A decorative graphic consisting of three vertical bars on the left, a large blue number '7' in the center, and three vertical bars on the right. Below the number '7' is the text 'Componentes principais' in green.

7

Componentes principais

7.1. Introdução

A análise de componentes principais está relacionada com a explicação da estrutura de covariância por meio de poucas combinações lineares das variáveis originais em estudo. Os objetivos dessa análise são: i) redução da dimensão original; e ii) facilitação da interpretação das análises realizadas. Em geral, a explicação de toda a variabilidade do sistema determinado por p variáveis só pode ser efetuada por p componentes principais. No entanto, uma grande parte dessa variabilidade pode ser explicada por um número k menor de componentes, $k \leq p$. Os componentes principais são uma técnica de análise intermediária e, portanto não se constituem em um método final e conclusivo. Esse tipo de análise se presta fundamentalmente como um passo intermediário em grandes investigações científicas.

Essa técnica pode ser aplicada, ainda, na análise de regressão múltipla, principalmente, nos casos de colinearidade ou de multicolinearidade; aplica-se também à análise de agrupamento e como estimadores de fatores nas técnicas multivariadas denominadas de análises fatoriais. Muitas outras

aplicações de componentes principais são encontradas nas literaturas aplicadas. A técnica AMMI (additive multiplicative interaction model) considera modelos lineares com interação entre dois fatores e aplica como base para seus procedimentos a análise de componentes principais.

7.2. Componentes principais populacionais

Algebricamente os componentes principais representam combinações lineares de p variáveis aleatórias X_1, X_2, \dots, X_p . Geometricamente, essas combinações lineares representam a seleção de novos eixos coordenados, os quais são obtidos por rotações do sistema de eixos original, representados por X_1, X_2, \dots, X_p . Os novos eixos representam as direções de máxima variabilidade.

Como pode ser demonstrado, os componentes principais dependem somente da matriz de covariância Σ (ou da matriz de correlação ρ) e de X_1, X_2, \dots, X_p . Seu desenvolvimento não requer pressuposições de normalidade multivariada, mas possuem interpretações úteis em termos da constante elipsóide de densidade, se a normalidade existir. A princípio, serão definidos os conceitos de componentes principais populacionais. Posteriormente, naturalmente esses conceitos serão estendidos para a situação amostral.

Seja o vetor aleatório $\underline{X}' = [X_1 \ X_2 \ \dots \ X_p]$ amostrado de uma população com covariância Σ , cujos autovalores são $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, então, os componentes principais (Y_1, Y_2, \dots, Y_p) são as combinações lineares dadas por

(7.1)

$$(\Sigma - \lambda_i I)\underline{e}_i = 0 \quad (7.5)$$

É fácil perceber que dessa equação surge a seguinte e óbvia relação, obtida no ponto máximo, dada por: $\Sigma \underline{e}_i = \lambda_i \underline{e}_i$. Portanto, a variância e a covariância de Y_i , especificadas em (7.2) e em (7.3) são dadas por:

$$\text{Var}(Y_i) = \underline{e}_i' \Sigma \underline{e}_i = \underline{e}_i' \lambda_i \underline{e}_i = \lambda_i \underline{e}_i' \underline{e}_i = \lambda_i \quad (7.6)$$

$$\text{Cov}(Y_i, Y_k) = \underline{e}_i' \Sigma \underline{e}_k = \underline{e}_i' \lambda_k \underline{e}_k = \lambda_k \underline{e}_i' \underline{e}_k = 0 \quad i \neq k \quad (7.7)$$

Utilizando algumas propriedades matriciais estudadas no capítulo 2, pode-se demonstrar que:

$$\sum_{i=1}^p \text{Var}(X_i) = \sum_{i=1}^p \text{Var}(Y_i)$$

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \lambda_1 + \lambda_2 + \dots + \lambda_p$$

A variação total existente nas variáveis X_i , $i=1, 2, \dots, p$ é igual à variação existente nos p componentes principais. Para demonstrar isso, seja Σ a matriz de covariância entre as p variáveis X , cujos pares de autovalores e autovetores são dados por $(\lambda_i, \underline{e}_i)$. O componente principal Y_i é definido por $Y_i = \underline{e}_i' X$, o qual possui variância igual a λ_i .

Da decomposição espectral de $\Sigma = P\Lambda P'$ e sabendo que $PP' = P'P = I$ verifica-se que:

$$\text{tr}(\Sigma) = \text{tr}(P\Lambda P')$$

Uma propriedade do traço de uma matriz é: $\text{tr}(AB) = \text{tr}(BA)$. Fazendo $A = P$ e $B = \Lambda P'$, então,

$$\text{tr}(\Sigma) = \sum_{i=1}^p \sigma_{ii} = \text{tr}(P\Lambda P') = \text{tr}(\Lambda P'P) = \text{tr}(\Lambda) = \sum_{i=1}^p \lambda_i$$

E, portanto, a porcentagem da variação total explicada pelo k-ésimo componente principal é dada por (7.8).

$$\% \text{VarExp}(Y_k) = \frac{\lambda_k}{\sum_{i=1}^p \lambda_i} \times 100 \quad (7.8)$$

Em muitas situações em que se aplicam os componentes principais se uma porcentagem de 70% ou mais for atribuída aos primeiros r componentes principais, então, esses podem substituir as p variáveis originais sem perda de uma quantidade demasiada de informações. A determinação dessa porcentagem da variação explicada pelos primeiros r componentes deve ser feita pelo pesquisador interessado e que possui maior conhecimento da área estudada. A

determinação do número r de componentes para que uma determinada porcentagem fixada da informação seja contemplada por eles é um dos problemas que dificulta o emprego dessa metodologia.

Os componentes do autovetor $\underline{e}'_i = [e_{i1} \ e_{i2} \ \dots \ e_{ip}]$ podem informar sobre a importância das variáveis para o i -ésimo componente principal, por meio de suas magnitudes. No entanto, esses componentes são influenciados pela escala das variáveis. Para contornar tal problema, os pesquisadores podem utilizar uma importante medida de associação, a qual não depende da magnitude das mensurações (escala) das variáveis originais, que é o coeficiente de correlação entre Y_i e X_k . Esse coeficiente de correlação está apresentado em (7.9)

$$\rho_{Y_i, X_k} = \frac{e_{ik} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}, \quad i, k = 1, 2, \dots, p \quad (7.9)$$

Demonstração: Para demonstrar (7.9), primeiro é apresentada a definição do coeficiente de correlação. Posteriormente, foi avaliado cada termo dessa expressão individualmente.

$$\rho_{Y_i, X_k} = \frac{\text{Cov}(Y_i, X_k)}{\sqrt{\text{Var}(Y_i) \text{Var}(X_k)}}$$

Mas,

$$\text{Cov}(Y_i, X_k) = \text{Cov}(\underline{e}'_i \underline{X}, X_k) = \text{Cov}(\underline{e}'_i \underline{X}, \underline{\ell}' \underline{X})$$

com, $\ell' = [0 \dots 1 \dots 0]$, vetor composto de valores 0 e com 1 na k-ésima posição.

Logo,

$$\text{Cov}(Y_i, X_k) = \text{Cov}(\underline{e}_i' \underline{X}, \underline{\ell}' \underline{X}) = \underline{e}_i' \Sigma \underline{\ell} = \underline{\ell}' \Sigma \underline{e}_i$$

Como $\Sigma \underline{e}_i = \lambda_i \underline{e}_i$, então,

$$\text{Cov}(Y_i, X_k) = \underline{\ell}' \Sigma \underline{e}_i = \underline{\ell}' \lambda_i \underline{e}_i = \lambda_i \underline{\ell}' \underline{e}_i = \lambda_i e_{ik}$$

Da mesma forma as variâncias de Y_i e X_k são:

$$\text{Var}(Y_i) = \text{Var}(\underline{e}_i' \underline{X}) = \underline{e}_i' \Sigma \underline{e}_i = \lambda_i \underline{e}_i' \underline{e}_i = \lambda_i$$

e,

$$\text{Var}(X_k) = \sigma_{kk}$$

Assim, a prova fica completa, conforme descrito a seguir:

$$\rho_{Y_i, X_k} = \frac{\text{Cov}(Y_i, X_k)}{\sqrt{\text{Var}(Y_i) \text{Var}(X_k)}} = \frac{\lambda_i e_{ik}}{\sqrt{\lambda_i} \sqrt{\sigma_{kk}}} = \frac{\sqrt{\lambda_i} e_{ik}}{\sqrt{\sigma_{kk}}}$$

Exemplo 7.1

Sejam as variáveis aleatórias X_1 , X_2 e X_3 com covariância dada por:

$$\Sigma = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Obter os componentes principais, a correlação das variáveis originais com os componentes e verificar a veracidade da afirmativa a seguir de forma numérica:

$$\sum_{i=1}^p \text{Var}(X_i) = \sum_{i=1}^p \text{Var}(Y_i)$$

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \lambda_1 + \lambda_2 + \dots + \lambda_p$$

