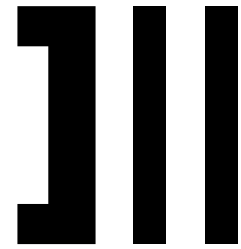


9

Análise de fatores



9.1. Introdução

A técnica dos componentes principais consiste em uma transformação ortogonal dos eixos coordenados do sistema multivariado buscando as orientações de maior variabilidade. Para o estudo de dependências estruturais multinormais, as técnicas de explicação das covariâncias das respostas são preferidas. Apesar das técnicas dos componentes principais poder ser usada para essa finalidade, esta não deve ser preferida por ser apenas uma transformação e não um resultado de um modelo fundamental da estrutura de covariância. Esse método possui alguns inconvenientes, tais como não ser invariante quanto às mudanças de escalas e não possuir um critério adequado para determinar quando uma proporção suficiente da variação total foi explicada pelos componentes retidos.

Nesse capítulo apresenta-se a técnica de análise de fatores com o propósito essencial de descrever, se possível, as relações de covariância entre diversas variáveis em função de poucas, não observáveis, quantidades aleatórias denominadas de fatores. Sob o modelo de fatores cada variável resposta é

representada por uma função linear de uma pequena quantidade de fatores comuns, não observáveis, e de uma simples variável latente específica. Os fatores comuns geram as covariâncias entre as variáveis observadas e os termos específicos contribuem somente para as variâncias de suas respostas relacionadas. Os coeficientes dos fatores comuns não são restritos a condição de ortogonalidade, o que confere generalidade, apesar de se exigir normalidade dos dados e a determinação, a priori, do número de fatores.

Nesse capítulo serão apresentados o modelo de fatores ortogonais, os métodos de estimação dos parâmetros desse modelo e brevemente o problema de rotação dos fatores. É abordado método de estimação que não exigem normalidade. Métodos de estimação dos escores dos fatores são abordados, o que ao contrário dos componentes principais não é uma tarefa simples.

9.2. Modelo de fatores ortogonais

Supondo que o sistema multivariado consiste de p resposta descritas pelas p variáveis observáveis aleatórias X_1, X_2, \dots, X_p . Assumindo que o vetor de observações multivariadas ${}_p\tilde{X}_1$ possui média $\underline{\mu}$ e covariância Σ , então, o modelo de fatores pressupõe que o vetor ${}_p\tilde{X}_1$ é linearmente dependente de algumas poucas variáveis não observáveis F_1, F_2, \dots, F_m chamadas de fatores comuns, e p fontes de variações adicionais $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ chamadas de erro ou de fatores específicos. O modelo de fatores pode ser especificado por:

elementos são independentemente distribuídos, ou seja, $\text{Cov}(\underline{\varepsilon}) = \Psi$ ($p \times p$). Sendo assim, definem-se:

$$E(\underline{F}) = \underline{0} \quad (9.3)$$

$$\text{Cov}(\underline{F}) = E(\underline{F}\underline{F}') = I \quad (9.4)$$

$$E(\underline{\varepsilon}) = \underline{0} \quad (9.5)$$

$$\text{Cov}(\underline{\varepsilon}) = E(\underline{\varepsilon}\underline{\varepsilon}') = \Psi = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix} \quad (9.6)$$

Finalmente, é assumido que \underline{F} e $\underline{\varepsilon}$ são independentes, portanto,

$$\text{Cov}(\underline{\varepsilon}, \underline{F}) = E(\underline{\varepsilon}\underline{F}') = \underset{(p \times m)}{0} \quad (9.7)$$

O modelo (9.2) e essas pressuposições definem o modelo de fatores ortogonal. Dessa forma a estrutura de covariância de \underline{X} pode ser dada por:

$$\text{Cov}(\underline{X}) = \Sigma = E(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})^t$$

Substituindo $\tilde{X} - \tilde{\mu}$ pelas definições dadas no modelo (9.2), verifica-se

que:

$$\begin{aligned} (\tilde{X} - \tilde{\mu})(\tilde{X} - \tilde{\mu})' &= (L\tilde{F} + \tilde{\varepsilon})(L\tilde{F} + \tilde{\varepsilon})' = (L\tilde{F} + \tilde{\varepsilon})((L\tilde{F})' + \tilde{\varepsilon}') = \\ &= L\tilde{F}(L\tilde{F})' + \tilde{\varepsilon}(L\tilde{F})' + L\tilde{F}\tilde{\varepsilon}' + \tilde{\varepsilon}\tilde{\varepsilon}' \end{aligned}$$

Então,

$$\begin{aligned} \text{Cov}(\tilde{X}) = \Sigma &= E(\tilde{X} - \tilde{\mu})(\tilde{X} - \tilde{\mu})' = \\ &= E[L\tilde{F}(L\tilde{F})' + \tilde{\varepsilon}(L\tilde{F})' + L\tilde{F}\tilde{\varepsilon}' + \tilde{\varepsilon}\tilde{\varepsilon}'] \\ &= LE(\tilde{F}\tilde{F}')L' + E(\tilde{\varepsilon}\tilde{F}')L' + LE(\tilde{F}\tilde{\varepsilon}') + E(\tilde{\varepsilon}\tilde{\varepsilon}') \end{aligned}$$

De acordo com as condições (9.4), (9.7), (9.6), tem-se:

$$\text{Cov}(\tilde{X}) = \Sigma = LL' + \Psi \quad (9.8)$$

Também podem ser obtidas as covariâncias entre os componentes de \tilde{X} e \tilde{F} a partir das suposições assumidas e apresentadas anteriormente. Assim,

$$\begin{aligned} \text{Cov}(\tilde{X}, \tilde{F}) &= E[(\tilde{X} - \tilde{\mu})\tilde{F}'] = E[(L\tilde{F} + \tilde{\varepsilon})\tilde{F}'] = E(L\tilde{F}\tilde{F}' + \tilde{\varepsilon}\tilde{F}') = \\ &= E(L\tilde{F}\tilde{F}') + E(\tilde{\varepsilon}\tilde{F}') = LE(\tilde{F}\tilde{F}') + E(\tilde{\varepsilon}\tilde{F}') = LI + 0 = L \end{aligned}$$

Logo,

$$\text{Cov}(\underline{X}, \underline{F}) = L \quad \text{ou} \quad \text{Cov}(X_i, F_j) = \ell_{ij} \quad (9.9)$$

Da relação (9.8) verifica-se que:

$$\text{Var}(X_i) = \sigma_{ii} = \sum_{j=1}^m \ell_{ij}^2 + \psi_i = \ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2 + \psi_i \quad (9.10)$$

$$\text{Cov}(X_i, X_k) = \sigma_{ik} = \sum_{j=1}^m \ell_{ij} \ell_{kj} = \ell_{i1} \ell_{k1} + \ell_{i2} \ell_{k2} + \dots + \ell_{im} \ell_{km}$$

A porção da i -ésima variável explicada por m fatores comuns é chamada de comunalidade e a porção de σ_{ii} devida aos fatores específicos é denominada de variância específica. Denotando a i -ésima comunalidade por h_i^2 é fácil observar de (9.10) que:

$$h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2 \quad (9.11)$$

Assim,

$$\sigma_{ii} = h_i^2 + \psi_i \quad i = 1, 2, \dots, p \quad (9.12)$$

Quando $m=p$ a matriz Σ pode ser reproduzida exatamente por LL' , de tal forma que $\Psi=0$. A utilidade da análise de fatores, no entanto, ocorre quando m é bem menor do que p . Dessa forma, o número de parâmetros na análise de fatores, $p(m+1)$,

é bem menor do que aqueles $p(p+1)/2$ parâmetros de Σ . Por exemplo, para $p=20$ existem $20 \times 21/2 = 210$ parâmetros em Σ . Se $m=2$ fatores são utilizados, então, o modelo de fatores possui $p(m+1) = 20(2+1) = 60$ parâmetros (ℓ_{ij} e ψ_i).

O grande problema da análise de fatores é a dificuldade ou a impossibilidade de fatorar a matriz Σ em $LL'+\Psi$, quando m é bem menor do que p . Algumas vezes, quando são obtidas soluções, estas são, em geral, inconsistentes com as interpretações estatísticas. A análise de fatores tem como propósito a determinação dos elementos da matriz de cargas fatoriais L e dos elementos de Ψ , obedecendo a restrição (9.12).

Quando $m > 1$ existem mais de uma solução para o modelo de fatores, todas consistentes com as interpretações estatísticas. Essa ambigüidade é a base para uma importante característica da análise de fatores que é a rotação fatorial. Para demonstrar essa propriedade, seja T uma matriz ortogonal $m \times m$, ou seja, $TT' = T'T = I$. A expressão (9.2) pode ser reescrita por:

$$\underline{X} - \underline{\mu} = L\underline{F} + \underline{\varepsilon} = LTT'\underline{F} + \underline{\varepsilon} = L^*\underline{F}^* + \underline{\varepsilon} \quad (9.13)$$

em que: $L^* = LT$ e $\underline{F}^* = T'\underline{F}$.

