns são descritos como combinações lineares das partes comuns das variáveis:

$$f_p = \sum_{i=1}^n w_{ip} Z_i'' = w_p z'' \quad (p = 1, 2, \dots, m)$$

De acordo com Kaiser e Caffrey, o coeficiente de generalização é:

$$\alpha = \frac{n}{n-1} \left[ 1 - \frac{w^t H^2 w}{w^t (R - D^2) w} \right] \quad \text{onde:}$$

$H^2$  é a matriz diagonal (nxn) formada pelas comunalidade  $h_i^2$  ( $i = 1, 2, \dots, n$ ).

$D^2$  é a matriz diagonal (nxn) formada pela parte específica das variâncias das variáveis  $d_i^2$  ( $i = 1, 2, \dots, n$ ). O problema de maximizar  $\alpha$  passa então a ser:

Determinar  $w$  tal que:

$$\min \frac{w^t H^2 w}{w^t (R - D^2) w}$$

A solução completa e uma discussão sobre o número de fatores para esse problema está em [19].

### - Análise Fatorial da Imagem

Nesse método considera-se ainda o conceito de universo de variáveis introduzido na descrição do método de Análise Alfa. Guttman em 1953 [7] desenvolveu esse método tendo em vista o problema da estimativa das comunalidades. Na Análise da Imagem a parte comum de uma variável é chamada imagem da variável e é definida como a estimativa da regressão múltipla sobre todas as  $n-1$  variáveis restantes. Como vimos na seção 3.3, através da regressão múltipla procura-se escrever uma variável ( $z_i$ ) em função de um conjunto ( $z_k, k = 1, 2, \dots, n, k \neq i$ ). Segundo o método da Imagem, a estimativa que se obtém nessa regressão múltipla é a parte comum da variável  $z_i$ . Assim, sendo  $\beta_{ik}$  os coeficientes dessa regressão temos:

$$\tilde{z}_i = \sum_{\substack{k=1 \\ k \neq i}}^n \beta_{ik} z_k$$

onde  $\tilde{z}_i$  é a estimativa da parte comum da variável  $z_i$ .

A variável observada pode ser escrita da seguinte forma:

$$z_i = \tilde{z}_i + e_i \quad \text{onde}$$

$e_i$  = resíduo da regressão múltipla = parte específica + erro, também chamada anti-imagem da variável  $z_i$ .

A matriz a ser analisada é a matriz de covariância estimada ou matriz de covariância da imagem.

$$G = \bar{z} \bar{z}^t$$

cuja diagonal contém o quadrado da correlação múltipla (S.M.C.) de cada variável com as  $n-1$  restantes. A relação entre a matriz de covariância da imagem  $G$  e a matriz de correlação das variáveis  $R$  foi estabelecida por Guttman:

$$R = G - \Gamma + 2S^2$$

onde  $\Gamma$  é a matriz de covariância dos resíduos ou das anti-imagens, e  $S^2$  é a matriz diagonal das variâncias dos resíduos ou das anti-imagens.

Foi demonstrado no desenvolvimento desse método, que a verdadeira comunalidade de uma variável será o quadrado da correlação múltipla (S.M.C.) desta com todas as demais no universo de variáveis. Se o conjunto das  $n$  variáveis representa satisfatoriamente esse universo, esses serão os valores das comunalidades. O método A.I. usa esses valores fixos para as comunalidades. No entanto qualquer matriz de correlação  $R$  será sempre simétrica semi definida positiva. Ao inserirmos os valores das S.M.C. na sua diagonal, pode-se obter uma matriz que não tenha essas características. Daí, o que a A.I. faz é fixar os valores na diagonal de  $R$  como sendo as S.M.C. das variáveis, e ajustar os elementos fora da diagonal, de tal maneira que a matriz permaneça simétrica semi definida positiva. Então ao contrário da A.F. tradicional, na A.I. a diagonal da matriz de correlação é fixa (S.M.C.), enquanto que os valores fora da diagonal são alterados.

#### 4.2 - ROTAÇÃO OBLÍQUA

Até aqui apresentamos métodos de solução e rotação nos quais os fatores comuns eram obrigatoriamente ortogonais. Do ponto de vista matemático é sempre muito melhor lidar com uma solução que envolva fatores ortogonais, o que garante resultados mais simples e facilidades de manipulação e análises posteriores. Mas na verdade, supor a ortogonalidade dos fatores comuns é supor que eles são não correlacionados, ou seja, estatisticamente independentes. E obviamente isso nem sempre ocorre.