Aplicando-se o power method, determinaram-se os pares de autovalores e autovetores de Σ , os quais são:

$$\lambda_1 = 5 \quad \xi'_1 = [0,7071 \quad 0,7071 \quad 0], \quad \lambda_2 = 3 \quad \xi'_2 = [0,7071 \quad -0,7071 \quad 0] \quad \text{e} \quad \lambda_3 = 2 \quad \xi'_3 = [0 \quad 0 \quad 1]$$

Os componentes principais são:

$$Y_1 = \xi'_1 \mathbf{X} = 0,7071X_1 + 0,7071X_2$$

$$Y_2 = \xi'_2 \mathbf{X} = 0,7071X_1 - 0,7071X_2$$

$$Y_3 = \underline{e}_3' \underline{X} = X_3$$

A variável X_3 é individualmente um de os componentes principais por não ser correlacionada com nenhuma das outras duas variáveis. As variâncias dos componentes principais são:

$$\text{Var}(Y_1) = \lambda_1 = 5, \quad \text{Var}(Y_2) = \lambda_2 = 3 \quad \text{e} \quad \text{Var}(Y_3) = \lambda_3 = 2$$

Pode-se mostrar, a título de ilustração, que:

$$\begin{aligned} \text{Var}(Y_1) &= \text{Var}\left(\frac{\sqrt{2}}{2}X_1 + \frac{\sqrt{2}}{2}X_2\right) = \text{Var}\left(\frac{\sqrt{2}}{2}X_1\right) + \text{Var}\left(\frac{\sqrt{2}}{2}X_2\right) + 2\text{Cov}\left(\frac{\sqrt{2}}{2}X_1, \frac{\sqrt{2}}{2}X_2\right) = \\ &= \frac{1}{2}\text{Var}(X_1) + \frac{1}{2}\text{Var}(X_2) + 2\frac{\sqrt{2}}{2}\frac{\sqrt{2}}{2}\text{Cov}(X_1, X_2) = \frac{1}{2}4 + \frac{1}{2}4 + 1 = 5 = \lambda_1 \end{aligned}$$

Verifica-se, também, que:

$$\sigma_{11} + \sigma_{22} + \sigma_{33} = \lambda_1 + \lambda_2 + \lambda_3$$

$$4+4+2=5+3+2$$

$$10=10 \text{ c.q.m.}$$

A percentagem da explicação dada por cada componente é apresentada na tabela seguinte.

Component e	Var(Y _i)=λ _i	% da variação explicada	% variação acumulada
Y ₁	5	50	50
Y ₂	3	30	80
Y ₃	2	20	100

Os coeficientes de correlação entre os componentes e as variáveis originais são:

Component e	X ₁	X ₂	X ₃
Y ₁	0,7906	0,7906	0,0000
Y ₂	0,6124	-0,6124	0,0000
Y ₃	0,0000	0,0000	1,0000

Para ilustrar um dos cálculos usando a expressão (7.9), apresentase a seguir a correlação entre Y₁ e X₁.

$$\rho_{Y_1, X_1} = \frac{e_{11}\sqrt{\lambda_1}}{\sqrt{\sigma_{11}}} = \frac{\sqrt{2}/2 \sqrt{5}}{\sqrt{4}} = 0,7906.$$

Para o componente principal mais importante (Y₁), concluiu-se que X₁ e X₂ são igualmente importantes.

Os componentes principais podem ser obtidos pela padronização das variáveis originais por:

$$Z_i = \frac{X_i - \mu_i}{\sqrt{\sigma_{ii}}} \quad (7.10)$$

Em notação matricial tem-se:

$$\underline{Z} = V^{-1/2} (\underline{X} - \underline{\mu}) \quad (7.11)$$

em $V^{-1/2}$ é uma matriz diagonal com os elementos da diagonal dados $1/\sqrt{\sigma_{ii}}$. É fácil verificar que:

$$E(\underline{Z}) = \underline{0} \text{ e } \text{Cov}(\underline{Z}) = V^{-1/2} \Sigma V^{-1/2} = \rho$$

Então, os componentes principais de \underline{Z} são dados pelos autovalores e autovetores de ρ , matriz de correlação de \underline{X} . Os autovalores e autovetores de Σ são, em geral, diferentes daqueles derivados de ρ .

Sejam as variáveis padronizadas Z_1, Z_2, \dots, Z_p disposta no vetor \underline{Z} com $\text{Cov}(\underline{Z}) = \rho$, então, os componentes principais são dados por:

$$Y_i = \underline{e}_i' \underline{Z} = \underline{e}_i' V^{-1/2} (\underline{X} - \underline{\mu}), \quad i=1, 2, \dots, p \quad (7.12)$$

Da mesma forma, verifica-se que:

$$\begin{aligned} \sum_{i=1}^p \text{Var}(Y_i) &= \sum_{i=1}^p \text{Var}(Z_i) = p \\ \sum_{i=1}^p \lambda_i &= p \end{aligned} \quad (7.13)$$

Também se verifica que:

$$\rho_{Y_i, Z_k} = e_{ik} \sqrt{\lambda_i} \quad (7.14)$$

Sendo que em todos esses casos $(\lambda_i, \underline{e}_i)$ são os autovalores e autovetores de ρ , com $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. As demonstrações de (7.12), (7.13) e (7.14) podem ser realizadas da mesma forma que as demonstrações anteriores, substituindo Σ por ρ .

Para algumas matrizes de covariância, com estruturas especiais, existem simples formas de se expressar os componentes principais. Serão tratados alguns desses casos, conforme apresentado em Johnson e Wichern, (1998) e em Morrison (1976). Para uma matriz Σ diagonal,

$$\Sigma = \begin{bmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{pp} \end{bmatrix} \quad (7.15)$$

Os autovalores e autovetores são dados por:

$$\lambda_i = \sigma_{ii} \quad \text{e} \quad \underline{e}_i^t = [0 \quad \dots \quad 0 \quad 1 \quad 0 \quad \dots \quad 0] \quad \text{com } 1 \text{ na } i\text{-ésima posição e } 0 \text{ nas demais.}$$

A demonstração disso pode ser facilmente realizada, uma vez que das equações de maximização de formas quadráticas verifica-se que : $\Sigma \underline{e}_i = \lambda_i \underline{e}_i$.

Assumindo-se as definições anteriores para os autovalores e autovetores verifica-se que:

$$\begin{aligned} \Sigma \underline{e}_i &= \lambda_i \underline{e}_i \\ &= \Sigma \underline{e}_i = \sigma_{ii} \underline{e}_i \\ &= \begin{bmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{pp} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \sigma_{ii} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \end{aligned}$$

Dessa forma, pode-se concluir que $(\sigma_{ii}, \underline{e}_i)$, com \underline{e}_i definido anteriormente, são os pares de autovalores e autovetores de Σ . Desde que os componentes principais são dados pelas combinações lineares $\underline{e}_i' \underline{X} = X_i$, então, os componentes principais são as próprias variáveis originais não correlacionadas, cujos autovalores são as próprias variâncias originais das respectivas variáveis aleatórias. Do ponto de vista de extração de componentes principais nada pode ser ganho, uma vez que os eixos originais já estão no sentido de maior variabilidade. Dessa forma não há necessidade para fazer rotação dos eixos originais. A standardização não altera a situação, uma vez que $\rho=1$, e o par autovalor e componente principal é dado por $(1, Z_i)$, em que Z_i é a i -ésima variável padronizada.

Outro tipo de matriz de covariância com determinado padrão é apresentado a seguir, o qual descreve muitas vezes o comportamento de

entidades biológicas, desempenha um papel importante na teoria dos componentes principais.

$$\Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{bmatrix} \quad (7.16)$$

A matriz de correlação correspondente é dada por:

$$\rho = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix} \quad (7.17)$$

que implica em uma estrutura de igualdade de correlação entre as p variáveis estudadas.

Morrison (1976) demonstra que os componentes principais de (7.16) são dados por dois grupos. O primeiro grupo com o primeiro componente e o segundo com os demais componentes principais. O primeiro componente principal de (7.16) é definido pelo par autovalor e autovetor apresentado a seguir.

$$\lambda_1 = \sigma^2 [1 + (p-1)\rho] \quad (7.18)$$

$$\underline{e}_1 = \left[\frac{1}{\sqrt{p}}, \frac{1}{\sqrt{p}}, \dots, \frac{1}{\sqrt{p}} \right] \quad (7.19)$$

Para a matriz de correlação definida em (7.17), pode-se demonstrar que 7.18 e 7.19 permanecem válidos, sendo necessário apenas fazer $\sigma^2=1$. A proporção da explicação do primeiro componente principal é dada por $100[1+(p-1)\rho]/p\%$ do total do conjunto de variáveis. Se ρ é próximo a 1 o primeiro componente principal terá uma elevada explicação da variação total.