Como $E(\underline{F}^*) = T'E(\underline{F}) = T'\underline{0} = \underline{0}$ e $\text{Cov}(\underline{F}^*) = T'\text{Cov}(\underline{F})T = T'IT = T'T = I$, então, é impossível distinguir as cargas de L das de L^* , ou seja, os fatores \underline{F} e $\underline{F}^* = T'\underline{F}$ possuem as mesmas propriedades, mesmo que as cargas fatoriais de L e de L^* sejam, em geral, diferentes, uma vez que geram a mesma matriz de covariância Σ . Assim,

$$\Sigma = LL' + \Psi = LTT'L' + \Psi = (L^*)(L^*)' + \Psi \quad (9.14)$$

A escolha da matriz T é direcionada por um critério de facilitação da interpretação dos fatores gerados, uma vez que as propriedades estatísticas não são alteradas.

9.3. Estimação das cargas fatoriais

Nas situações reais, os parâmetros do modelo de fatores são desconhecidos e devem ser estimados das observações amostrais. A análise de fatores é justificável quando Σ difere de uma matriz diagonal, ou quando matriz ρ de correlações difere da identidade. Para uma amostra X_1, X_2, \dots, X_n de tamanho n em p variáveis correlacionadas a matriz S é um estimador de Σ , bem como R é de ρ . Com base em uma estimativa de Σ é possível realizar o teste de hipótese de igualdade de Σ a uma matriz diagonal, conforme descrição realizada no capítulo 7. Se a hipótese não for rejeitada, os fatores específicos possuem papel dominante, sendo que a análise de fatores é determinar alguns poucos fatores comuns. Nesse caso, a análise de fatores não terá grande utilidade.

Se a hipótese de a estrutura de Σ ser igual a uma matriz diagonal for rejeitada, então, o modelo de fatores será útil e o problema inicial será o de estimar as

cargas fatoriais ℓ_{ij} e as variâncias específicas ψ_i . Nessa seção são considerados dois métodos de estimação para os parâmetros do modelo de fatores: o método dos componentes principais e o método da máxima verossimilhança apresentado por Lawley (1940, 1942 e 1943). Qualquer que seja o método aplicado, as soluções podem sofrer rotações com a finalidade de simplificar as interpretações dos fatores. É prudente, também, tentar mais de uma solução.

9.3.1. Método dos componentes principais

A decomposição espectral vista nos capítulos 2 e 7, representa um importante método de fatoração de Σ . Sejam as matrizes $P = [\underline{e}_1 \ \underline{e}_2 \ \dots \ \underline{e}_p]$ e $\Lambda = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$, compostas dos autovetores e autovalores de Σ , com $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, então:

$$\Sigma = P\Lambda P' = P\Lambda^{1/2}\Lambda^{1/2}P' = LL' \quad (9.15)$$

em que, $L = P\Lambda^{1/2}$ é uma matriz $p \times p$ de cargas fatoriais.

A equação (9.15) reflete um ajuste da estrutura de covariância por um modelo de fatores tendo tantos fatores quanto variáveis ($m=p$) e variâncias específicas ψ_i nulas para todo $i=1, 2, \dots, p$. Nesse modelo as cargas fatoriais do j -ésimo fator representam os coeficientes do j -ésimo componente principal (autovetor) multiplicado pelo fator de escala $\sqrt{\lambda_j}$. Embora a relação (9.15) seja exata, esta não é útil por utilizar

tantos fatores quanto variáveis e por não deixar variação alguma para os fatores específicos.

Uma solução para o problema é considerar um número m , de fatores comuns, menor do que o de variáveis p . Com esse critério $p-m$ autovalores e os respectivos autovetores são desconsiderados. Esses autovalores são àqueles $(p-m)$ menores. Dessa forma a contribuição de $\lambda_{m+1}z_{m+1}z_{m+1}^t + \lambda_{m+2}z_{m+2}z_{m+2}^t + \dots + \lambda_p z_p z_p^t$ para Σ é negligenciada. Desprezando essa contribuição, a seguinte aproximação de Σ pode ser obtida:

$$\Sigma \cong \begin{bmatrix} \sqrt{\lambda_1} e_1 & & & \\ & \sqrt{\lambda_2} e_2 & & \\ & & \dots & \\ & & & \sqrt{\lambda_p} e_p \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} e_1 \\ \sqrt{\lambda_2} e_2 \\ \vdots \\ \sqrt{\lambda_p} e_p \end{bmatrix} = LL' \quad (9.16)$$

em que L é uma matriz $p \times m$. A representação (9.16), no entanto, não considera a contribuição dos fatores específicos. A contribuição desses fatores pode ser estimada tomando-se a diagonal de $\Sigma - LL'$, sendo LL' definida em (9.16).

Dessa forma a matriz Σ pode ser aproximada por:

$$\Sigma \cong LL' + \Psi \quad (9.17)$$

$$\Psi = \text{Diag}(\Sigma - LL') \text{ ou } \psi_i = \sigma_{ii} - \sum_{j=1}^m \ell_{ij}^2 \text{ para } i=1, 2, \dots, p.$$

É comum trabalhar com a representação das variáveis em uma escala padronizada. Nessa situação a variável Z_i possui média 0 e variância 1. A padronização pode ser realizada por:

$$\tilde{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix} = V^{-1/2} (\tilde{X} - \tilde{\mu}) = \begin{bmatrix} \frac{X_1 - \mu_1}{\sqrt{\sigma_{11}}} \\ \vdots \\ \frac{X_p - \mu_p}{\sqrt{\sigma_{pp}}} \end{bmatrix} \quad (9.18)$$

em que:

$$V^{-1/2} = \begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{\sigma_{pp}}} \end{bmatrix}$$

A matriz de covariância de \tilde{Z} é dada por ρ . O processo de obtenção dos parâmetros do modelo de fatores é o mesmo descrito nas equações de (9.17), considerando $\Sigma = \rho$ e $L = P\Lambda^{1/2}$, sendo P a matriz $p \times m$ com as colunas compostas pelos m primeiros autovetores de ρ e $\Lambda^{1/2}$ uma matriz $m \times m$ com diagonal igual a $\sqrt{\lambda_i}$. Como $\sigma_{ii} = 1$, é fácil perceber que $\psi_i = 1 - \sum_{j=1}^m \ell_{ij}^2$. A padronização evita que uma variável com elevada variação influencie indevidamente a determinação das cargas fatoriais.

A representação apresentada em (9.17), quando Σ ou ρ são substituídos pelos seus estimadores S ou R , é conhecida como solução dos componentes principais para a análise de fatores. O nome se origina do fato de os fatores serem derivados dos primeiros componentes principais amostrais. O resumo dos principais resultados desse método de estimação é doravante apresentado.

A análise de fatores por componentes principais obtidos da covariância amostral S é especificada em função dos pares de autovalores e autovetores $(\hat{\lambda}_i, \hat{e}_i)$, $i=1, 2, \dots, p$, em que $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$. Seja $m < p$, o número de fatores comuns. A matriz das cargas fatoriais estimadas $(\hat{\ell}_{ij})$ é dada por:

$$\tilde{L} = \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{e}_1 & \sqrt{\hat{\lambda}_2} \hat{e}_2 & \dots & \sqrt{\hat{\lambda}_m} \hat{e}_m \end{bmatrix} = \hat{P}_1 \hat{\Lambda}_1^{1/2} \quad (9.19)$$

em que \hat{P}_1 é uma matriz $p \times m$ dos autovetores amostrais de S e $\hat{\Lambda}_1$ é uma matriz diagonal $m \times m$ dos autovalores amostrais de S .

Os estimadores das variâncias específicas são dados pela matriz diagonal resultante da seguinte operação matricial.

$$\tilde{\Psi} = \begin{bmatrix} \tilde{\psi}_1 & 0 & \dots & 0 \\ 0 & \tilde{\psi}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \tilde{\psi}_p \end{bmatrix} = \text{Diag}(S - \tilde{L}\tilde{L}') \quad (9.20)$$

De (9.20) verifica-se que:

$$\tilde{\Psi}_i = S_{ii} - \sum_{j=1}^m \tilde{\ell}_{ij}^2 = S_{ii} - \tilde{h}_i^2 \quad (9.21)$$

Sendo que o estimador da comunalidade é dado por:

$$\tilde{h}_i^2 = \tilde{\ell}_{i1}^2 + \tilde{\ell}_{i2}^2 + \dots + \tilde{\ell}_{im}^2 \quad (9.22)$$

A análise de fatores por componentes principais da matriz R, por sua vez, é obtida substituindo S por R nas equações de (9.19) a (9.22). Na solução dos componentes principais as estimativas das cargas fatoriais não se alteram com o aumento do número m de fatores.