Obtendo-se uma solução oblíqua, os grupos de variáveis correspondentes aos fatores comuns ficam melhor definidos, e como podemos calcular as correlações entre os fatores, teremos as correlações entre os diversos grupos de variáveis. Caso esses grupos não se correlacionem, cairemos numa solução ortogonal, que pode ser entendida como um caso particular da solução oblíqua. A Figura (4.1) mostra um caso em que os grupos de variáveis são correlacionados. Os fatores que melhor representam os grupos formados são os obtidos pela rotação oblíqua.

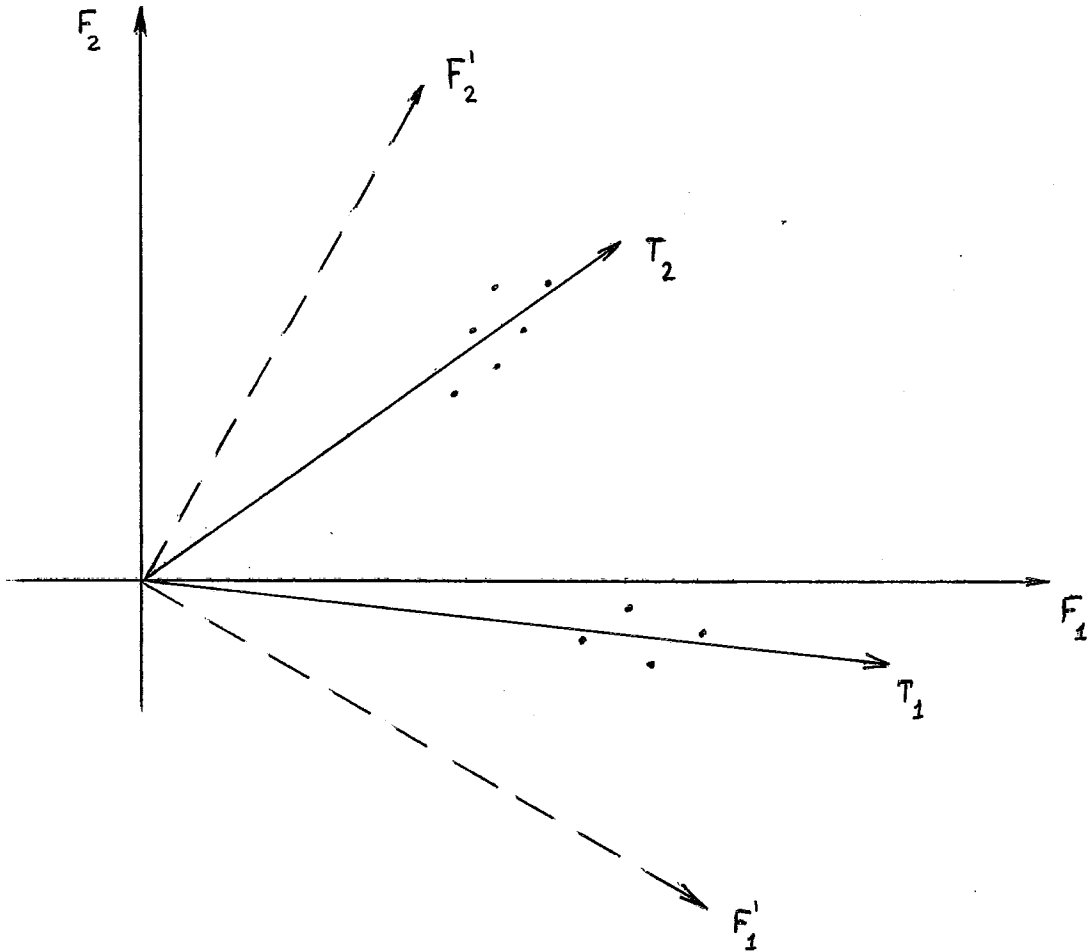


Figura 4.1 - Dois grupos de variáveis correlacionadas.

$F_1 \times F_2 \rightarrow$  Solução inicial

$F_1' \times F_2' \rightarrow$  Solução com rotação ortogonal

$T_1 \times T_2 \rightarrow$  Solução com rotação oblíqua

A desvantagem que a solução oblíqua apresenta, é em relação a interpretação, que como veremos mais tarde, consiste da análise de duas matrizes (padrão e estrutura) contra uma (padrão) no caso ortogonal.