Os demais $(p-1)$ componentes principais possuem valores característicos iguais, dados por:

$$\sigma^2(1-\rho) \quad (7.20)$$

e seus respectivos autovetores são iguais a:

$$\left\{ \begin{array}{l} \hat{e}'_2 = \left[\frac{1}{\sqrt{1 \times 2}}, \frac{-1}{\sqrt{1 \times 2}}, 0, \dots, 0 \right] \\ \hat{e}'_3 = \left[\frac{1}{\sqrt{2 \times 3}}, \frac{1}{\sqrt{2 \times 3}}, \frac{-2}{\sqrt{2 \times 3}}, 0, \dots, 0 \right] \\ \vdots \\ \hat{e}'_i = \left[\frac{1}{\sqrt{(i-1) \times i}}, \dots, \frac{1}{\sqrt{(i-1) \times i}}, \frac{-(i-1)}{\sqrt{(i-1) \times i}}, 0, \dots, 0 \right] \\ \vdots \\ \hat{e}'_p = \left[\frac{1}{\sqrt{(p-1) \times p}}, \dots, \frac{1}{\sqrt{(p-1) \times p}}, \frac{-(p-1)}{\sqrt{(p-1) \times p}} \right] \end{array} \right. \quad (7.21)$$

Finalmente é tratada a situação em que o vetor \underline{X} é uma variável aleatória da distribuição normal multivariada, ou seja, $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$. Nesse caso os componentes principais têm uma atrativa interpretação. Foi demonstrado no capítulo 4 que a densidade de \underline{X} é constante na elipsóide centrada em $\underline{\mu}$,

$$(\underline{X} - \underline{\mu})' \Sigma^{-1} (\underline{X} - \underline{\mu}) = c^2 = \chi_p^2(\alpha)$$

cujos eixos são dados por $\pm \sqrt{\chi_p^2(\alpha)} \sqrt{\lambda_i} \underline{e}_i$, $i = 1, 2, \dots, p$, em que $(\lambda_i, \underline{e}_i)$ são os pares de autovalor-autovetor de Σ . É possível verificar, fazendo $\underline{\mu} = \underline{0}$ por conveniência de algumas demonstrações que se seguem, que:

$$\chi_p^2(\alpha) = \underline{X}' \Sigma^{-1} \underline{X} = \frac{1}{\lambda_1} (\underline{e}_1' \underline{X})^2 + \frac{1}{\lambda_2} (\underline{e}_2' \underline{X})^2 + \dots + \frac{1}{\lambda_p} (\underline{e}_p' \underline{X})^2$$

em que $\underline{e}_i' \underline{X}$, $i = 1, 2, \dots, p$ são os componentes principais de \underline{X} . Fazendo $Y_i = \underline{e}_i' \underline{X}$, $i = 1, 2, \dots, p$ tem-se

$$\chi_p^2(\alpha) = \underline{X}' \Sigma^{-1} \underline{X} = \frac{1}{\lambda_1} Y_1^2 + \frac{1}{\lambda_2} Y_2^2 + \dots + \frac{1}{\lambda_p} Y_p^2$$

Essa última equação define uma elipsóide com os eixos coordenados Y_1, Y_2, \dots, Y_p dispostos nas direções de $\underline{e}_1, \underline{e}_2, \dots, \underline{e}_p$, respectivamente.

Como λ_1 é o maior autovalor, o maior eixo tem a direção definida por \underline{e}_1 , os eixos remanescentes têm a direção definida por $\underline{e}_2, \dots, \underline{e}_p$.

Foi assumido que $\underline{\mu} = \underline{0}$. No entanto, é pouco provável que isso aconteça em uma situação real. Todavia, as interpretações definidas anteriormente são válidas da mesma forma, apenas sendo necessário definir o i -ésimo componente principal centrado na média, por:

$$Y_i = \underline{e}_i^t (\underline{X} - \underline{\mu}), \quad i = 1, 2, \dots, p \quad (7.22)$$

o qual tem média zero e direção definida por \underline{e}_i . Na Figura 7.1 ilustram-se os componentes principais bivariados com densidade fixa de 95%. A rotação dos eixos X_1 e X_2 nos novos eixos Y_1 e Y_2 são a essência dos componentes principais.

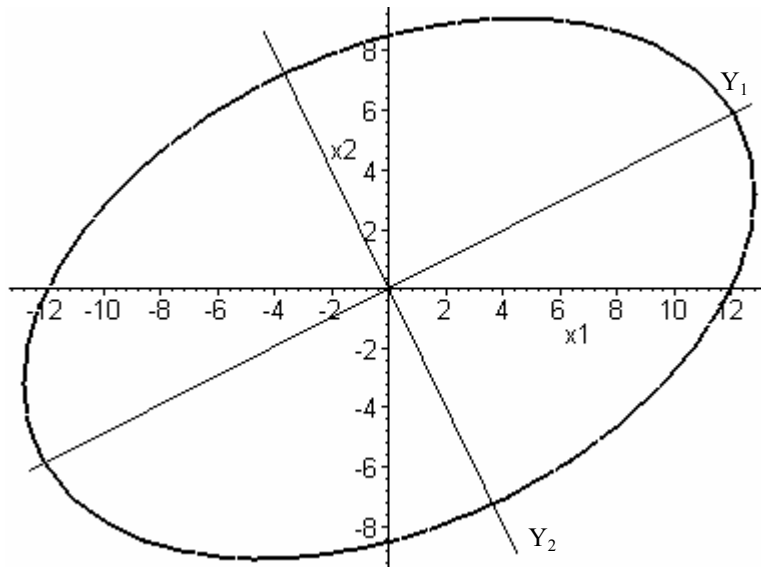


Figura 7.1. A elipse de 95% de densidade constante e os componentes principais Y_1 e Y_2 para a distribuição normal bivariada com média $\underline{\mu} = \underline{0}$.

7.3. Componentes principais amostrais

Seja $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n$ uma amostra aleatória retirada de uma população p -variada qualquer com média $\underline{\mu}$ e covariância Σ . O vetor de médias amostrais é $\bar{\underline{X}}$, a matriz de covariância amostral é S e a matriz de correlação amostral é R . O objetivo dessa seção é apresentar os conceitos de componentes principais para a estrutura de covariância amostral.

As combinações lineares das variáveis mensuradas que maximizam a variação total da amostra e que são mutuamente ortogonais são chamadas de componentes principais amostrais. Seja a forma quadrática

$$Q = \hat{\text{Var}}(\hat{Y}) = \hat{\text{Var}}(\underline{e}' X) = \underline{e}' S \underline{e}$$

O máximo de Q não existe, pois quanto maior for o comprimento de \underline{e} maior será o valor de Q . É conveniente tomar-se o máximo de Q restrito ao comprimento unitário de \underline{e} . Dessa forma, o máximo tem que ser obtido da forma quadrática restrita seguinte.

$$\lambda = \frac{\underline{e}' S \underline{e}}{\underline{e}' \underline{e}}$$

O máximo é obtido tomando-se a derivada em relação a \underline{e} e igualando-se a derivada a zero. O sistema obtido é resolvido em relação a \underline{e} e as soluções obtidas referem-se ao máximo.

$$\frac{\partial \lambda}{\partial \underline{e}} = \frac{2S\underline{\hat{e}}(\underline{\hat{e}}'\underline{\hat{e}}) - 2(\underline{\hat{e}}'S\underline{\hat{e}})\underline{\hat{e}}}{(\underline{\hat{e}}'\underline{\hat{e}})^2} = \frac{2}{\underline{\hat{e}}'\underline{\hat{e}}} \left(S - \frac{\underline{\hat{e}}'S\underline{\hat{e}}}{\underline{\hat{e}}'\underline{\hat{e}}} I \right) \underline{\hat{e}} = \underline{0}$$

$$\left(S - \frac{\underline{\hat{e}}'S\underline{\hat{e}}}{\underline{\hat{e}}'\underline{\hat{e}}} I \right) \underline{\hat{e}} = \underline{0}$$

A equação resultante é dada por:

$$(S - \hat{\lambda}I)\underline{\hat{e}} = \underline{0} \tag{7.23}$$

A solução de (7.23) conduz aos pares de autovalores e autovetores $(\hat{\lambda}_i; \hat{e}_i)$ de S, que correspondem a variância amostral e combinação linear que definem os componentes principais amostrais, para $i=1, 2, \dots, p$.

Portanto, o i-ésimo componente principal amostral é:

$$\hat{Y}_i = \hat{e}_i^t \mathbf{X} = \hat{e}_{i1} X_1 + \hat{e}_{i2} X_2 + \dots + \hat{e}_{ip} X_p, \quad i = 1, 2, \dots, p \quad (7.24)$$

em que $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p \geq 0$ são os autovalores amostrais de S correspondentes.