É fácil perceber por meio das definições apresentadas que a matriz S não é fielmente reproduzida pela solução de componentes principais. A diagonal de S é exatamente reproduzida pelo modelo de fatores, mas os elementos fora da diagonal principal não são. Assim,

$$S \cong \tilde{L}\tilde{L}' + \tilde{\Psi} \quad (9.23)$$

Se o número de fatores não é especificado por considerações a priori, como por teoria ou por trabalhos anteriores de outros pesquisadores, a escolha de m para uma decomposição de maior acurácia de S pode ser baseada nos autovalores estimados, da mesma forma que o número de componentes principais a serem retidos é determinado. Analiticamente, Johnson e Wichern (1998) demonstram que a soma de

quadrados dos elementos da matriz de resíduos $S - \tilde{L}\tilde{L}' - \tilde{\Psi}$ é menor ou igual a $\sum_{i=m+1}^p \hat{\lambda}_i^2$.

Assim, um pequeno valor da soma de quadrados dos últimos $(p-m)$ autovalores negligenciados implica em uma pequena soma de quadrados do erro da aproximação realizada por m componentes. O ideal é obter uma elevada contribuição dos primeiros fatores para a variação total amostral. Assim, verifica-se que:

$$\sum_{i=1}^m \tilde{\ell}_{ij}^2 = \tilde{\ell}_{1j}^2 + \tilde{\ell}_{2j}^2 + \dots + \tilde{\ell}_{pj}^2 = \sqrt{\hat{\lambda}_i} \hat{e}_i \sqrt{\hat{\lambda}_i} \hat{e}_i^t = \hat{\lambda}_i \quad (9.24)$$

Logo, a porcentagem da variação total devida ao j -ésimo fator é dada por:

$$\% \text{VarExp} = \begin{cases} \frac{\hat{\lambda}_j}{\text{Tr}(S)} \times 100 & \text{para fatores de S} \\ \frac{\hat{\lambda}_j}{p} \times 100 & \text{para fatores de R} \end{cases} \quad (9.25)$$

O critério (9.25) é usado como um artifício heurístico para determinar o valor apropriado de m . O número de fatores comuns retidos deve aumentar até que um fração “adequada” da variação amostral tenha sido contemplada.

Exemplo 9.1. Em 24 tartarugas fêmeas foram mensuradas $p=3$ variáveis X_1 , X_2 e X_3 , quais sejam, comprimento, largura e altura de carapaças transformadas por logaritmo.

A matriz de covariâncias amostrais é apresentada a seguir. Obter a análise de fatores com $m=1$ e $m=2$ usando o método dos componentes principais.

$$S = \begin{bmatrix} 4,9810 & 3,8063 & 4,7740 \\ 3,8063 & 3,0680 & 3,7183 \\ 4,7740 & 3,7183 & 4,8264 \end{bmatrix}$$

Inicialmente foi testada a hipótese:

$$H_0 : \Sigma = \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{pp} \end{bmatrix}; \quad \sigma_{ii} > 0$$

O valor de qui-quadrado obtido foi de $\chi_c^2 = 127,9805$ com $v=2$ graus de liberdade. Como $\Pr(\chi^2 > 127,9805) = 0,0024$ rejeita-se H_0 de independência entre todas as variáveis. Portanto, a análise de fatores deve ser eficiente.

A solução para $m=1$ é apresentada a seguir. A solução de 1 fator explica 98,2% da variação total e pode ser julgada satisfatória. A soma de quadrados dos dois últimos autovalores, dada por $\hat{\lambda}_2^2 + \hat{\lambda}_3^2 = 0,0291$, foi considerada muito pequena e indica que a soma de quadrados dos elementos da matriz de resíduos não deve ultrapassar esse valor. Os resultados obtidos são:

Variáveis	Cargas fatoriais		Comunalidades	Variâncias específicas
	F ₁		\tilde{h}_i^2	$\tilde{\psi}_i$
X ₁	2,2165		4,9129	0,0681
X ₂	1,7277		2,9849	0,0831
X ₃	2,1770		4,7394	0,0870
% explicação	98,1500			

A matriz de resíduos é dada por:

$$S - \tilde{L}\tilde{L}' - \tilde{\Psi} =$$

$$= \begin{bmatrix} 4,9810 & 3,8063 & 4,7740 \\ 3,8063 & 3,0680 & 3,7183 \\ 4,7740 & 3,7183 & 4,8264 \end{bmatrix} - \begin{bmatrix} 2,2165 \\ 1,7277 \\ 2,1770 \end{bmatrix} \begin{bmatrix} 2,2165 & 1,7277 & 2,1770 \end{bmatrix} - \begin{bmatrix} 0,0681 & 0 & 0 \\ 0 & 0,0831 & 0 \\ 0 & 0 & 0,0870 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & -0,0228 & -0,0515 \\ -0,0228 & 0 & -0,0429 \\ -0,0515 & -0,0429 & 0 \end{bmatrix}$$

A soma de quadrados dos elementos dessa matriz de resíduos é de apenas 0,01003, que é menor do que 0,0291 conforme já era esperado.

Para m=2 a solução é dada por:

Variáveis	Cargas fatoriais		Comunalidades	Variâncias específicas
	F ₁	F ₁	\tilde{h}_i^2	$\tilde{\psi}_i$
X ₁	2,2165	0,1630	4,9394	0,0418
X ₂	1,7277	0,1608	3,0108	0,0575
X ₃	2,1770	-0,2935	4,8255	0,0003
% explicação acumulada	98,15	99,23		

A soma de quadrados de resíduos para esse caso (m=2) é igual a 0,0049, a qual é limitada por 0,0099. Uma vez que os ganhos foram muito pequenos, o modelo

de 1 fator pode ser julgado adequado. O fator 1 pode ser interpretado como um fator de volume.

Uma aproximação modificada do método dos componentes principais é denominada solução fatorial principal. O procedimento é válido tanto para R quanto para S . A descrição que é realizada a seguir utiliza a matriz R . No modelo de fatores $\rho = LL' + \Psi$ é perfeitamente especificado. Os m fatores comuns reconstituem perfeitamente os elementos fora da diagonal principal de ρ , bem como os elementos da diagonal com a participação da variância específica: $1 = h_i^2 + \psi_i$.

Supondo que a contribuição dos fatores específicos seja removida da reconstituição de ρ , então, a matriz resultante é $\rho - \Psi = LL'$. Suponha, também, que estimativas iniciais ψ_i^* tenham sido obtidas por um meio qualquer, então, é possível definir a matriz de correlação amostral reduzida (R_r) eliminando o efeito dos fatores específicos por $R_r = R - \Psi^*$. Esse processo é equivalente a substituir a diagonal de R por $h_i^{*2} = 1 - \psi_i^*$. A matriz R_r é definida por:

$$R_r = R - \Psi^* = \begin{bmatrix} h_1^{*2} & r_{12} & \cdots & r_{1p} \\ r_{21} & h_2^{*2} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & h_p^{*2} \end{bmatrix} \quad (9.26)$$

Teoricamente, desconsiderando a variação amostral, é possível estabelecer que a matriz R_r pode ser recomposta pelos m fatores comuns. Dessa forma, R_r é fatorada em:

$$R_r \cong L^* L_r^{*t} \quad (9.27)$$

em que L_r^* é a matriz dos estimadores das cargas fatoriais ℓ_{ij}^* .

O método fatorial principal de análise de fatores utiliza os estimadores:

$$\begin{cases} L_r^* = \begin{bmatrix} \sqrt{\hat{\lambda}_1^*} \hat{\xi}_1^* & \sqrt{\hat{\lambda}_2^*} \hat{\xi}_2^* & \dots & \sqrt{\hat{\lambda}_m^*} \hat{\xi}_m^* \end{bmatrix} \\ \psi_i^* = 1 - \sum_{j=1}^m \ell_{ij}^{*2} \end{cases} \quad (9.28)$$

em que $(\hat{\lambda}_i^*; \hat{\xi}_i^*)$, $i=1, 2, \dots, m$ são os (maiores) pares de autovalor-autovetor obtidos de R_r .

As communalidades devem ser re-estimadas por:

$$\tilde{h}_i^{*2} = \sum_{j=1}^m \ell_{ij}^{*2} \quad (9.29)$$

O método, então, é aplicado iterativamente, considerando as communalidades estimadas em (9.29) para recalcular a matriz R_r em (9.26). Os autovalores e autovetores dessa nova matriz R_r são obtidos e as estimativas das cargas fatoriais e variâncias específicas utilizando (9.28) são novamente obtidas. Novas

comunalidades, também, são obtidas utilizando (9.29) e o processo é repetido em novos estágios sucessivos, até que não haja alterações nas estimativas das cargas fatoriais e das variâncias específicas para uma dada precisão.