Considerando o caso geral da solução oblíqua, refaremos alguns cálculos apresentados no Capítulo 2, onde par-

tíamos da hipótese da ortogonalidade dos fatores. Como foi visto então, a parte comum do modelo, segundo a notação matricial introduzida na seção 2.5 tem a seguinte forma:

$$Z'' = A F$$

Para calcular as correlações entre as variáveis, usando a definição de coeficiente de correlação:

$$r_{z_i z_k} = \frac{1}{N} \sum_{j=1}^N z_{ij} z_{kj}$$

Usando a notação matricial temos:

$$R = \frac{1}{N} Z Z^t$$

Para obtermos as correlações reproduzidas, substituímos as variáveis observadas pelos seus valores obtidos pelo modelo. Como sō estamos levando em conta a parte comum do modelo, obteremos a matriz de correlação reduzida, ou seja:

$$R^* = \frac{1}{N} Z'' (Z'')^t$$

$$R^* = \frac{1}{N} A F F^t A^t$$

Designando por  $\phi$  a matriz de correlação entre os fatores, isto é:

$$\phi = \frac{1}{N} FF^t$$

Podemos escrever:

$$R^* = A \phi A^t$$

Observamos agora, que no caso dos fatores serem ortogonais, a matriz  $\phi = I$  e caímos no resultado obtido no Capítulo 2:

$$R^* = AA^t.$$

Outra informação importante é a respeito da correlação entre cada variável  $z_i^1$  e cada fator  $f_p$ , que indica quanto cada variável é relevante para a determinação de cada fator comum. Ainda pela definição de coeficiente de correlação podemos escrever:

$$r_{z_i^1 f_p}^1 = \frac{1}{N} \sum_{j=1}^N z_{ij}^1 f_{pj}$$

Essas correlações fornecem o que chamamos de estrutura dos fatores, e a matriz de correlações assim formada é a matriz estrutura.

Chamando  $r_{z_i^1 f_p}^1 = s_{ip}$  e tomando a parte comum da variável, temos:

$$S = \frac{1}{N} Z'' F^t$$

$$S = \frac{1}{N} A F F^t$$

$$S = A \phi$$

Mais uma vez lembramos que no caso ortogonal temos  $\phi = I$ , o que implica em  $S = A$ . Ou seja, quando os fatores comuns não são correlacionados, as correlações entre as variáveis  $z_i^!$  ( $i = 1, 2, \dots, n$ ) e os fatores  $f_p$  ( $p = 1, 2, \dots, m$ ) são os próprios coeficientes da matriz padrão fatorial. Por isso na solução ortogonal a matriz padrão nos fornece toda a informação necessária. Já na solução oblíqua o padrão fornecerá os coeficientes dos fatores na descrição das variáveis, que são as coordenadas dessas variáveis no sistema de eixos formado pelos fatores comuns. Enquanto que a matriz estrutura fornecerá as correlações entre as variáveis e os fatores.

Apresentaremos agora métodos através dos quais a partir de uma solução ortogonal inicial, obtemos uma solução oblíqua. A transformação envolvida é tal que aplicada ao padrão fatorial ortogonal obtenho a estrutura oblíqua da nova solução. Isto é:

$S = AT$ , onde:

$A \rightarrow$  padrão ortogonal

$T \rightarrow$  transformação oblíqua

$S \rightarrow$  estrutura da nova solução que fornece a correlação entre os novos fatores  $t_1, t_2, \dots, t_m$  oblíquos e as variáveis  $z_i^!$ .



Mas como acabamos de ver, uma solução oblíqua completa contém além da estrutura, um padrão fatorial que fornece a descrição linear das variáveis segundo o novo referencial formado pelos fatores oblíquos. Enquanto A era o padrão ortogonal, P agora é o novo padrão oblíquo. Usando a relação entre padrão e estrutura fatorial, temos:

$$S = P \Phi$$

onde  $\Phi$  é a matriz de correlação entre os fatores comuns oblíquos, e dessa forma completamos as informações necessárias. Os dois métodos apresentados a seguir, tem como objetivo obter a estrutura fatorial da solução oblíqua.