O estimador da variância amostral dos componentes principais é:

$$\hat{\text{Var}}(\hat{Y}_k) = \hat{\lambda}_k, \quad k = 1, 2, \dots, p \quad (7.25)$$

e a covariância entre dois componentes principais (i e k) é:

$$\text{Cov}(\hat{Y}_i, \hat{Y}_k) = 0, \quad i \neq k = 1, 2, \dots, p \quad (7.26)$$

Pela mesma razão apresentada para os componentes principais populacionais, verifica-se que a variação total explicada pelos componentes principais amostrais é igual a $\sum_{i=1}^p \hat{\lambda}_i = \sum_{i=1}^p S_{ii}$. A partir da decomposição espectral de

S, dada por $S = \hat{P}\hat{\Lambda}\hat{P}^t$ e da propriedade que $\text{tr}(AB) = \text{Tr}(BA)$ demonstra-se que:

$$\text{tr}(S) = \sum_{i=1}^p S_{ii} = \text{tr}(\hat{P}\hat{\Lambda}\hat{P}^t) = \text{tr}(\hat{\Lambda}\hat{P}^t\hat{P}) = \text{tr}(\hat{\Lambda}) = \sum_{i=1}^p \hat{\lambda}_i$$

Dessa forma, a explicação do k-ésimo componente principal amostral da variação total do sistema é:

$$\% \text{VarExp}(\hat{Y}_k) = \frac{\hat{\lambda}_k}{\sum_{i=1}^p \hat{\lambda}_i} \times 100 \quad (7.27)$$

A correlação amostral entre \hat{Y}_i e X_k é definida por:

$$r_{\hat{Y}_i, X_k} = \frac{\hat{e}_{ik} \sqrt{\hat{\lambda}_i}}{\sqrt{S_{kk}}}, \quad i, k = 1, 2, \dots, p \quad (7.28)$$

Os componentes principais podem ser definidos por componentes principais amostrais centrados na média amostral $\bar{\underline{X}}$, da seguinte forma:

$$\hat{Y}_i = \hat{e}_i' (\underline{X} - \bar{\underline{X}}) = \hat{e}_{i1} (X_1 - \bar{X}_1) + \hat{e}_{i2} (X_2 - \bar{X}_2) + \dots + \hat{e}_{ip} (X_p - \bar{X}_p), \quad i = 1, 2, \dots, p \quad (7.29)$$

Se o vetor \underline{X} for substituído em (7.29) por \underline{X}_j (vetor de observações amostrais), pode-se obter os escores dos componentes principais. Esses escores são plotados, muitas vezes, com o intuito de agrupar objetos ou itens, simplificar a representação para uma ou duas dimensões, entre outras aplicações.

Os componentes principais, em geral, não são invariantes com relação a transformações nas escalas. A mudança de escala mais usual é aquela

que transforma as escalas de as variáveis para uma outra sem dimensão, cuja média é igual a zero e a variância é igual a 1. A padronização é obtida por:

$$\underline{z}_j = D^{-1/2} (\underline{X}_j - \bar{X}), \quad j=1,2,\dots,n \quad (7.30)$$

em que $D^{-1/2} = \text{Diag}(1/\sqrt{S_{11}}, 1/\sqrt{S_{22}}, \dots, 1/\sqrt{S_{pp}})$. O estimador de a covariância de \underline{Z} é dado por:

$$\widehat{\text{Cov}}(\underline{Z}) = D^{-1/2} \widehat{\text{Cov}}(\underline{X}) D^{-1/2} = D^{-1/2} S D^{-1/2} = R \quad (7.31)$$

Os componentes principais obtidos de R são definidos pelos pares de autovalores e autovetores de R $(\hat{\lambda}_i; \hat{e}_i)$. Assim, o i-ésimo componente principal amostral obtido da matriz de correlação amostral é dado por:

$$\hat{Y}_i = \hat{e}_i' \underline{Z} = \hat{e}_{i1} Z_1 + \hat{e}_{i2} Z_2 + \dots + \hat{e}_{ip} Z_p, \quad i=1,2,\dots,p \quad (7.32)$$

A variação total explicada pelo k-ésimo componente principal é dada por:

$$\% \text{VarExp}(\hat{Y}_k) = \frac{\hat{\lambda}_k}{p} \times 100 \quad (7.33)$$

A correlação amostral entre \hat{Y}_i e Z_k é definida por:

$$r_{\hat{Y}_i, Z_k} = \hat{e}_{ik} \sqrt{\hat{\lambda}_i}, \quad i, k = 1, 2, \dots, p \quad (7.34)$$

Pequenos valores para os últimos autovalores, tanto de S como de R, indicam, em geral, a presença de dependência linear no conjunto de dados. Neste contexto pelo menos uma variável é redundante e pode ser eliminada do conjunto de variáveis originais.

Existe sempre a questão importante de o número de componentes que deve ser retido. Não existe uma resposta definitiva para essa questão. Os aspectos que devem ser considerados incluem a quantidade da variação amostral explicada, o tamanho relativo dos autovalores e a interpretação subjetiva dos componentes. Uma ferramenta visual importante para auxiliar a determinação de o número suficiente de componentes a ser retido é o “scree plot”. O termo “scree” refere-se ao acúmulo de rochas nas bases de um penhasco, portanto os “scree plots” serão considerados “gráficos de cotovelos”. Na Figura 7.2 observa-se que um cotovelo é formado aproximadamente na posição $i=4$. Isso significa que os componentes acima de $\hat{\lambda}_3$ possuem aproximadamente a mesma magnitude e são relativamente pequenos. Isso indica que os três primeiros, talvez os quatro primeiros componentes são suficientes para resumir a variação amostral total.

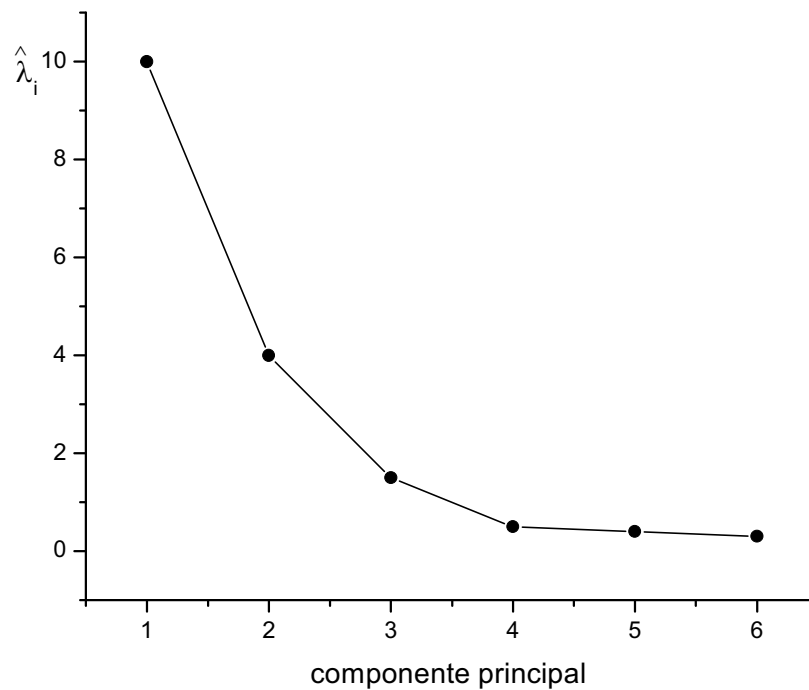


Figura 7.2. “Scree plot” de um exemplo com $p=6$ componentes principais para ilustrar o processo de determinação de o número apropriado de componentes a ser retido.

7.4. Gráficos dos componentes principais

Os gráficos provenientes dos componentes principais podem ser reveladores de diversos aspectos presentes nos dados de interesse do pesquisador. Em muitas áreas os pesquisadores utilizam os primeiros e mais importantes componentes para agrupar objetos e itens de acordo com a representação em duas ou no máximo três dimensões retidas. Os gráficos dos componentes principais podem revelar observações suspeitas, como também

permitir uma avaliação da suposição de normalidade. Por se tratarem de combinações lineares de p variáveis, supostamente normais, é possível assumir a normalidade para os componentes principais. O teste de normalidade pode ser feito em apenas alguns poucos componentes, o que pode simplificar a complexidades das análises necessárias e reduzir o número de testes a ser realizado.

Os valores amostrais de os componentes principais obtidos a partir dos dados amostrais originais são chamados de escores. A equação (7.35) refere-se a definição do escore do k -ésimo componente principal, para a j -ésima observação amostral.

$$\hat{Y}_{jk} = \hat{e}_k^t \underline{X}_j = \hat{e}_{k1} X_{j1} + \hat{e}_{k2} X_{j2} + \dots + \hat{e}_{kp} X_{jp}, \quad k = 1, 2, \dots, p; j = 1, 2, \dots, n \quad (7.35)$$

De uma forma geral, os escores dos p componentes principais, representados pelo vetor $\hat{\underline{Y}}_j^t = [\hat{Y}_{j1} \hat{Y}_{j2} \dots \hat{Y}_{jp}]$ para a j -ésima observação amostral

$\underline{X}_j^t = [X_{j1} \ X_{j2} \ \dots \ X_{jp}]$, são dados por:

$$\hat{\underline{Y}}_j = \hat{\mathbf{P}}^t \underline{X}_j = \begin{bmatrix} \hat{e}_1^t \\ \hat{e}_2^t \\ \vdots \\ \hat{e}_p^t \end{bmatrix} \underline{X}_j \quad (7.36)$$

Para o agrupamento de objetos e também para avaliar desvios de normalidade obtêm-se gráficos dos primeiros componentes retidos em um diagrama contendo pares de componentes. Também, é possível obter os Q-Q plots para cada componente, conforme descrição realizada no capítulo 4. Desvios de normalidade podem ser verificados e o teste da correlação Q-Q plot pode ser realizado.

Para a verificação de observações suspeitas os gráficos dos últimos componentes principais obtidos da plotagem dos mesmos tomados dois a dois são utilizados. Esse tipo de gráfico pode ajudar a identificar observações suspeitas. Também, com esse intuito os Q-Q plots desses componentes, de menor importância para a variação total, são utilizados.