Um problema que pode surgir nesse procedimento é o aparecimento de autovalores de R_r negativos. Recomenda-se utilizar o número de fatores comuns igual ao posto da matriz reduzida (R_r). Uma das causas dos autovalores negativos é devida aos valores iniciais das variâncias específicas utilizadas. Algumas alternativas existem para a escolha desses valores iniciais. A mais popular é utilizar $\psi_i^* = 1/r^{ii}$, em que r^{ii} é o elemento da i -ésima diagonal da matriz R^{-1} . As communalidades iniciais são, então, dadas por:

$$h_i^{*2} = 1 - \psi_i^* = 1 - \frac{1}{r^{ii}} \quad (9.30)$$

que é igual ao coeficiente de determinação parcial múltiplo entre a i -ésima variável (X_i) e as $(p-1)$ demais variáveis. Essa relação é útil, pois permite que h_i^{*2} seja obtida pelo coeficiente de determinação múltiplo, mesmo quando R não tiver posto completo. Usando S , a variância específica inicial é função de S^{ii} , o elemento da i -ésima posição da diagonal de S^{-1} , da seguinte forma:

$$h_i^{*2} = S_{ii} - \left(1 - \frac{m}{2p}\right) S^{ii} \quad (9.31)$$

9.3.2. Método da máxima verossimilhança

Se os fatores comuns \underline{F} e os fatores específicos $\underline{\varepsilon}$ possuem distribuição normal, estimativas de máxima verossimilhança podem ser obtidas. Do modelo de fatores e da consideração de que as variáveis \underline{F} e $\underline{\varepsilon}$ possuem distribuição normal pode concluir que $\underline{X}_j - \underline{\mu} = \underline{L}\underline{F}_j + \underline{\varepsilon}_j$ também é normalmente distribuído e portanto a função de verossimilhança é:

$$\begin{aligned}
 L(\underline{\mu}, \Sigma) &= (2\pi)^{-np/2} |\Sigma|^{-n/2} \\
 &\times \exp \left\{ -\left(\frac{1}{2}\right) \text{tr} \left[\Sigma^{-1} \left(\sum_{j=1}^n (\underline{X}_j - \bar{\underline{X}})(\underline{X}_j - \bar{\underline{X}})^t + n(\bar{\underline{X}} - \underline{\mu})(\bar{\underline{X}} - \underline{\mu})^t \right) \right] \right\} = \\
 &= (2\pi)^{-(n-1)p/2} |\Sigma|^{-(n-1)/2} \exp \left\{ -\left(\frac{1}{2}\right) \text{tr} [\Sigma^{-1} \underline{S}_n] \right\} \\
 &\times (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp \left\{ -\left(\frac{n}{2}\right) \text{tr} \left[(\bar{\underline{X}} - \underline{\mu})^t \Sigma^{-1} (\bar{\underline{X}} - \underline{\mu}) \right] \right\}
 \end{aligned} \tag{9.32}$$

a qual depende de \underline{L} e Ψ por meio de $\Sigma = \underline{L}\underline{L}^t + \Psi$.

Devido à multiplicidade de escolhas para \underline{L} dadas por transformações ortogonais é imperativo impor uma restrição de unicidade computacional por:

$$\underline{L}^t \Psi^{-1} \underline{L} = \Delta \quad \text{uma matriz diagonal} \tag{9.33}$$

Os estimadores de máxima verossimilhança $\hat{\underline{L}}$ e $\hat{\Psi}$ devem ser obtidos por maximização numérica de (9.32). A maximização de (9.32) sujeita a condição de unicidade (9.33) deve satisfazer:

$$(\hat{\Psi}^{-1/2} S_n \hat{\Psi}^{-1/2})(\hat{\Psi}^{-1/2} \hat{L}) = \hat{\Psi}^{-1/2} \hat{L} (I + \hat{\Delta}) \quad (9.34)$$

Lawley (1940, 1942, 1943) mostra que o estimador $\hat{\Delta}$ é dado por:

$$\hat{\Delta} = \hat{L}^t \hat{\Psi}^{-1} \hat{L} \quad (9.35)$$

Assim, a equação (9.34) pode ser rescrita de outra forma, procedendo as seguintes operações:

$$(\hat{\Psi}^{-1/2} S_n \hat{\Psi}^{-1/2}) \hat{\Psi}^{-1/2} \hat{L} = \hat{\Psi}^{-1/2} \hat{L} (I + \hat{L}^t \hat{\Psi}^{-1} \hat{L})$$

$$(\hat{\Psi}^{-1/2} S_n \hat{\Psi}^{-1/2}) \hat{\Psi}^{-1/2} \hat{L} - \hat{\Psi}^{-1/2} \hat{L} = \hat{\Psi}^{-1/2} \hat{L} \hat{L}^t \hat{\Psi}^{-1} \hat{L}$$

$$(\hat{\Psi}^{-1/2} S_n \hat{\Psi}^{-1/2} - I) \hat{\Psi}^{-1/2} \hat{L} = \hat{\Psi}^{-1/2} \hat{L} \hat{L}^t \hat{\Psi}^{-1} \hat{L}$$

Logo,

$$\left[\hat{\Psi}^{-1/2} (S_n - \hat{\Psi}) \hat{\Psi}^{-1/2} \right] \hat{\Psi}^{-1/2} \hat{L} = \hat{\Psi}^{-1/2} \hat{L} \hat{L}^t \hat{\Psi}^{-1} \hat{L} \quad (9.36)$$

Como $\hat{L}^t \hat{\Psi}^{-1} \hat{L}$ é uma matriz diagonal para garantir que os elementos de \hat{L} sejam únicos, então, as autovalores de $(S_n - \hat{\Psi}) \hat{\Psi}^{-1}$, e portanto $\hat{\Psi}^{-1/2} (S_n - \hat{\Psi}) \hat{\Psi}^{-1/2}$, são iguais aos valores correspondentes a diagonal de $\hat{\Delta}$. Dessa forma, a i -ésima coluna de

$\hat{\Psi}^{-1/2}\hat{\mathbf{L}}$ é o vetor característico correspondente ao i -ésimo autovalor de $\hat{\Psi}^{-1/2}(S_n - \hat{\Psi})\hat{\Psi}^{-1/2}$. O cálculo desses vetores não é um processo direto, uma vez que os elementos de $\hat{\Psi}$ são também desconhecidos, os quais devem ser obtidos da relação $\hat{\Psi} = \text{Diag}(S - \hat{\mathbf{L}}\hat{\mathbf{L}}^t)$. Sendo assim, o processo de estimação deve ser executado iterativamente estimando-se os vetores característicos correspondentes a valores iniciais de os elementos de $\hat{\Psi}$, e então, utilizá-los para obter novas estimativas mais precisas das variâncias específicas sucessivamente.

Para o modelo com m fatores os vetores característicos correspondentes aos m maiores autovalores de S_n podem ser utilizados como valores iniciais do processo iterativo. Os elementos desses vetores devem ser re-escalados para que as somas de seus quadrados sejam iguais aos respectivos autovalores. O processo iterativo é descrito a seguir:

1. Calcular as m raízes características $(\hat{\lambda}_{10}, \hat{\lambda}_{20}, \dots, \hat{\lambda}_{m0})$ de S_n e os vetores característicos correspondentes $(\hat{\xi}_{10}, \hat{\xi}_{20}, \dots, \hat{\xi}_{m0})$, de tal sorte que seus elementos sejam re-escalados para que tenham norma quadrática igual a $\hat{\lambda}_{i0}$, $i=1, 2, \dots, m$. Seja a matriz \hat{Q}_0 ($p \times m$) definida por $\hat{Q}_0 = [\hat{\xi}_{10} \ \hat{\xi}_{20} \ \dots \ \hat{\xi}_{m0}]$. Dessa forma, é possível definir as matrizes $\hat{\Lambda}_0$ ($m \times m$) e \hat{P}_0 ($p \times m$) por:

$$\hat{\Lambda}_0 = \begin{bmatrix} \hat{\lambda}_{10} & 0 & \cdots & 0 \\ 0 & \hat{\lambda}_{20} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\lambda}_{m0} \end{bmatrix}$$

e

$$\hat{P}_0 = \hat{Q}_0 \hat{\Lambda}_0^{1/2}$$

2. Aproximar as variâncias específicas por:

$$\hat{\Psi}_0 = \text{Diag}(\mathbf{S}_n - \hat{P}_0 \hat{P}_0^t) \quad (9.37)$$