#### - Método Oblimax

Esse método desenvolvido no trabalho de Pinzka e Saunders [26], considera a transformação  $S = AT$  na qual, seguindo os critérios de estrutura simples de Thurstone, os coeficientes  $s_{ip}$  da matriz S devem satisfazer a seguinte condição:

$$\text{Max } K = \sum_{i=1}^n \sum_{p=1}^m (s_{ip}^4 / (\sum_{i=1}^n \sum_{p=1}^m s_{ip}^2)^2)$$

Podemos observar que essa função K, se considerássemos o caso de fatores ortogonais onde  $\sum_{p=1}^m s_{ip}^2 = \sum_{p=1}^m a_{ip}^2 =$  constante, se reduziria a:

$$K = \frac{1}{H} \sum_{i=1}^n \sum_{p=1}^m s_{ip}^4, \text{ onde } H = \left( \sum_{i=1}^n \sum_{p=1}^m s_{ip}^2 \right)^2$$

e a função objetivo seria equivalente a apresentada no método Quartimax. Por isso pode-se dizer que o método Oblimax é a adaptação, para o caso oblíquo, do Quartimax. A solução completa desse método está em [26].

### - Métodos Oblimin

Citaremos agora uma classe de métodos de rotação oblíqua que envolvem a minimização de um certo critério. Essa classe de métodos é chamada Oblimin. A transformação envolvida ainda é do mesmo tipo isto é,  $S = AT$  onde  $S$  é a nova estrutura fatorial oblíqua e  $A$  o padrão fatorial da solução ortogonal inicial. A forma geral da função objetivo para os métodos Oblimin, segundo Carroll [2] é:

$$\min \sum_{p=1}^{m-1} \sum_{q=p+1}^m \left[ n \sum_{i=1}^n (s_{ip}^2/h_i^2)(s_{iq}^2/h_i^2) - \gamma \sum_{i=1}^n s_{ip}^2/h_i^2 \sum_{i=1}^n s_{iq}^2/h_i^2 \right]$$

onde o valor de  $\gamma$  varia para cada método desejado, sabendo-se que:

$\gamma = 1 \rightarrow$  método Covarimin - menos oblíquo

$\gamma = 0.5 \rightarrow$  método Biquartimin - medianamente oblíquo

$\gamma = 0 \rightarrow$  método Quartimin - mais oblíquo.

Qualquer outro valor de  $\gamma$  entre 0 e 1 pode ser utilizado no critério geral dos métodos Oblimin. Esse valor deverá ser escolhido de acordo com o problema, que apresentará grupos de variáveis mais ou menos correlacionados, e a solução desejada deverá ser mais ou menos oblíqua.

## CONSIDERAÇÕES FINAIS

Ao terminar esse Capítulo 4 onde foram focalizados métodos muito diferentes daqueles anteriormente apresentados, com muitos conceitos novos, tudo isso muito rapidamente, apenas com o objetivo de informar melhor o leitor, achamos conveniente, para fechar o trabalho, voltar a discutir alguns pontos sobre o modelo geral de Análise Fatorial.

1 - A A. F. pode ser vista como uma técnica exploratória dos dados observados, sendo assim, é o primeiro passo numa sequência de investigações a respeito das relações entre variáveis.

2 - A hipótese fundamental do modelo é que os dados iniciais apresentam um certo número de fatores comuns. Caso essa hipótese não seja verdadeira, a análise perde sua validade. Então caso se encontre uma solução não interpretável, pode estar ocorrendo alguns desses diferentes problemas: o número de fatores não está correto, a rotação efetuada não foi conveniente, a hipótese de que existem fatores comuns é falha, ou ainda, as relações entre variáveis e fatores, não são lineares. Todos esses problemas devem ser lembrados ao se fazer uma Análise Fatorial.

3 - Em relação a dificuldade de estimar o número de fatores comuns ressaltamos mais uma vez que para as aplicações dos testes estatísticos que verificam a validade de uma solução com um determinado número de fatores, parte-se da hipótese da multinormalidade das variáveis observadas. Nos métodos

em que essa hipótese não é necessária, caímos no problema da estimativa das comunalidades (como no método dos Fatores Principais).

4 - Uma vez obtida a solução inicial, nos deparamos com a indeterminação do modelo. Essa indeterminação como vimos, é que nos possibilita efetuarmos as rotações necessárias para melhor interpretação dos fatores. Mas como essa rotação é arbitrária (existem vários métodos ortogonais e oblíquos), pode-se chegar a soluções diferentes a partir dos mesmos dados. Portanto a escolha da solução final é da maior importância e depende fundamentalmente da experiência do analista e do seu conhecimento em relação ao problema analisado.