Da equação (7.36) e lembrando que \hat{P} é uma matriz ortogonal, pois $\hat{P}\hat{P}^t = \hat{P}^t\hat{P} = I$, portanto $(\hat{P}^t)^{-1} = \hat{P}$, pode-se demonstrar que:

$$\underline{X}_j = \hat{P}\hat{Y}_j = \begin{bmatrix} \hat{e}_1 & \hat{e}_2 & \cdots & \hat{e}_p \end{bmatrix} \hat{Y}_j \quad (7.37)$$

$$\underline{X}_j = \hat{Y}_{j1}\hat{e}_1 + \hat{Y}_{j2}\hat{e}_2 + \cdots + \hat{Y}_{jp}\hat{e}_p$$

Essa é uma importante equação que mostra que a observação amostral multivariada \underline{X}_j pode ser recuperada dos escores dos componentes principais correspondentes. Essa é uma importante forma de identificar com elevada precisão as observações suspeitas. Para isso um número q de componentes principais $q \leq p$ é retido para ajustar as n observações amostrais

multivariadas. Dessa forma, uma medida da qualidade desse ajuste é obtida avaliando quanto $\hat{Y}_{j1}\hat{\varepsilon}_{j1} + \hat{Y}_{j2}\hat{\varepsilon}_{j2} + \dots + \hat{Y}_{jq}\hat{\varepsilon}_{jq}$ difere de \underline{X}_j , tendo como desvio o valor dado por $\hat{Y}_{jq+1}\hat{\varepsilon}_{jq+1} + \hat{Y}_{jq+2}\hat{\varepsilon}_{jq+2} + \dots + \hat{Y}_{jp}\hat{\varepsilon}_{jp}$. Essa medida é feita tomando-se o quadrado desse desvio, o qual refere-se ao seu comprimento quadrático, ou seja, por $\hat{Y}_{jq+1}^2 + \hat{Y}_{jq+2}^2 + \dots + \hat{Y}_{jp}^2$. As observações consideradas suspeitas são aquelas que possuem pelo menos uma das coordenadas de $\hat{Y}_{jq+1}, \hat{Y}_{jq+2}, \dots, \hat{Y}_{jp}$ que contribui para o comprimento quadrático total com grande valor.

7.5. Inferências para grandes amostras

Foram apresentados os conceitos fundamentais dos componentes principais. A essência dos componentes principais está na obtenção dos autovalores e autovetores da matriz de covariância (correlação). Os autovetores determinam a rotação a ser realizada nos eixos coordenados originais nos sentidos de maior variabilidade e os autovalores determinam as variâncias desses novos eixos coordenados. As decisões com relação aos componentes principais devem ser tomadas com base nos pares de autovalores-autovetores, $(\hat{\lambda}_i; \hat{\varepsilon}_i)$, estimados na amostra. Esses autovalores e autovetores são diferentes dos respectivos valores populacionais devido às variações amostrais. Derivações à respeito das distribuições amostrais de $\hat{\lambda}_i$ e de $\hat{\varepsilon}_i$ são apresentadas em Anderson

(1963). Os resultados relativos aos resultados de grandes amostras são apresentados a seguir, de uma forma resumida.

Suponha que $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n$ seja uma amostra aleatória retirada de uma população p-variada qualquer com média $\underline{\mu}$ e covariância Σ . O vetor de médias amostrais é $\bar{\underline{X}}$, a matriz de covariância amostral é S e a matriz de correlação amostral é R. Suponha que Σ possui autovalores (desconhecidos) distintos e positivos, quais sejam, $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$ com correspondentes autovetores (desconhecidos) $\underline{e}_1, \underline{e}_2, \dots, \underline{e}_p$. O estimador amostral de Σ é S, sendo que os estimadores de λ_i e \underline{e}_i são $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots > \hat{\lambda}_p > 0$ e $\hat{\underline{e}}_1, \hat{\underline{e}}_2, \dots, \hat{\underline{e}}_p$.

Girshik (1939), Lawley (1956) e Anderson (1963) demonstraram que os seguintes resultados se verificam para grandes amostras. Dessa forma, os resultados que se seguem referem-se a teoria de distribuições de grandes amostras para os autovalores $\hat{\underline{\lambda}}^t = [\hat{\lambda}_1 \hat{\lambda}_2 \dots \hat{\lambda}_p]$ e para os autovetores $\hat{\underline{e}}_1, \hat{\underline{e}}_2, \dots, \hat{\underline{e}}_p$ de S. Fazendo Λ uma matriz diagonal dos autovalores $\lambda_1, \lambda_2, \dots, \lambda_p$ de Σ , então,

1. $\sqrt{n}(\hat{\underline{\lambda}} - \underline{\lambda})$ tem distribuição aproximadamente $N_p(\underline{0}, 2\Lambda^2)$.

2. Seja

$$E_i = \lambda_i \sum_{\substack{k=1 \\ k \neq i}}^p \left[\frac{\lambda_k}{(\lambda_k - \lambda_i)^2} \underline{e}_k \underline{e}_k^t \right] \quad (7.38)$$

então, $\sqrt{n}(\hat{\underline{e}}_i - \underline{e}_i) \simeq N_{p-1}(\underline{0}, E_i)$.

3. Cada $\hat{\lambda}_i$ tem distribuição independente dos elementos do vetor característico associado $\hat{\xi}_i$.

4. A covariância do r-ésimo elemento de $\hat{\xi}_i$ e o s-ésimo elemento de $\hat{\xi}_j$ ($i \neq j$) é:

$$\text{Cov}(\hat{e}_{ir}, \hat{e}_{js}) = \frac{\lambda_i \lambda_j e_{is} e_{jr}}{n(\lambda_i - \lambda_j)^2} \quad (i \neq j) \quad (7.39)$$

Os resultados 1 a 4 são referentes às propriedades distribucionais de grandes amostras e válidas para o caso de p distintas raízes características. Entretanto, Anderson (1963) aponta que o resultado 2 requer somente que λ_i seja distinto dos demais $p-1$ valores característicos, os quais podem ter qualquer multiplicidade. Esses resultados podem ser utilizados para construir testes de hipóteses e intervalos de confiança para os autovalores e autovetores populacionais.

O resultado 1 implica, em grande amostras, que os $\hat{\lambda}_i$'s são independentemente distribuídos com distribuição aproximadamente $N(\lambda_i, 2\lambda_i^2/n)$. As inferências podem ser derivadas desse resultado. O intervalo de confiança para λ_i pode ser obtido a partir da afirmativa probabilística:

$$P \left[\frac{|\hat{\lambda}_i - \lambda_i|}{\lambda_i \sqrt{\frac{2}{n}}} \leq Z(\alpha/2) \right] = 1 - \alpha \quad (7.40)$$

O intervalo de confiança resultante é dado por:

$$IC_{\hat{\lambda}_i}(1-\alpha) : \left[\frac{\hat{\lambda}_i}{1+Z(\alpha/2)\sqrt{\frac{2}{n}}}; \frac{\hat{\lambda}_i}{1-Z(\alpha/2)\sqrt{\frac{2}{n}}} \right] \quad (7.41)$$

Obviamente os valores de α e de n devem ser apropriados para que o limite superior de (7.41) seja válido. Testes de hipóteses de o tipo $H_0 : \lambda_i = \lambda_0$ pode ser realizado calculando-se o escore normal padrão:

$$Z_c = \frac{\hat{\lambda}_i - \lambda_0}{\lambda_0 \sqrt{\frac{n}{2}}} \quad (7.42)$$

Uma inferência importante e mais geral sobre a estrutura de dependência é apresentada por Anderson (1963). O teste de hipótese de que os r autovalores intermediários de Σ sejam iguais é apresentado. A hipótese de interesse é:

$$H_0 : \lambda_{q+1} = \lambda_{q+2} = \dots = \lambda_{q+r} \quad (7.43)$$

Aos q maiores e aos $(p-q-r)$ menores autovalores não são impostas restrições quanto aos seus valores ou multiplicidades. A hipótese alternativa é

especificada da seguinte forma: H_1 : pelo menos um dos r autovalores difere dos demais intermediários. O teste de razão de verossimilhança conduz a estatística

$$\chi_c^2 = -(n-1) \sum_{j=q+1}^{q+r} \ln(\hat{\lambda}_j) + (n-1)r \ln \left(\frac{\sum_{j=q+1}^{q+r} \hat{\lambda}_j}{r} \right) \quad (7.44)$$

que tem distribuição aproximadamente de qui-quadrado sob H_0 com $v=r(r+1)/2 - 1$ graus de liberdade para grandes amostras.

Um caso especial importante deste teste de hipótese ocorre quando $q+r=p$ ou quando a variação das últimas q dimensões é esférica.