3. Obter a matriz

$$\hat{\Psi}_0^{-1/2} (\mathbf{S}_n - \hat{\Psi}_0) \hat{\Psi}_0^{-1/2} \quad (9.38)$$

e extrair os m autovetores $(\hat{e}_{11}, \hat{e}_{21}, \dots, \hat{e}_{m1})$ e os correspondentes autovalores

$(\hat{\lambda}_{11}, \hat{\lambda}_{21}, \dots, \hat{\lambda}_{m1})$ dessa matriz. Formar a matriz $\hat{Q}_1 = [\hat{e}_{11} \ \hat{e}_{21} \ \dots \ \hat{e}_{m1}]$ e definir as

matrizes:

$$\hat{\Lambda}_1 = \begin{bmatrix} \hat{\lambda}_{11} & 0 & \cdots & 0 \\ 0 & \hat{\lambda}_{21} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\lambda}_{m1} \end{bmatrix}$$

e

$$\hat{P}_1 = \hat{Q}_1 \hat{\Lambda}_1^{1/2}$$

A primeira aproximação de L (\hat{L}_1) é dada por:

$$\hat{L}_1 = \hat{\Psi}_0^{1/2} \hat{P}_1 \quad (9.39)$$

4. Calcular

$$\hat{\Psi}_1 = \text{Diag}(S_n - \hat{L}_1 \hat{L}_1^t) \quad (9.40)$$

Repetir os passos 3 e 4 até que os correspondentes elementos de sucessivas iterações de \hat{L}_i e \hat{L}_{i+1} não difiram por um valor superior a uma quantidade pré-determinada (critério de convergência). O resultado final do processo iterativo conterà as estimativas de máxima verossimilhança para as cargas fatoriais L e das variâncias específicas para o modelo m-fatorial. É apresentado a seguir um programa SAS no procedimento de matrizes IML para a obtenção de estimativas de máxima verossimilhança do modelo m-fatorial.


```

options ps=5000 ls=80 nodate nonumber;;
proc iml;
  S={4.9810 3.8063 4.7740,
      3.8063 3.0680 3.7183,
      4.7740 3.7183 4.8264};
  p=ncol(S);n=24;alpha=0.05;
  L0=Diag(eigval(S));P0=eigvec(S);
  numfac=1;numIt=100;
  L0=L0[1:numfac,1:numfac];
  P0=P0[1:p,1:numfac];P0=P0*root(L0);
  print L0 P0; Psi0=diag(S-P0*P0`);
  print psi0;
  psii=psi0;
  do i=1 to numIt;
    Print '_____';
    print 'iteracao ' i;
    Print '_____';
    Delta=inv(root(psii))*(S-psii)*inv(root(psii));
    *print delta;
    Li=Diag(eigval(delta));Pi=eigvec(delta);
    Li=Li[1:numfac,1:numfac]; Pi=Pi[1:p,1:numfac];
    Pi=root(psii)*Pi*root(Li);
    *print Li Pi;
    Psii=diag(S-Pi*Pi`);
    /*soma de quadrados dos residuos do modelo*/
    resi=S-pi*pi`-psii;
    print 'Soma de quadrados dos residuos';
    SQResiduo=sum(resi#resi);
    print sqresiduo;
    *print psii;
    Print '_____';
  end;
  Print 'Solucao final do modelo de fatores';
  Print 'Cargas fatoriais';
  print Pi;
  print 'Variancias especificas';
  print psii; resi=S-pi*pi`-psii;
  print 'matriz de residuos';
  print resi;
  print 'Soma de quadrados dos residuos';
  SQResiduo=sum(resi#resi);
  print sqresiduo;
  print 'Cargas fatoriais de Z-variaveis padronizadas';
  D=root(inv(diag(S))); PiZ=D*Pi;
  print PiZ;
  print 'Variancias especificas fatoriais de Z-variaveis padronizadas';
  PsiZ=D*psii*D;
  print PsiZ;
  Li=Diag(eigval(delta));
  print Li;
quit;

```

As cargas fatoriais e as variâncias específicas da matriz R podem ser obtidas diretamente de \hat{L} e $\hat{\Psi}$ realizando as seguintes transformações. Formar a matriz diagonal (D) a partir dos elementos S_{ii} de S. Então obter as estimativas de máxima verossimilhança de R para as cargas fatoriais (\hat{L}_Z) e para as variâncias específicas $\hat{\Psi}_Z$. Esses estimadores são:

$$\hat{L}_Z = D^{-1/2} \hat{L} \quad (9.41)$$

$$\hat{\Psi}_Z = D^{-1/2} \hat{\Psi} D^{-1/2} \quad (9.42)$$

As estimativas de máxima verossimilhança das comunalidades são dadas por:

$$\hat{h}_i^2 = \hat{\ell}_{i1}^2 + \hat{\ell}_{i2}^2 + \dots + \hat{\ell}_{im}^2 \quad \text{para } i = 1, 2, \dots, p \quad (9.43)$$

Dessa forma, a proporção explicada pelo j-ésimo fator é dada por:

$$\% \text{VarExp} = \begin{cases} \frac{\sum_{i=1}^p \hat{\ell}_{ij}^2}{\text{Tr}(S)} \times 100 & \text{para fatores de S} \\ \frac{\sum_{i=1}^p \hat{\ell}_{Z(ij)}^2}{p} \times 100 & \text{para fatores de R} \end{cases} \quad (9.44)$$

O processo descrito anteriormente para a obtenção das soluções de máxima verossimilhança possui convergência lenta. Aitken (1937) propôs uma técnica conhecida por processo δ^2 de aceleração dos esquemas iterativos de convergência. Seja $\tilde{\ell}_{jt}$ os elementos do t-ésimo processo iterativo, referente a j-ésima coluna da matriz de cargas fatoriais L_t do estágio t. O processo de Aitken (1937) prevê para 3 consecutivos valores de $\tilde{\ell}_{jt}$ o ajuste pela razão:

$$\ell_{ijt}^2 = \frac{\begin{vmatrix} \tilde{\ell}_{ij(t-1)} & \tilde{\ell}_{ijt} \\ \tilde{\ell}_{ijt} & \tilde{\ell}_{ij(t+1)} \end{vmatrix}}{\tilde{\ell}_{ij(t+1)} - 2\tilde{\ell}_{ijt} + \tilde{\ell}_{ij(t-1)}} \quad (9.45)$$

em que $\tilde{\ell}_{ijt}$ é o i-ésimo elemento de $\tilde{\ell}_{jt}$. Se o denominador de (9.45) for nulo o valor de ℓ_{ijt}^2 deve ser feito igual a $\tilde{\ell}_{ijt}$.

Aitken (1937) mostra que os termos de ℓ_{ijt}^2 convergem mais rapidamente do que àqueles de $\tilde{\ell}_{jt}$.

Exemplo 9.2. Utilizando a matriz de covariâncias amostral das 24 tartarugas fêmeas que foram mensuradas em $p=3$ variáveis X_1 , X_2 e X_3 , as quais são: comprimento, largura e altura de carapaças transformadas por logaritmo, determinar o modelo de fatores com $m=1$. Ajustar o modelo por meio de estimativas de máximas verossimilhanças.

$$S = \begin{bmatrix} 4,9810 & 3,8063 & 4,7740 \\ 3,8063 & 3,0680 & 3,7183 \\ 4,7740 & 3,7183 & 4,8264 \end{bmatrix}$$

- i) Inicialmente foram obtidos os autovalores e autovetores de S e compostas as matrizes $\hat{\Lambda}_0(1 \times 1)$, $\hat{Q}_0(3 \times 1)$ e $\hat{P}_0(3 \times 1)$ por:

$$\hat{\Lambda}_0 = 12,637147 \quad \hat{Q}_0 = \begin{bmatrix} 0,6234937 \\ 0,4859812 \\ 0,612436 \end{bmatrix} \quad \hat{L}_0 = \hat{P}_0 = \hat{Q}_0 \hat{\Lambda}_0^{1/2} = \begin{bmatrix} 2,2164432 \\ 1,727603 \\ 2,1771344 \end{bmatrix}$$

- ii) As variâncias específicas iniciais foram obtidas por:

$$\hat{\Psi}_0 = \text{Diag}(S_n - \hat{P}_0 \hat{P}_0^t) = \begin{bmatrix} 0,0683794 & 0 & 0 \\ 0 & 0,0833879 & 0 \\ 0 & 0 & 0,0864857 \end{bmatrix}$$

- iii) Foi obtida a seguinte matriz e desta extraídos os autovalores e autovetores. O m=1 primeiro autovalor e autovetor correspondente foram usados para compor as matrizes $\hat{\Lambda}_1(1 \times 1)$, $\hat{Q}_1(3 \times 1)$ e $\hat{P}_1(3 \times 1)$.