Com esse trabalho o que procuramos foi antes de tudo mostrar como funciona a A.F., alertando contra o uso mecânico das técnicas, uma vez que tendo-se acesso aos pacotes de estatística, conseguimos o resultado de uma A.F. sem nem ao menos termos conhecimento do modelo e suas hipóteses iniciais, não percebendo a fragilidade dessa técnica e chegando a conclusões confirmatórias em relação aos dados que a A.F. por si não pode garantir.

APÊNDICE IMULTIPLICADORES DE LAGRANGE | 20 |

Os multiplicadores de Lagrange usados para se encontrar um ponto extremo (mínimo ou máximo) de uma função real  $f(x)$  sujeita a uma ou mais restrições do tipo  $h(x) = 0$ , aparecem quando procura-se estabelecer condições necessárias para que um determinado ponto  $x^*$ , obedecendo a todos os vínculos impostos, seja solução ótima do problema. Faremos agora um resumo do estudo dessas condições necessárias.

Seja o seguinte problema de otimização com vínculos de igualdade.

$$\min f(x)$$

s.a.:

$$h_1(x) = 0$$

$$h_2(x) = 0$$

⋮

$$h_m(x) = 0$$

Onde:

$$x \in \mathbb{R}^n$$

$$f(x) \in \mathbb{R}$$

$$h_i(x) \in \mathbb{R} \quad \forall i$$

$$m < n.$$

O conjunto de vínculos  $h_i(x) = 0$  para  $i = 1, 2, \dots, m$  define uma superfície  $S$ , sub-conjunto de  $R^n$ . Essa superfície é a região viável do problema, isto é, a solução do problema tem que pertencer a  $S$ . Vamos supor que todos os vínculos  $h_i(x)$  são deriváveis. Para determinarmos as condições necessárias para que um ponto  $x^*$  seja solução do problema, precisamos introduzir dois conceitos fundamentais.

- Definição: Ponto Regular - um ponto  $x$  pertencente a  $S$  é regular se os gradientes  $\nabla h_1(x), \nabla h_2(x), \dots, \nabla h_m(x)$  são linearmente independentes.
- Definição: Plano Tangente - seja a superfície  $S$  definida por  $h_i(x) = 0$  para  $i = 1, 2, \dots, m$ . Uma curva nessa superfície é um conjunto de pontos  $x(t) \in S$  continuamente parametrizadas por  $t \in |a, b|$ . A curva  $x(t)$  passa por um ponto  $x^*$  se  $x^* = x(t^*)$  onde  $t^* \in |a, b|$ . Uma curva  $x(t)$  é diferenciável se existe  $dx(t)/dt \forall t \in |a, b|$ . Consideremos todas as curvas diferenciáveis de  $S$  passando por um ponto  $x^*$ . Um plano tangente a superfície  $S$  no ponto  $x^*$ , é definido como a coleção das derivadas de todas essas curvas, no ponto  $x^*$ .

Utilizando esses dois conceitos chegamos ao seguinte teorema:

Em um ponto regular  $x^*$  da superfície  $S$  definida por  $h_i(x) = 0, i = 1, 2, \dots, m$ , o plano tangente é:

$$M = \{y | \langle \nabla h_i(x^*), y \rangle = 0 \quad \forall i.$$

A partir desse resultado, podemos estabelecer a condição necessária para que um ponto  $x^*$  seja solução ótima do problema inicialmente proposto:

Lema: Seja  $x^*$  um ponto regular de  $S$  definida por  $h_i(x) = 0$   $i = 1, 2, \dots, m$  e ponto extremo local de  $f$  sujeito a  $h_i(x) = 0$   $i = 1, 2, \dots, m$ . Então todo  $y \in \mathbb{R}^n$  satisfazendo

$$\langle \nabla h_i(x^*), y \rangle = 0 \quad \forall i \quad (1)$$

deve também satisfazer

$$\langle \nabla f(x^*), y \rangle = 0$$

Isto é, como o conjunto dos pontos que satisfazem (1) define o plano tangente passando por  $x^*$ , o lema acima afirma que  $\nabla f(x^*)$  é ortogonal a esse plano. Isso implica que  $\nabla f(x^*)$  é uma combinação linear dos gradientes de  $h_i$   $i = 1, 2, \dots, m$  em  $x^*$ , o que conduz a introdução dos multiplicadores de Lagrange ( $\lambda_i$ ).