Outro importante teste refere-se aos autovetores. A hipótese de que o i -ésimo autovetor populacional de Σ é igual a um vetor de constantes com norma 1 é apresentada a seguir.

$$H_0 : \underline{e}_i = \underline{e}_0 \quad (7.45)$$

O teste da hipótese nula (7.45) é realizado com base no resultado 2 dessa seção e na matriz de covariância E_i definida em (7.38) devidamente substituída pelo seu estimador \hat{E}_i , o qual é obtido pela substituição de λ_i e \underline{e}_i pelos seus estimadores $\hat{\lambda}_i$ e $\hat{\underline{e}}_i$. Assim, Anderson (1963) demonstra que o teste estatístico dado por:

$$\chi_c^2 = n \left(\hat{\lambda}_i \mathbf{e}_0^t \mathbf{S}^{-1} \mathbf{e}_0 + \frac{1}{\hat{\lambda}_i} \mathbf{e}_0^t \mathbf{S} \mathbf{e}_0 - 2 \right) = n (\hat{\mathbf{e}}_i - \mathbf{e}_0)^t \hat{\mathbf{E}}_i^g (\hat{\mathbf{e}}_i - \mathbf{e}_0) \quad (7.46)$$

tem distribuição assintótica de qui-quadrado com $p-1$ graus de liberdade se H_0 for verdadeira. Em que $\hat{\mathbf{E}}_i^g$ é uma inversa generalizada de $\hat{\mathbf{E}}_i$.

Demonstração: A matriz E_i do resultado 2 pode ser reescrita na forma matricial como se segue. Para isso, serão definidas as seguintes matrizes:

$$\Lambda_i = \begin{bmatrix} \frac{\lambda_1}{(\lambda_1 - \lambda_i)^2} & 0 & \cdots & 0 \\ 0 & \frac{\lambda_2}{(\lambda_2 - \lambda_i)^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\lambda_p}{(\lambda_p - \lambda_i)^2} \end{bmatrix} \quad \text{uma matriz } (p-1) \times (p-1) \text{ originária da}$$

eliminação da i -ésima linha e i -ésima coluna de uma matriz $\text{Diag} \left(\frac{\lambda_j}{(\lambda_j - \lambda_i)^2} \right)$, $p \times p$.

$P_i = [\mathbf{e}_1 \quad \mathbf{e}_2 \quad \cdots \quad \mathbf{e}_p]$ $p \times (p-1)$, sendo \mathbf{e}_j os autovetores de Σ , com $j \neq i$ e dimensão $p \times (p-1)$.

Assim, pode-se definir E_i por:

$$E_i = \lambda_i P_i \Lambda_i P_i^t = \lambda_i \sum_{\substack{j=1 \\ j \neq i}}^p \left[\frac{\lambda_j}{(\lambda_i - \lambda_j)^2} \mathbf{e}_j \mathbf{e}_j^t \right]$$

e sua inversa generalizada, devido a E_i ter posto (dimensão) $p-1$, por:

$$E_i^g = \lambda_i P_i \Lambda_i^{-1} P_i^t = \lambda_i \sum_{\substack{j=1 \\ j \neq i}}^p \left[\frac{(\lambda_i - \lambda_j)^2}{\lambda_j} \underline{e}_j \underline{e}_j^t \right]$$

No capítulo 4 foi visto que sob normalidade ou para grandes amostras a forma quadrática

$$\sqrt{n}(\underline{e}_i - \underline{e}_0)^t E_i^g \sqrt{n}(\underline{e}_i - \underline{e}_0) \sim \chi_{p-1}^2$$

Os graus de liberdade são iguais a (p-1) e não a p devido a E_i ter posto incompleto (p-1). Devido aos autovetores de E_i^g e o autovetor \underline{e}_i serem ortogonais, a forma quadrática anterior pode ser simplificada por:

$$\begin{aligned} n(\underline{e}_i - \underline{e}_0)^t E_i^g (\underline{e}_i - \underline{e}_0) &= n \underline{e}_0^t E_i^g \underline{e}_0 = \frac{n}{\lambda_i} \underline{e}_0^t \sum_{\substack{j=1 \\ j \neq i}}^p \left[\frac{(\lambda_i - \lambda_j)^2}{\lambda_j} \underline{e}_j \underline{e}_j^t \right] \underline{e}_0 = \\ &= \frac{n}{\lambda_i} \underline{e}_0^t \sum_{\substack{j=1 \\ j \neq i}}^p \left[\frac{(\lambda_j^2 - 2\lambda_i \lambda_j + \lambda_i^2)}{\lambda_j} \underline{e}_j \underline{e}_j^t \right] \underline{e}_0 = \frac{n}{\lambda_i} \underline{e}_0^t \left\{ \sum_{\substack{j=1 \\ j \neq i}}^p [\lambda_j \underline{e}_j \underline{e}_j^t] - 2\lambda_i \sum_{\substack{j=1 \\ j \neq i}}^p [\underline{e}_j \underline{e}_j^t] + \lambda_i^2 \sum_{\substack{j=1 \\ j \neq i}}^p \left[\frac{1}{\lambda_j} \underline{e}_j \underline{e}_j^t \right] \right\} \underline{e}_0 = \end{aligned}$$

Como $\sum_{j=1}^p [\lambda_j \underline{e}_j \underline{e}_j^t] = \Sigma$, além disso, somando e subtraindo $\lambda_i \underline{e}_i \underline{e}_i^t$ ao

termo da expressão $\sum_{\substack{j=1 \\ j \neq i}}^p [\lambda_j \underline{e}_j \underline{e}_j^t]$, tem-se que:

$$\sum_{\substack{j=1 \\ j \neq i}}^p [\lambda_j \underline{e}_j \underline{e}_j^t] + \lambda_i \underline{e}_i \underline{e}_i^t - \lambda_i \underline{e}_i \underline{e}_i^t = \Sigma - \lambda_i \underline{e}_i \underline{e}_i^t$$

Utilizando o mesmo raciocínio para $\Sigma^{-1} = \sum_{j=1}^p \left[\frac{1}{\lambda_j} \underline{e}_j \underline{e}_j^t \right]$ somando e

subtraindo ao termo $\sum_{\substack{j=1 \\ j \neq i}}^p \left[\frac{1}{\lambda_j} \underline{e}_j \underline{e}_j^t \right]$ a quantidade dada por $\frac{1}{\lambda_i} \underline{e}_i \underline{e}_i^t$, tem-se:

$$\sum_{\substack{j=1 \\ j \neq i}}^p \left[\frac{1}{\lambda_j} \underline{e}_j \underline{e}_j^t \right] + \frac{1}{\lambda_i} \underline{e}_i \underline{e}_i^t - \frac{1}{\lambda_i} \underline{e}_i \underline{e}_i^t = \Sigma^{-1} - \frac{1}{\lambda_i} \underline{e}_i \underline{e}_i^t$$

Finalmente, o termo $\sum_{\substack{j=1 \\ j \neq i}}^p \left[\underline{e}_j \underline{e}_j^t \right]$ é equivalente a seguinte expressão,

lembrando que os autovetores tem norma 1 e são ortogonais e ainda aplicando-se o mesmo tipo de artifício:

$$\sum_{\substack{j=1 \\ j \neq i}}^p \left[\underline{e}_j \underline{e}_j^t \right] = \mathbf{I} - \underline{e}_i \underline{e}_i^t$$

Assim, retornando ao desenvolvimento anterior da aproximação de qui-quadrado tem-se:

$$\begin{aligned} & \frac{n}{\lambda_i} \underline{e}_0^t \left\{ \sum_{\substack{j=1 \\ j \neq i}}^p \left[\lambda_j \underline{e}_j \underline{e}_j^t \right] - 2\lambda_i \sum_{\substack{j=1 \\ j \neq i}}^p \left[\underline{e}_j \underline{e}_j^t \right] + \lambda_i^2 \sum_{\substack{j=1 \\ j \neq i}}^p \left[\frac{1}{\lambda_j} \underline{e}_j \underline{e}_j^t \right] \right\} \underline{e}_0 = \\ & = \frac{n}{\lambda_i} \underline{e}_0^t \left[\Sigma - \lambda_i \underline{e}_i \underline{e}_i^t - 2\lambda_i (\mathbf{I} - \underline{e}_i \underline{e}_i^t) + \lambda_i^2 \left[\Sigma^{-1} - \frac{1}{\lambda_i} \underline{e}_i \underline{e}_i^t \right] \right] \underline{e}_0 = \\ & = n \left[\frac{\underline{e}_0^t \Sigma \underline{e}_0}{\lambda_i} - \frac{\underline{e}_0^t \underline{e}_i \underline{e}_i^t \underline{e}_0}{\lambda_i} \lambda_i - \frac{\underline{e}_0^t \mathbf{I} \underline{e}_0}{\lambda_i} 2\lambda_i + 2\lambda_i \frac{\underline{e}_0^t \underline{e}_i \underline{e}_i^t \underline{e}_0}{\lambda_i} + \lambda_i^2 \frac{\underline{e}_0^t \Sigma^{-1} \underline{e}_0}{\lambda_i} - \lambda_i^2 \frac{\underline{e}_0^t \underline{e}_i \underline{e}_i^t \underline{e}_0}{\lambda_i \times \lambda_i} \right] = \end{aligned}$$

$$\begin{aligned}
&= n \left[\frac{\underline{e}_0^t \Sigma \underline{e}_0}{\lambda_i} - \underline{e}_0^t \underline{e}_i \underline{e}_i^t \underline{e}_0 - 2 \underline{e}_0^t \underline{e}_0 + 2 \underline{e}_0^t \underline{e}_i \underline{e}_i^t \underline{e}_0 + \lambda_i \underline{e}_0^t \Sigma^{-1} \underline{e}_0 - \underline{e}_0^t \underline{e}_i \underline{e}_i^t \underline{e}_0 \right] = \\
&= n \left[\frac{\underline{e}_0^t \Sigma \underline{e}_0}{\lambda_i} + \lambda_i \underline{e}_0^t \Sigma^{-1} \underline{e}_0 - 2 \right]
\end{aligned}$$

Substituindo nessa última expressão Σ pelo estimador S , a distribuição ainda continua aproximadamente de qui-quadrado para grandes amostras. Dessa forma, a prova fica completa.