$$\hat{\Psi}_0^{-1/2} (S_n - \hat{\Psi}_0) \hat{\Psi}_0^{-1/2} = \begin{bmatrix} 71,843527 & 50,406739 & 62,079406 \\ 50,406739 & 35,791891 & 43,784534 \\ 62,079406 & 43,784534 & 54,805777 \end{bmatrix}$$

$$\hat{\Lambda}_1 = 161,45963 \quad \hat{Q}_1 = \begin{bmatrix} 0,6657947 \\ 0,4691915 \\ 0,5801523 \end{bmatrix} \quad \hat{P}_1 = \hat{Q}_1 \hat{\Lambda}_1^{1/2} = \begin{bmatrix} 8,4600381 \\ 5,9618652 \\ 7,3718074 \end{bmatrix}$$

Finalmente a primeira aproximação \hat{L}_1 é feita por:

$$\hat{L}_1 = \hat{\Psi}_0^{1/2} \hat{P}_1 = \begin{bmatrix} 2,2122546 \\ 1,721606 \\ 2,167934 \end{bmatrix}$$

iv) Foi calculado o segundo valor $\hat{\Psi}_1$ por:

$$\hat{\Psi}_1 = \text{Diag}(S_n - \hat{L}_0 \hat{L}_0^t) = \begin{bmatrix} 0,0869296 & 0 & 0 \\ 0 & 0,1040727 & 0 \\ 0 & 0 & 0,1264622 \end{bmatrix}$$

Os procedimentos 3 e 4 foram repetidos 41 vezes até que as trocas na matriz (vetor) \hat{L} fosse da ordem de $1e-7$ ou menos. O resultado final foi:

$$\hat{L}_{41} = \hat{\Psi}_{40}^{1/2} \hat{P}_{41} = \begin{bmatrix} 2,2106526 \\ 1,7217993 \\ 2,1595433 \end{bmatrix} e$$

$$\hat{\Psi}_{41} = \text{Diag}(S_n - \hat{L}_{41} \hat{L}_{41}^t) = \begin{bmatrix} 0,0940152 & 0 & 0 \\ 0 & 0,1034073 & 0 \\ 0 & 0 & 0,1627727 \end{bmatrix}$$

A matriz de resíduos (R) foi:

$$R = \begin{bmatrix} 0 & 2,9835E-8 & 3,7474E-8 \\ 2,9835E-8 & 0 & -7,05E-8 \\ 3,7474E-8 & -7,05E-8 & 0 \end{bmatrix}$$

E a soma de quadrados dos resíduos foi:

SQResíduos= 1,453E-14

As cargas fatoriais obtidas das variáveis padronizadas são:

$$\hat{L}_Z = D^{-1/2} \hat{L} = \begin{bmatrix} \frac{1}{\sqrt{4,9810}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{3,0680}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{4,8264}} \end{bmatrix} \begin{bmatrix} 2,2106526 \\ 1,7217993 \\ 2,1595433 \end{bmatrix} = \begin{bmatrix} 0,9905177 \\ 0,983003 \\ 0,9829926 \end{bmatrix}$$

E as variâncias específicas são:

$$\hat{\Psi}_Z = D^{-1/2} \hat{\Psi} D^{-1/2} = \begin{bmatrix} 0,0188748 & 0 & 0 \\ 0 & 0,0337051 & 0 \\ 0 & 0 & 0,0337255 \end{bmatrix}$$

Exemplo 9.3. A matriz de correlação entre 10 escores das respectivas 10 provas do declato, medidas em n=160 atletas, está apresentada a seguir. Obter os m=4 fatores pelo método da máxima verossimilhança. As dez variáveis mensuradas são: i) corrida de 100 m rasos; ii) salto em distância; iii) lançamento de peso; iv) salto em altura; v) corrida dos 400m livres; vi) 110 m com barreiras; vii) arremesso de disco; viii) salto com vara; ix) arremesso de dardos; e x) corrida de 1500 m. A matriz de correlação dos escores dos 160 competições.

$$R = \begin{bmatrix} 1,00 & 0,59 & 0,35 & 0,34 & 0,63 & 0,40 & 0,28 & 0,20 & 0,11 & -0,07 \\ & 1,00 & 0,42 & 0,51 & 0,49 & 0,52 & 0,31 & 0,36 & 0,21 & 0,09 \\ & & 1,00 & 0,38 & 0,19 & 0,36 & 0,73 & 0,24 & 0,44 & -0,08 \\ & & & 1,00 & 0,29 & 0,46 & 0,27 & 0,39 & 0,17 & 0,18 \\ & & & & 1,00 & 0,34 & 0,17 & 0,23 & 0,13 & 0,39 \\ & & & & & 1,00 & 0,32 & 0,33 & 0,18 & 0,00 \\ & & & & & & 1,00 & 0,24 & 0,34 & -0,02 \\ & & & & & & & 1,00 & 0,24 & 0,17 \\ & & & & & & & & 1,00 & -0,00 \\ & & & & & & & & & 1,00 \end{bmatrix}$$

A solução de $m=4$ fatores obtida por Johnson e Wichern (1998) foi re-estimada pelo algoritmo apresentado nesse trabalho para obter as estimativas de máxima verossimilhança. Após 100 mil iterações o algoritmo convergiu.

Variáveis	Estimativas de máxima verossimilhança				Variâncias específicas
	Cargas fatoriais estimadas				$\hat{\psi}_i = 1 - \hat{h}_i^2$
	F_1	F_2	F_3	F_4	
Corrida 100m	-0,0869	0,3449	0,8290	-0,1685	0,157935
Salto em distância	0,0688	0,4352	0,5931	0,2746	0,378693
Lançamento de peso	-0,1294	0,9911	-0,0038	-0,0007	0,001053
Salto em altura	0,1603	0,4059	0,3343	0,4451	0,499688
corrida 400m	0,3787	0,2437	0,6702	-0,1372	0,329262
110m com barreira	-0,0178	0,3629	0,4234	0,3878	0,538310
Arremesso de disco	-0,0563	0,7294	0,0268	0,0182	0,463815
Salto com vara	0,1573	0,2640	0,2275	0,3937	0,698795
Arremesso de dardos	-0,0218	0,4411	-0,0115	0,0971	0,795340
1500m rasos	0,9986	0,0496	-0,0004	-0,0001	0,000408
Proporção cumulativa da variância explicada	0,12	0,37	0,55	0,61	

9.4. Rotação fatorial

A fatoração de Σ em $LL'+\Psi$ não é única, conforme discussão realizada na seção 9.2. A pós-multiplicação da matriz de cargas fatoriais L por qualquer matriz ortogonal conformável (T) conduz a uma fatoração igualmente válida. A solução numérica de Rao-Maxwell para as equações de verossimilhança remove essa indeterminação por adotar a restrição de que $\hat{L}'\hat{\Psi}^{-1}\hat{L}$ seja uma matriz diagonal. Não obstante, após a obtenção da solução de máxima verossimilhança, qualquer transformação ortogonal pode ser realizada. A idéia é aplicar tal transformação rígida dos eixos coordenados, a qual conduz a um padrão que tornam as cargas fatoriais mais facilmente interpretáveis. Essa rotação rígida dos eixos coordenados das m -dimensões fatoriais é chamada de rotação das cargas fatoriais.

Citado por Morrison (1974) Thurstone sugere um critério de resposta de simples estrutura para a realização da rotação fatorial. Estruturas como a sugerida raramente existe em dados reais e não será descrito o procedimento de Thurstone. Outra técnica de uso limitado é a obtenção de rotação graficamente dos fatores plotados dois a dois. A rotação analítica é o procedimento mais comumente empregado. Na rotação ortogonal rígida as propriedades estatísticas dos fatores ficam inalteradas, embora a matriz de cargas fatoriais não seja a mesma. Supondo que a matriz $p \times m$ de cargas fatoriais seja submetida a uma rotação rígida pela matriz ortogonal T ($m \times m$) por meio da seguinte operação: $L^* = LT$. A ortogonalidade de T , isto é, $TT'=T'T=I$, faz com que as comunalidade fiquem inalteradas:

$$\sum_{j=1}^m \ell_{ij}^{*2} = \sum_{j=1}^m \ell_{ij}^2 \Leftrightarrow h_i^{*2} = h_i^2$$

bem como a soma de seus quadrados:

$$\sum_{i=1}^p \left(\sum_{j=1}^m \ell_{ij}^2 \right)^2 = \sum_{i=1}^p \sum_{j=1}^m \ell_{ij}^4 + 2 \sum_{i=1}^p \sum_{j=1}^{m-1} \sum_{k=j+1}^m \ell_{ij}^2 \ell_{ik}^2 \quad (9.46)$$

também é invariante.