Teorema: Seja  $x^*$  um ponto extremo de  $f$  sujeito à restrição  $h_i(x) = 0$   $i = 1, 2, \dots, m$  e  $x^*$  é um ponto regular de  $S$ . Então existem  $\lambda_i$   $i = 1, 2, \dots, m$  nem todos nulos, tais que:

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) = 0 \quad (2)$$

Se introduzirmos a função Lagrangeana



$$L(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i h_i(x)$$

obtemos um problema de otimização sem vínculos onde a função a ser minimizada não é mais  $f(x)$  mas  $L(x, \lambda)$ .

A condição necessária para se obter uma solução ótima dada em (2) pode ser escrita:

$$\nabla_x L(x, \lambda) = 0$$

$$\nabla_{\lambda} L(x, \lambda) = 0$$

onde a segunda dessas duas equações, reproduz as equações dos vínculos.



$$\det|A - \lambda I| = 0 \quad (2)$$

Expandindo esse determinante chegaremos a um polinômio em  $\lambda$  de grau  $n$ . Para obter os valores de  $\lambda$  que satisfazem a equação (2), calculamos as  $n$  raízes desse polinômio. Essas raízes  $\lambda_i$  ( $i = 1, 2, \dots, n$ ) são os auto-valores da matriz  $A$ . O vetor  $x$  que resolve o sistema de equações quando  $\lambda$  é substituído pelo valor  $\lambda_i$ , chama-se auto vetor associado ao auto valor  $\lambda_i$ .

Pela equação característica (1) temos:

$$Ax = \lambda x .$$

Então se  $y = cx$ ,  $c$  escalar, temos:

$$Ay = Acx = cAx \text{ que por (1) é:}$$

$$cAx = c\lambda x = \lambda cx = \lambda y.$$

Concluimos portanto que  $y = cx$  ( $c$  escalar) também é auto vetor associado ao auto valor  $\lambda$ . Isto é ao resolvermos a equação de auto valores e auto vetores, os auto valores são determinados (pela equação (2)), enquanto que os auto vetores não. Temos uma família deles representada por uma direção. Para a determinação de um auto vetor, precisamos arbitrar uma norma para ele. No problema estudado na seção 3.2, esse valor já é fornecido pelo próprio problema pela equação (3.7).

$$\sum_{i=1}^n a_{i1}^2 = \lambda^1 = ||a_1||^2 \text{ então}$$

$$||a_1|| = \sqrt{\lambda^1}.$$

Dessa forma, no método dos Fatores Principais, os auto vetores que são as colunas da matriz padrão A, estão bem determinados.

### - Auto valores e auto vetor de matrizes simétricas

Na determinação dos Fatores Principais, caímos no problema de determinar os auto valores e auto vetores da matriz  $R_1$  que é a matriz de correlação reduzida, ou seja uma matriz simétrica. Vamos então estudar as propriedades dos auto valores e auto vetores dessa matriz especial.

1. Os auto valores de uma matriz real simétrica são reais.

Prova:

Suponhamos que  $Ax = \lambda x$  tem uma solução complexa. Seja  $x^*$  e  $\lambda^*$  os complexos conjugados de  $\lambda$  e  $x$  (isto é, se  $\lambda = a + bi$  então  $\lambda^* = a - bi$ ).

Multiplicando a equação característica por  $(x^*)^t$  (transposto do complexo conjugado de  $x$ ):

$$(x^*)^t Ax = (x^*)^t \lambda x \quad (3)$$

Considerando  $x^*$  e  $\lambda^*$  na equação (1):

$Ax^* = \lambda^* x^*$ , multiplicando por  $x^t$ :

$$x^t Ax^* = x^t \lambda^* x^* \quad (4)$$

Subtraindo (3) e (4):

$$(x^*)^t Ax - x^t Ax^* = (x^*)^t \lambda x - x^t \lambda^* x^*$$

Como  $(x^*)^t x = x^t x^*$ , temos:

$$(x^*)^t Ax - x^t Ax^* = (\lambda - \lambda^*) x^t x^*$$

mas como  $A = A^t$  porque  $A$  é simétrica:

$$x^t Ax^* = (x^t Ax^*)^t = (x^*)^t Ax$$

Então:

$$(\lambda - \lambda^*) x^t x^* = 0$$

se  $x \neq 0$

$$x^t x^* \neq 0 \rightarrow \lambda - \lambda^* = 0 \rightarrow \lambda = \lambda^*$$

Como  $\lambda = \lambda^*$ , isto significa que a variável complexa é igual a seu complexo conjugado, o que significa que a parte imaginária é nula (se  $a + bi = a - bi \rightarrow b = 0$ ). Logo  $\lambda = \lambda^* \rightarrow \lambda$  é real.