Um outro importante teste de interesse é o da hipótese de mesma estrutura de correlação, ou seja, $\text{Cov}(X_i, X_k) = \sqrt{\sigma_{ii} \sigma_{kk}} \rho$ ou $\text{Corr}(X_i, X_k) = \rho$, para todo $i \neq k$. Nesse caso, os autovalores de Σ não são distintos e os resultados anteriores não se aplicam. Embora as distribuições amostrais dos componentes principais obtidos da matriz R sejam difíceis de derivar, esse caso especial conduz a resultados tratáveis (Morrison, 1976).

O teste proposto por Lawley (1963) de igualdade dos últimos $(p-1)$ valores característicos da matriz de correlação populacional ρ ($p \times p$) é equivalente ao teste da hipótese de igualdade de todas as $p(p-1)/2$ correlações (ρ_{ij}) . A hipótese de interesse é dada por:

$$H_0 : \rho = \rho_0 = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & \rho & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix} \quad \text{Vs} \quad H_0 : \rho \neq \rho_0 \quad (7.47)$$

Essa hipótese pode ser escrita na forma equivalente $H_0 : \rho_{ij} = \rho$ para todos os subscritos $i \neq j$. O procedimento de Lawley (1963) requer as seguintes quantidades:

$$\bar{r}_k = \frac{1}{p-1} \sum_{\substack{i=1 \\ i \neq k}}^p r_{ik}; \quad k = 1, 2, \dots, p \quad (7.48)$$

$$\bar{r} = \frac{2}{p(p-1)} \sum_{i=1}^{p-1} \sum_{k=i+1}^p r_{ik} \quad (7.49)$$

$$\hat{\gamma} = \frac{(p-1)^2 [1 - (1 - \bar{r})^2]}{p - (p-2)(1 - \bar{r})^2} \quad (7.50)$$

Verifica-se facilmente que \bar{r}_k de (7.48) é a média dos elementos fora da diagonal para as k colunas de R e \bar{r} de (7.49) é a média de todos os elementos fora da diagonal principal de R . Lawley (1963) mostrou que quando n tende para infinito o teste estatístico:

$$\chi_c^2 = \frac{n-1}{(1-\bar{r})^2} \left[\sum_{i=1}^{p-1} \sum_{k=i+1}^p (r_{ik} - \bar{r})^2 - \hat{\gamma} \sum_{k=1}^p (\bar{r}_k - \bar{r})^2 \right] \quad (7.51)$$

tem distribuição de qui-quadrado com $\nu = (p+1)(p-2)/2$ graus de liberdade.

Finalmente, o teste, denominado de teste de esfericidade, é apresentado. Para isso é necessário obter a seguinte estatística.

$$\Lambda_1 = \frac{|S|^{n/2}}{[\text{tr}(S)/p]^{np/2}} = \frac{\left(\sqrt[p]{\prod_{i=1}^p \hat{\lambda}_i}\right)^{np/2}}{\left(\sum_{i=1}^p \hat{\lambda}_i / p\right)^{np/2}} = \frac{\left(\sqrt[p]{\prod_{i=1}^p \hat{\lambda}_i}\right)^{np/2}}{\left(\sum_{i=1}^p \hat{\lambda}_i / p\right)^{np/2}} \quad (7.52)$$

Bartlett (1954) sugere que para grandes amostras a estatística dada por:

$$\chi_c^2 = -2 \left[1 - \frac{(2p^2 + p + 2)}{6pn} \right] \ln(\Lambda_1) \quad (7.53)$$

tem distribuição aproximadamente de qui-quadrado com $v=(p+1)(p-2)/2$ graus de liberdade sob H_0 dada em (7.54).

$$H_0 : \Sigma = \sigma^2 I \quad (7.54)$$

O teste (7.53) da hipótese nula (7.54) é denominado de teste de esfericidade, porque os contornos da densidade são esferas quando $\Sigma = \sigma^2 I$.

Um teste mais geral do que o teste (7.53) para a hipótese de que todas as variáveis sejam independentes é dado pelo teste de razão de verossimilhança. Seja a hipótese

$$H_0 : \Sigma = \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{pp} \end{bmatrix}; \quad \sigma_{ii} > 0 \quad (7.55)$$

A seguinte estatística deve ser calculada inicialmente:

$$\Lambda_2 = \frac{|S|^{n/2}}{\prod_{i=1}^p S_{ii}^{n/2}} = |R|^{n/2} \quad (7.56)$$

Para grandes amostras, sob H_0 , o teste estatístico:

$$\chi_c^2 = -2 \left[1 - \frac{(2p+11)}{6n} \right] \ln(\Lambda_2) \quad (7.57)$$

tem distribuição aproximadamente de qui-quadrado com $\nu=p(p-1)/2$ graus de liberdade sob H_0 dada em (7.55). Essa aproximação é devida a Bartlett (1954) em substituição a aproximação usual $-2\ln(\Lambda_2)$. O resultado (7.57) melhora a aproximação qui-quadrado usual.

Lawley (1940) mostra que o teste (7.57) pode ser aproximado por:

$$\chi_c^2 \cong \left[n - \frac{(2p+11)}{6} \right] \sum_{i=1}^{p-1} \sum_{k=i+1}^p r_{ik}^2 \quad (7.58)$$

Essa expressão representa uma melhor aproximação de (7.57) para pequenas correlações e para grandes amostras é pouco provável que conduza a diferentes resultados dos obtidos pela fórmula determinantal exata (7.57), Morrison (1976).

É apresentado a seguir um programa SAS no procedimento de matrizes IML para a realização de todas as inferências propostas nessa seção.

Um exemplo com comentários é apresentado para que o usuário possa reproduzir os testes e os procedimentos de estimação propostos.

```

options ps=5000 ls=75 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;
  print 'Valor de p tamanho da amostra e alpha';
  print p n alpha;
  print 'Matriz de covariancias amostral: S';
  print S;
  Ls=diag(eigval(s));
  Ps=eigvec(S);
  print 'Matriz de autovalores de S';
  print Ls;
  print 'Matriz de autovetores de S';
  print Ps;
  D=diag(S);
  D_12=inv(root(D));
  *print D_12;
  Rs=D_12*S*D_12;
  print 'Matriz de correlacoes amostrais R';
  print Rs;
  Lr=diag(eigval(Rs));
  print 'Matriz de autovalores de R';
  print Lr;
  Pr=eigvec(Rs);
  print 'Matriz de autovetores de R';
  print Pr;
  /*intervalo de confianca para autovalores de S - equacao 7.41*/
  za2=probit(1-alpha/2);
  print 'Intervalos de confianca para os autovalores de S, sendo 1-
alpha=' alpha;
  print 'Autovalor      Li      Ls';
  do i=1 to p;
    lin=ls[i,i]/(1+za2*(2/n)**0.5);
    lsu=ls[i,i]/(1-za2*(2/n)**0.5);
    print i  lin  lsu;
  end;
  /*Testar a hipotese de que o maior autovalor de S e igual a 10=12.35 -
equacao 7.42 */
  /* este teste eh motivado pelo fato de l1=sig2(1+(p-1)rho), com
sig2=4.2 e rho=0.97 */
  l0=12.35;
  Zc=(ls[1,1]-l0)/(l0*(n/2))**0.5;
  przc=2*(1-probnorm(abs(zc)));
  print 'Teste de H0: l1=12.35 (igual correlacao). Esse valor eh apenas
um exemplo';
  print 'Valor de Zc valor de prob>|zc|';
  print 'Se [prob>|zc|]>valor de alpha Ho nao deve ser rejeitada';
  print Zc przc;
  /* teste 7.43 igualdade de r autovalores intermediarios*/