Com esse resultado em evidência é possível especificar critérios de simplicidade ou parcimônia propostos pelos analistas de fatores (Morrison, 1976). Fergusson (1954) sugeriu minimizar o termo dos duplos produtos de (9.46) como uma medida de parcimônia, por meio de uma escolha adequada de T. Esse resultado foi determinado quase que ao mesmo tempo e independentemente por Carroll (1953).

Neuhaus e Wrigley (1954) propuseram a maximização da variância do quadrado das pm cargas fatoriais para definir T. A variância do quadrado das cargas fatoriais é:

$$V = \sum_{i=1}^p \sum_{j=1}^m \ell_{ij}^4 - \frac{1}{pm} \left(\sum_{i=1}^p \sum_{j=1}^m \ell_{ij}^2 \right)^2 \quad (9.47)$$

Como o termo de correção é meramente soma das comunalidades tomada ao quadrado, então, a maximização de V é equivalente a maximizar a soma da quarta potência das cargas fatoriais, ou equivalentemente, minimizar a medida de parcimônia de Fergusson (1954) e Carroll (1953). Por argumentos diferentes Sanders (1960) obteve o mesmo critério de Neuhaus e Wrigley (1954). Esse critério determina o

método denominado de “quartimax” por maximizar a soma da quarta potencia das cargas fatoriais.

Kaiser (1958, 1959) propôs uma medida de estrutura simples relacionada a soma das variâncias das cargas fatoriais quadráticas dentro de cada coluna da matriz L de fatores. O critério de “varimax” de linha de Kaiser é:

$$v^* = \frac{1}{p^2} \sum_{j=1}^m \left[p \sum_{i=1}^p \ell_{ij}^4 - \left(\sum_{i=1}^p \ell_{ij}^2 \right)^2 \right] \quad (9.48)$$

Esse critério dá pesos iguais às respostas com grandes e com pequenas comunalidades e Kaiser sugere a melhora desse critério pelo uso do critério alternativo:

$$v = \frac{1}{p^2} \sum_{j=1}^m \left[p \sum_{i=1}^p x_{ij}^4 - \left(\sum_{i=1}^p x_{ij}^2 \right)^2 \right] \quad (9.49)$$

em que:

$$x_{ij} = \frac{\ell_{ij}}{\sqrt{\sum_{j=1}^m \ell_{ij}^2}} \quad (9.50)$$

é j-ésima carga fatorial do i-ésima variável resposta dividida pela raiz quadrada de sua comunalidade. Na seqüência da rotação os valores de x_{ij} devem ser multiplicados pela

raiz quadrada de sua comunalidade respectiva para restaurar a dimensão original. Esse critério foi nomeado por Kaiser de “varimax”.

O processo computacional para a rotação varimax é descrito a seguir. Considere o par de fatores r e s , com cargas normalizadas x_{ir} e x_{is} . A rotação desses fatores envolve o simples ângulo ϕ , e diferenciando (9.49) com relação a ϕ Kaiser mostrou que o ângulo deve satisfazer a relação:

$$\operatorname{tg}(\phi) = \frac{2 \left[2p \sum_{i=1}^p (x_{ir}^2 - x_{is}^2) x_{ir} x_{is} - \sum_{i=1}^p (x_{ir}^2 - x_{is}^2) \left(2 \sum_{i=1}^p x_{ir} x_{is} \right) \right]}{p \sum_{i=1}^p \left[(x_{ir}^2 - x_{is}^2) - (2x_{ir} x_{is})^2 \right] - \left\{ \left[\sum_{i=1}^p (x_{ir}^2 - x_{is}^2) \right]^2 - 2 \left(\sum_{i=1}^p x_{ir} x_{is} \right)^2 \right\}} \quad (9.51)$$

Para que a segunda derivada seja negativa é necessário que 4ϕ seja colocado no quadrante correto. A escolha é designada pelos sinais do numerador e denominador de (9.51). A Tabela 9.1 especifica o quadrante de 4ϕ em função destes sinais.

A solução iterativa para a rotação é realizada de acordo com os seguintes procedimentos: a rotação do primeiro e segundo fator é realizada como ângulo ϕ determinado conforme descrição anterior; o novo primeiro fator é rotado com o terceiro fator original, e assim por diante, até que $m(m-1)/2$ pares de rotações tenham sido executadas. Essa seqüência de rotações é repetida até que “todos” os ângulos sejam menores que um critério de convergência especificado ε , dentro de um ciclo.

Tabela 9.1. Quadrante do ângulo 4ϕ em função dos sinais do numerador e denominador da equação (9.51).

Sinal do denominador	Sinal do numerador	
	+ (positivo)	- (negativo)
+ (positivo)	I: $0^0 \leq 4\phi < 90^0$	IV: $-90^0 \leq 4\phi < 0^0$
- (negativo)	II: $90^0 \leq 4\phi < 180^0$	III: $-180^0 \leq 4\phi < -90^0$

Exemplo 9.4. Efetuar a rotação varimax dos $m=3$ fatores obtidos por Morrison (1974) apresentados a seguir.

(incompleto)

9.5. Teste da falta de ajuste do modelo de fatores

A natureza das estimativas de máxima verossimilhança das cargas fatoriais conduz a um teste formal para o m -ésimo modelo fatorial. A hipótese nula é:

$$\begin{cases} H_0 : \Sigma = LL^t + \Psi \\ H_1 : \Sigma \text{ uma matriz } p \times p \text{ p.d. sim.} \end{cases} \quad (9.52)$$

Usando a distribuição de Wishart, Morrison (1976) mostra que a razão de verossimilhança fornece o seguinte teste, com a correção de Bartlett (1954):

$$\chi_c^2 = \left[n - 1 - \frac{(2p + 4m + 5)}{6} \right] \ln \frac{|\hat{L}\hat{L}^t + \hat{\Psi}|}{|S_n|} \quad (9.53)$$

o qual tem distribuição qui-quadrado para grandes amostras com:

$$v = \frac{1}{2} \left[(p - m)^2 - p - m \right] \quad (9.54)$$

graus de liberdade.

Pela propriedade da invariância das cargas e das variâncias específicas estimadas segue-se que o valor do teste seria o mesmo da solução de fatores da matriz de correlação R. Para a aplicação do teste da falta de ajuste é necessário que os graus de liberdade sejam positivos. Isso significa que o número de fatores comuns m não pode exceder o maior inteiro que satisfaz a equação:

$$m < \frac{1}{2} (2p + 1 - \sqrt{8p + 1}) \quad (9.55)$$

O teste de razão de verossimilhança compara as variâncias generalizadas $|\hat{L}\hat{L}^t + \hat{\Psi}|$ e $|S_n|$. Se m for pequeno em relação a p, geralmente H_0 é rejeitada, conduzindo a um modelo com um maior número de fatores comuns. Por outro lado, quando m for grande em relação a hipótese tende a ser não rejeitada, principalmente para grandes valores de n. Isso acontece devido ao fato de $\hat{L}\hat{L}^t + \hat{\Psi}$ aproximar de S_n , de tal sorte que o acréscimo de novos fatores não traga novas melhoras ao modelo. A diminuição de m pode, ainda, pelas mesmas razões levar a não rejeição de H_0 . Algum tipo de bom senso deve ser aplicado na escolha de m.

Para demonstrar que a padronização das variáveis não afeta o teste apresentado seja $D^{-1/2}$ definida anteriormente a matriz diagonal com o recíproco dos desvios padrões das p variáveis na diagonal principal. Então, a razão que aparece na equação (9.53) pode ser operada por:

$$\frac{|\hat{L}\hat{L}^t + \hat{\Psi}|}{|S_n|} = \frac{|D^{-1/2}||\hat{L}\hat{L}^t + \hat{\Psi}||D^{-1/2}|}{|D^{-1/2}||S_n||D^{-1/2}|}$$

uma vez que a multiplicação do numerador e denominador não altera o resultado final.

Pela propriedade do determinante $|AB|=|A|\times|B|$, verifica-se que:

$$\frac{|\hat{L}\hat{L}^t + \hat{\Psi}|}{|S_n|} = \frac{|D^{-1/2}\hat{L}\hat{L}^tD^{-1/2} + D^{-1/2}\hat{\Psi}D^{-1/2}|}{|D^{-1/2}S_nD^{-1/2}|} = \frac{|\hat{L}_z\hat{L}_z^t + \hat{\Psi}_z|}{|R|}$$

Dessa forma o teste de qui-quadrado é exatamente o mesmo, quando for aplicado a partir da matriz S_n ou da matriz R , com os dados padronizados.