Daí decorre que  $x$  terá componentes reais, porque será calculado a partir de um sistema linear de equações com coeficientes reais.

2. Os auto vetores de uma matriz simétrica, associados a auto valores diferentes são ortogonais.

Prova:

Sejam dois auto valores diferentes  $\lambda_i$  e  $\lambda_j$  associados aos auto vetores  $x_i$  e  $x_j$ . Então:

$$Ax_i = \lambda_i x_i \quad (5)$$

$$Ax_j = \lambda_j x_j \quad (6)$$

Multiplicando (5) por  $x_j^t$  e (6) por  $x_i^t$  temos:

$$x_j^t Ax_i = \lambda_i x_j^t x_i$$

$$x_i^t Ax_j = \lambda_j x_i^t x_j$$

Subtraindo:

$$x_j^t Ax_i - x_i^t Ax_j = \lambda_i x_j^t x_i - \lambda_j x_i^t x_j$$

Como  $x_j^t Ax_i = x_i^t Ax_j$  porque  $A$  é simétrica, temos:

$$0 = \lambda_i x_j^t x_i - \lambda_j x_i^t x_j$$

$$\text{Mas } x_j^t x_i = \sum_{k=1}^n x_{jk} x_{ik} = \sum_{k=1}^n x_{ik} x_{jk} = x_i^t x_j \dots$$

Então:

$$0 = (\lambda_i - \lambda_j) x_i^t x_j$$

Como  $\lambda_i \neq \lambda_j \rightarrow (\lambda_i - \lambda_j) \neq 0$ , o que acarreta que

$x_i^t x_j = 0 \rightarrow x_i$  e  $x_j$  são ortogonais.

REFERÊNCIAS

- | 1| CARROL, J. B. - "An Analytical Solution for Approximating Simple Structure in Factor Analysis. Psychometrika, 18 pg. 23-28, 1953.
  
- | 2| CARROL, J. B. - "Biquatimin Criterion for Rotation to Oblique Simple Structure in Factor Analysis". Science, 126 pg. 1114-1115, 1957.
  
- | 3| CATTEL, R. B. - "The Scree Test for the Number of Factor". Multivariate Behavioural Research, 1 pg. 245-276, 1966.
  
- | 4| DWYER, P. S. - "The Contribution of an Orthogonal Multiple Factor Solution to Multiple Correlation". Psychometrika, 4 pg. 163-171, 1939.
  
- | 5| FERGUSON, G. A. - "The Concept of Parsimony in Factor Analysis". Psychometrika, 19 pg. 281-290, 1954.
  
- | 6| FULKS, W. - "Advanced Calculus - An Introduction to Analysis". John Wiley and Sons, 1961.
  
- | 7| GUTTMAN, L. - "Image Theory for the Structure of Quantitative Variates". Psychometrika, 18 pg. 277-296, 1953.



- | 8| GUTTMAN, L. - "Some Necessary Conditions for Common Factor Analysis". Psychometrika, 19 pg. 149-161, 1954.
- | 9| HADLEY, G. - "Linear Algebra" - Addison-Wesley, 1961.
- |10| HARMAN, H. H. - "Modern Factor Analysis". University of Chicago Press, 1975.
- |11| HARMAN, H. H. e WAYNE, H. JONES - "Factor Analysis by Minimizing Residuals (Minres)". Psychometrika, 31 pg. 351-368, 1966.
- |12| HARMAN, H. H. e YOICHIRO FUKUDA - "Resolution of the Heywood Case in Minres Solution". Psychometrika, 31 pg. 563-571, 1966.
- |13| HARRIS, C. W. - "Some Ras-Guttman Relationships". Psychometrika, 27 pg. 247-263, 1962.
- |14| HERIG, L. C. T. - "O Modelo da Análise das Componentes Principais". Tese de Mestrado, Eng. Sistemas - COPPE/UFRJ, 1979.
- |15| HOLZINGER, K. J. e SWINEFORD, F. - "A Study in Factor Analysis: The Stability of a Bi-Factor Solution". Supplementary Educational Monographs, 48. University of Chicago, Dept. of Education, 1937.