```

```

/* neste exemplo sera testado Ho: l2 = l3 */
/*q=1, r=2, p=3 -teste 7.44 */
aux1=0;aux2=0;q=1;r=2;
do i=q+1 to q+r;
  aux1=aux1+log(ls[i,i]);
  aux2=aux2+ls[i,i]/r;
end;
qui2c=-(n-1)*aux1+(n-1)*r*log(aux2);
print 'Valores dos somatorios auxiliares para teste H0: l2 = l3';
print 'aux1 = soma ln(lj) e aux2 = media dos lj intermediarios';
print aux1 aux2;
v=r*(r+1)/2-1;
prqui2c=1-probchi(qui2c,v);
print 'Teste da hipotese de que Ho: l2 = l3 ';
print 'Qui-quadrado GL Pr>qui-Quadr';
print qui2c v prqui2c;
/* teste para a hipotese de igualdade de um autovetor a um vator de
constantes*/
/* Para ilustrar sera testado que e1=[1/3^0.5 1/3^0.5 1/3^0.5], ou
seja, igual*/
/* estrutura de correlacao da matriz Sigma que originou a S
*/

e0=j(p,1,1/3**0.5);
E1=j(p,p,0);
do i=1 to p;
  ek=Ps[,i];
  if i^=1 then
  do;
    E1=E1+(ls[i,i]/(ls[i,i]-ls[1,1])**2)*ek*t(ek);
  end;
end;
E1=ls[1,1]*E1;
Le=eigval(e1);
*print E1 le;
eil=Ps[,1];
print e0 eil;
qui2c=n*(ls[1,1]*e0`*inv(S)*e0+e0`*S*e0/ls[1,1]-2);
qui2c2=n*t(Ps[,1]-e0)*ginv(E1)*(Ps[,1]-e0);
v=p-1;
prqui2c=1-probchi(qui2c,v);
print 'Teste da hipotes e1=e0=t([1/3^0.5 1/3^0.5 1/3^0.5])';
print 'Qui-quadrado1 qui-quad2 GL Pr>qui-Quadr';
print qui2c qui2c2 v prqui2c;
/*teste da H0:phoij=pho-igual estrutura de correlacao */
rbar=(sum(Rs)-trace(Rs))/(p*(p-1));
rk=j(p,1,0);
do i=1 to p;
  rk[i]=(sum(Rs[,i])-1)/(p-1);
end;
gama=(p-1)**2*(1-(1-rbar)**2)/(p-(p-2)*(1-rbar)**2);
aux1=(Rs-j(p,p,rbar))#(Rs-j(p,p,rbar));
aux2=(sum(aux1)-trace(aux1))/2;
aux3=(rk-j(p,1,rbar))#(rk-j(p,1,rbar));
aux4=sum(aux3);
qui2c=(n-1)/(1-rbar)**2*(aux2-gama*aux4);
v=(p+1)*(p-2)/2;
if qui2c<=0 then qui2c=1e-14;
prqui2=1-probchi(qui2c,v);
print 'Teste da hipotes phij=pho: igual estrutura de correlacao';
print 'Qui-quadrado GL Pr>qui-Quadr';

```



```

print qui2c v prqui2;
print 'Valores utilizados no teste-para simples conferencia';
print 'media geral dos rij, vetor de medias de cada coluna de R e gama
chapeu';
print rbar rk gama;
/*teste de esfericidade-H0: Sigma=Sig^2*I*/
Lamb1=((det(S)**(1/p))/(trace(S)/p));
qui2c=-2*(n*p/2)*log(lamb1)*(1-(2*p**2+p+2)/(6*p*n));
v=(p+1)*(p-2)/2;
prqui2=1-probchi(qui2c,v);
print 'Teste de esfericidade - H0: Sigma=Sig^2*I';
print 'Qui-quadrado GL          Pr>qui-Quadr   Lambida 1^(2/(np))';
print qui2c v prqui2 lamb1;
/*teste de independencia de variaveis mais geral - H0: Sigma =
Diag(sig11 sig22 ... sigpp)*/
Lamb2=det(Rs);
qui2c=-2*(n/2)*log(lamb2)*(1-(2*p+1)/(6*n));
v=(p+1)*(p-2)/2;
prqui2=1-probchi(qui2c,v);
print 'Teste de independencia - H0: Sigma = Diag(sig11 sig22 ...
sigpp)';
print 'Qui-quadrado GL          Pr>qui-Quadr   Lambida 2^2/n';
print qui2c v prqui2 lamb2;
/*teste de independencia de variaveis - uso da aproximacao de Lawley-
pior*/
aux1=Rs#Rs;
aux2=(sum(aux1)-trace(aux1))/2;
qui2c=aux2*(n-(2*p+1))/6;
v=(p+1)*(p-2)/2;
prqui2=1-probchi(qui2c,v);
print 'Teste de independencia aproximado de Lawley (1940)';
print 'para a hipotese H0: Sigma = Diag(sig11 sig22 ... sigpp)';
print 'Qui-quadrado GL          Pr>qui-Quadr   Soma de rij^2=aux2';
print 'Obs. para grandes valores de rij essa eh uma pessima
aproximacao';
print qui2c v prqui2 aux2;
quit;

```

7.6. Exercícios

7.6.1. Extrair os componentes principais da matriz S obtida das mensurações de três variáveis em carapaças de tartarugas. As variáveis X_1 , X_2 , e X_3 são referentes ao comprimento, largura e altura transformadas por logaritmo natural, respectivamente. Uma amostra de 24 fêmeas foi realizada. A matriz S é apresentada a seguir, juntamente com o vetor de médias das variáveis

transformadas. Obter os componentes principais de S e interpretá-los, quando for possível. Obter a matriz R e os respectivos componentes principais. Obter em ambos os casos: a) a porcentagem de informação explicada por cada componente; b) a correlação entre as variáveis originais transformadas e os componentes principais. Observando o primeiro componente principal de R com mais profundidade, o que pode ser afirmado sobre a matriz R (sem a realização de teste).

$$\bar{\underline{X}} = \begin{bmatrix} 2,128 \\ 2,008 \\ 1,710 \end{bmatrix} \quad \text{e} \quad S = \begin{bmatrix} 4,9810 & 3,8063 & 4,7740 \\ 3,8063 & 3,0680 & 3,7183 \\ 4,7740 & 3,7183 & 4,8264 \end{bmatrix}$$

- 7.6.2. Com os dados do exercício 7.6.1, determine os intervalos de 95% de confiança assintótico para os 3 autovalores de Σ (3x3).
- 7.6.3. Com os dados do exercício 7.6.1 teste a hipótese de que o primeiro autovetor de ρ seja igual a $\underline{e}_1^t = [1/\sqrt{3} \quad 1/\sqrt{3} \quad 1/\sqrt{3}]$. Qual é sua conclusão com relação à decisão tomada?
- 7.6.4. Com os dados do exercício 7.6.1 reproduza a matriz S a partir do primeiro componente principal e a matriz de resíduos.
- 7.6.5. Teste a hipótese de que os $r=2$ últimos valores característicos de Σ , sejam iguais, utilizando os dados do exemplo 7.6.1.
- 7.6.6. Teste a hipótese de independência geral entre 3 variáveis, para as quais uma amostra de $n=50$ observações apresentou a seguinte matriz de covariância.

$$S = \begin{bmatrix} 24,9811 & 0,0796 & 0,0574 \\ 0,0796 & 5,2762 & 0,0020 \\ 0,0574 & 0,0020 & 3,0655 \end{bmatrix}$$

7.6.7. Os dados a seguir referem a uma amostra de 30 elementos em uma população normal trivariada. Obtenha os componentes principais e verifique a normalidade por meio dos dois primeiros componentes. Faça os Q-Q plots e os gráficos de dispersão dos escores do componente 1 vs 2. Utilize o último componente para verificar a possibilidade de observações suspeitas. Caso alguma observação suspeita seja observada, elimine-a da amostra e refaça o exercício.

U.A.	X_1	X_2	X_3
1	12,80	29,56	45,19
2	14,12	26,54	49,29
3	19,09	33,26	49,79
4	15,98	31,00	51,73
5	16,00	28,94	50,30
6	16,51	31,67	48,06
7	14,05	30,11	55,15
8	14,34	26,47	46,84
9	16,87	29,00	52,16
10	21,93	38,00	39,24
11	15,21	30,68	54,02
12	15,54	27,37	51,52
13	17,71	30,20	51,66
14	14,42	29,99	52,50
15	13,38	31,61	52,33
16	13,91	29,59	44,19
17	15,53	29,30	53,71
18	16,40	28,96	46,56
19	18,35	30,15	52,18
20	13,59	27,70	52,33
21	19,08	31,26	48,59
22	13,95	29,94	54,73

23	16,11	34,52	52,69
24	17,10	29,39	52,03
25	18,81	31,48	49,79
26	15,27	29,54	43,11
27	14,80	31,88	48,08
28	17,39	28,88	50,69
29	18,02	34,02	49,58
30	9,52	25,23	45,89

7.7. Referências

ANDERSON, T.W. The asymptotic theory for principal components analysis,

Annals of Mathematical Statistics, v.34, p.122-148, 1963.

BARTLETT, M.S. A note on the multiply factors for various chi-square approximation. **Journal of the Royal Statistical Society**, Series B. v.16, p.296-298, 1954.

BOCK, R.D. **Multivariate statistical methods in behavioral research**. McGraw Hill, 1975.

GIRSHICK, M.A. On the sampling theory of roots of determinantal equations. **Annals of Mathematical Statistics**. v.10, p.203-224, 1939.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. 4th edition. Prentice Hall, New Jersey, 1998. 816p.

LAWLEY, D.N. The estimation of factor loadings by the method of maximum likelihood. **Proceedings of the Royal Society of Edinburgh**, v.40, p.64-82, 1940.

LAWLEY, D.N. Tests of significance for the latent roots of covariance and correlation matrices. **Biometrika**, v.43, p.128-136, 1956.

LAWLEY, D.N. On testing a set of correlation coefficients for equality, **Annals of Mathematical Statistics**, v.34, p.149-151, 1963.

MORRISON, D.F. **Multivariate statistical methods**. New York: McGraw-Hill, 2d ed., 1976. 307p.