9.6. Escores fatoriais

Os fatores são variáveis não observáveis, muito embora seus valores possam ser estimados. Os valores estimados dos fatores são denominados de escores. Dois métodos de estimação são propostos. Ambos tratam as cargas fatoriais e as variâncias específicas estimadas como se fossem os verdadeiros valores desconhecidos. Se ocorrer rotação, os escores são obtidos a partir das cargas fatoriais que sofreram rotação e não a partir das originais. Não obstante, as fórmulas não distinguirão entre as situações em que ocorreu rotação daquelas em não ocorreu, uma vez que estas fórmulas não são alteradas pelas rotações.

9.6.1. Método dos mínimos quadrados ponderados

Suponha que $\underline{\mu}$, L e Ψ sejam conhecidos para o modelo fatorial:

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

Como $\text{Var}(\varepsilon_i) = \psi_i$, não necessariamente igual para todo i , Bartlett (1937) sugeriu o uso dos quadrados mínimos ponderados, usando como peso o recíproco das variâncias específicas. A soma de quadrados de resíduos do modelo fatorial ponderada é dada por:

$$\sum_{i=1}^p \frac{\varepsilon_i^2}{\Psi_i} = \underline{\varepsilon}' \Psi^{-1} \underline{\varepsilon} = (\underline{X} - \underline{\mu} - L\underline{F})' \Psi^{-1} (\underline{X} - \underline{\mu} - L\underline{F}) \quad (9.56)$$

Bartlett (1937) propôs a solução $\hat{\underline{F}}$ que minimiza (9.56). A solução é:

$$\hat{\underline{F}} = (L' \Psi^{-1} L)^{-1} L' \Psi^{-1} (\underline{X} - \underline{\mu}) \quad (9.57)$$

Como L , Ψ e $\underline{\mu}$ são desconhecidos, os respectivos estimadores são utilizados para a obtenção dos escores fatoriais:

$$\hat{\underline{F}}_j = (\hat{L}' \hat{\Psi}^{-1} \hat{L})^{-1} \hat{L}' \hat{\Psi}^{-1} (\underline{X}_j - \bar{\underline{X}}) \quad j=1, 2, \dots, n \quad (9.58)$$

Se a matriz de correlação for utilizada, então:

$$\hat{\underline{F}}_j = (\hat{L}'_z \hat{\Psi}_z^{-1} \hat{L}_z)^{-1} \hat{L}'_z \hat{\Psi}_z^{-1} \underline{Z}_j \quad j=1, 2, \dots, n \quad (9.59)$$

Se as cargas fatoriais que sofreram rotação são usadas $\hat{L}^* = \hat{L}T$, então,

$\hat{\underline{F}}_j$ se relaciona com $\hat{\underline{F}}_j^*$ por:

$$\hat{F}_j^* = T' \hat{F}_j \quad (9.60)$$

9.6.2. Método de regressão

A partir do modelo de fatores originais:

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

Considerando que L e Ψ são conhecidas, e que \underline{F} e $\underline{\varepsilon}$ possuem distribuição normal multivariada com média e variâncias dadas pelas equações de (9.3) a (9.6), a combinação linear $\underline{X} - \underline{\mu} = L\underline{F} + \underline{\varepsilon}$ tem distribuição $N_p(\underline{0}, LL' + \Psi)$. A distribuição conjunta de $\underline{X} - \underline{\mu}$ e \underline{F} é, também, $N_{m+p}(\underline{0}, \Sigma^*)$; em que:

$$\Sigma^* = \begin{bmatrix} \Sigma = LL' + \Psi & L \\ L' & I \end{bmatrix} \quad (9.61)$$

A média $\underline{0}$ é um vetor $[(m+p) \times 1]$ de zeros. A distribuição condicional de $\underline{F} / \underline{x}$ é normal com média e variância dados por:

$$E(\underline{F} / \underline{x}) = L' \Sigma^{-1} (\underline{x} - \underline{\mu}) = L' (LL' + \Psi)^{-1} (\underline{x} - \underline{\mu}) \quad (9.62)$$

e

$$\text{Cov}(\tilde{F}/\tilde{x}) = I - L'\Sigma^{-1}L = I - L'(LL' + \Psi)^{-1}L \quad (9.63)$$

Os coeficientes $L'(LL' + \Psi)^{-1}$ são os coeficientes de uma regressão multivariada dos fatores com as variáveis originais. As estimativas desses coeficientes produzem os escores fatoriais. Dados as observações \tilde{X}_j e tomando-se os estimadores de máxima verossimilhança \hat{L} e $\hat{\Psi}$ os escores dos fatores são dados por:

$$\hat{F}_j = \hat{L}'(\hat{L}\hat{L}' + \hat{\Psi})^{-1}(\tilde{X}_j - \bar{\tilde{X}}) \quad j=1, 2, \dots, n \quad (9.64)$$

O uso da identidade de matrizes:

$$\hat{L}'(\hat{L}\hat{L}' + \hat{\Psi})^{-1} = (\mathbf{I} + \hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1}\hat{L}'\hat{\Psi}^{-1} \quad (9.65)$$

pode simplificar o cálculo dos escores dos fatores, os quais são dados por:

$$\hat{F}_j = (\mathbf{I} + \hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1}\hat{L}'\hat{\Psi}^{-1}(\tilde{X}_j - \bar{\tilde{X}}) \quad j=1, 2, \dots, n \quad (9.66)$$

A comparação dos escores fatoriais obtidos por regressão (LS) e por mínimos quadrados ponderados (WLS) pode ser realizada subtraindo os estimadores (9.66) e (9.58). Assim, simbolizando os estimadores de regressão por \hat{F}_j^{LS} e o de mínimos quadrados ponderados por \hat{F}_j^{WLS} , tem-se:

$$\begin{aligned}\hat{\xi}_j^{LS} - \hat{\xi}_j^{WLS} &= (\hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1} \hat{L}'\hat{\Psi}^{-1} (\underline{X}_j - \bar{X}) - (I + \hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1} \hat{L}'\hat{\Psi}^{-1} (\underline{X}_j - \bar{X}) = \\ &= \left[(\hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1} - (I + \hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1} \right] \hat{L}'\hat{\Psi}^{-1} (\underline{X}_j - \bar{X})\end{aligned}$$

Pelas estimativas de máxima verossimilhança verifica-se que $(\hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1}$ é uma matriz diagonal, se o seu valor for próximo de zero a diferença anterior será próxima de zero, ou seja, os estimadores anteriores fornecerão aproximadamente os mesmos escores.

9.7. Exercícios

- 9.7.1. Teste a hipótese de que o modelo com $m=1$ fator, apresentado no exemplo 9.1, é adequado utilizando o teste de qui-quadrado para falta de ajuste do modelo.
- 9.7.2. Para o exemplo 9.3 testar a aderência do modelo com $m=4$ fatores.
- 9.7.3. Obter estimativas de máxima verossimilhança para $m=1$ e $m=2$ dos dados apresentados no exemplo 7.6.7 e calcular os escores pelos dois métodos apresentados. Para o caso de $m=2$ fatores plotar os escores dos dois fatores obtidos.

9.8. Referências

BARTLETT, M.S. A note on multiplying factors for various Chi-Square approximations.

Journal of the royal Statistical Society –Series B. v.16, p.296-298, 1954.

BARTLETT, M.S. The statistical conception of mental factors. **British Journal of**

Psychology. v.28, p.97-104, 1937.

CARROLL, J.B. An analytical solution for approximating simple structure in factor

analysis. **Psychometrika.** v.18, pp.23-28, 1953.

FERGUSON, G.A. The concept of parsimony in factor analysis. **Psychometrika.** v.19,

pp.281-290, 1954.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis.** 4th

edition. Prentice Hall, New Jersey, 1998. 816p.

KAISER, H.F. The varimax criterion for analytic rotation in factor analysis.

Psychometrika. v.23, pp.187-200, 1958.

KAISER, H.F. Computer program for varimax rotation in factor analysis. **Journal of**

Educational and Psychological Measurement. v.19, pp.413-420, 1959.

LAWLEY, D.N. The estimation of factor loadings by the method of the maximum

likelihood. **Proceedings of the Royal Society of Edinburgh, Series A.** v.60,

pp.64-82, 1940.

LAWLEY, D.N. Further estimation in factor analysis. **Proceedings of the Royal Society**

of Edinburgh, Series A. v.61, pp.176-185, 1942.

LAWLEY, D.N. The application of the maximum likelihood method to factor analysis.

British Journal of Psychology. v.33, pp.172-175, 1943.

MORRISON, D.F. **Multivariate statistical methods.** New York: McGraw-Hill, 2d ed.,

1976. 307p.

NEHAUS, J.; WRIGLEY, C. The quartimax method: an analytical approach to orthogonal simple structure. **British Journal of Psychology**. v.7, pp.81-91, 1954.