- [16] KAISER, H. F. - "The Varimax Criterion for Analytic Rotation in Factor Analysis". Psychometrika, 23 pg. 187-200, 1958.
- [17] KAISER, H. F. - "Computer Program for Varimax Rotation in Factor Analysis". Educational and Psychological Measurement, 19 pg. 413-420, 1959.
- [18] KAISER, H. F. - "The Application of Electronic Computers to Factor Analysis". Educational and Psychological Measurement, 20 pg. 141-151, 1960.
- [19] KAISER, H. F. e CAFFREY - "Alpha Factor Analysis". Psychometrika, 30 pg. 1-14, 1965.
- [20] LUENBERGER, D. G. - "Introduction to Linear and Nonlinear Programming". Addison-Wesley, 1973.
- [21] McDONALD, R. P. - "Numerical Methods for Polynomial Models in Nonlinear Factor Analysis". Psychometrika, 32 1967.
- [22] McDONALD, R. P. - "Factor Interaction in Nonlinear Factor Analysis". British Journal of Mathematical and Statistical Psychology, 20, 1967.
- [23] MORRINSON, D. F. - "Multivariate Statistical Methods". McGraw-Hill, 1967.

- |24| NEUHAUS, J. D. e WRIGLEY, C. - "The Quartimax Method: An Analytical Approach to Orthogonal Simple Structure" . British Journal of Mathematical and Statistical Psychology, 7 pg. 81-91, 1954.
- |25| OVERALL e KLETT - "Applied Multivariate Analysis". McGraw-Hill, 1972.
- |26| PINZKA, C. e SAUNDERS, D. R. - "Analytic Rotation to Simple Structure, II: Extension to an Oblique Solution". Research Bulletin RB-54-36, Princeton, N. J. Educational Testing Service, 1954.
- |27| RIPPE, D. D. - "Application of a Large Sampling Criterion to Some Sampling Problems in Factor Analysis". Psychometrika, 18 pg. 191-205, 1953.
- |28| SAUNDERS, D. R. - "A Computer Program to Find the Best Fitting Orthogonal Factor for a Given Hypothesis. Psychometrika, 25 pg. 199-205, 1960.
- |29| SPEARMAN, C. - "General Intelligence, Objectively Determined and Measured". American Journal of Psychology, 15 pg. 201-293, 1904.
- |30| THOMSON, G. H. - "Hotelling's Method Modified to Give Spearman's G." - Journal of Educational Psychology, 25 pg. 366-374, 1934.

- [31] THURSTONE, L. L. - "Multivariate Factor Analysis". Psychological Review, 38 pg. 406-427, 1931.
- [32] THURSTONE, L. L. - "The Vector of Mind". University of Chicago Press, 1935.
- [33] THURSTONE, L. L. - "Multiple Factor Analysis". University of Chicago Press, 1947.

BIBLIOGRAFIA ADICIONAL

- AFIFI, A. A. e AZEN, S. P. - "Statistical Analysis - A Computer Oriented Approach". Academic Press, 1974.
- ANDERSON, T. W. - "Introduction to Multivariate Statistical Analysis". John Wiley and Sons, 1958.
- BENNET e BOWERS - "Multivariate Techniques for Social and Behavioural Sciences". MacMillan, 1976.
- BRYANT, E. H. e ATCHLEY, W. R. (Ed.) - "Multivariate Statistical Methods Within Groups Covariation - Benchmark Papers in Systematic and Evolution Biology", Vol. 2. Dowden, Hutchinson and Ross, 1975.
- COOLEY, W. W. e LOHNES, P. R. - "Multivariate Data Analysis" John Wiley and Sons, 1971.
- HARRIS, R. J. - "Primer of Multivariate Statistics". Academic Press, 1975.
- HORST, P. - "Factor Analysis of Data Matrices". Holt, Rinehart and Winston, 1965.
- KENDALL, M. - "Multivariate Analysis". Charles Griffin, 1975.

- LAWLEY, D. N. e MAXWELL, A. E. - "Factor Analysis as a Statistical Method". London Butterworths, 1971.
- LEFEBVRE, J. - "Introduction Aux Analysis Statistique Multidimensionnelles". Masson, 1976.
- MARRIOT, F. H. C. - "The Interpretation of Multiple Observations". Academic Press, 1974.
- PRESS, S. J. - "Applied Multivariate Analysis". Holt, Rinehart and Winston, 1972.
- RUMMELL, R. J. - "Applied Factor Analysis". Evanston Northwestern University Press, 1970.
- TORRENS, J. IBERN. - "Modele et Methods de L'Analyse Factorielle". Dunod, 1972.