

**Disciplina: Análise Multivariada I**  
**Prof. Dr. Admir Antonio Betarelli Junior**

**AULA 2**

**1 ANÁLISE DE COMPONENTES PRINCIPAIS (ACP)**

Essa técnica expressa a estrutura de variância e covariância por meio de (poucas) combinações lineares das variáveis originais em estudo. Essas combinações lineares são chamadas de componentes principais, não são correlacionados entre si, e são ordenados pela parcela da informação total que cada um retém. Geralmente é útil como um passo intermediário em econometria, por exemplo.

**1.1 Objetivos principais**

Tem por principais objetivos:

- a) redução dos dados: busca-se indicar variáveis importantes e criar um conjunto de novas variáveis, muito menor, para substituir parcial ou completamente o conjunto original de variáveis, mantendo ainda a maior parte das informações da amostra;
- b) interpretação de relações: fornece uma clara compreensão sobre quais variáveis estão juntas e quantas variáveis podem ser consideradas.

**1.2 Quando usar a técnica**

Deve-se usar essa técnica quando a preocupação principal é o número mínimo de novas variáveis (componentes) necessárias para expressar uma grande parte da variância total das variáveis originais. Como regra básica procure selecionar um número mínimo que atinge pelo menos **70% da variância total** da amostra. A eficiência do método está relacionada à correlação entre as variáveis originais, seja ela positiva ou negativa. Recomenda-se que os valores das correlações sejam maiores que **0,30**, caso contrário, a técnica é provavelmente inapropriada. No caso de haver muitas observações com dados

omissos em algumas das variáveis, o usuário deverá avaliar até que ponto suas análises são válidas. Alguns softwares, automaticamente, eliminam o elemento amostral com dados omissos da execução dos cálculos correspondentes.

## 2 DESCRIÇÃO E ÁLGEBRA

Em geral, a variabilidade inicial do sistema, determinado por  $p$  variáveis pode ser reproduzido por  $p$  componentes principais. Contudo, dependendo da correlação entre as variáveis originais, um número  $k$  menor de componentes,  $k \leq p$ , pode reproduzir grande parte da variabilidade original. Sua definição não requer que seja assumido um modelo a priori, não há uma hipótese a ser testada, não é necessário postular causalidades (apenas correlações) e não requer a adoção da hipótese de distribuição normal multivariada. Entretanto, componentes principais derivados de uma população normal ajudam na interpretação das elipsóides de densidade constante.

A escolha entre matriz de covariância ou correlação deve levar em conta a escala das variáveis originais.

### 2.1 Álgebra geral

#### 2.1.1 Dados populacionais

Considerando que a matriz  $\Sigma_{p \times p}$  possui os pares de autovalores e autovetores,

$(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$ , tal que  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ , e o vetor aleatório é

$\mathbf{X}' = [X_1, X_2, \dots, X_p]$ , então os componentes principais são:

$$Y_1 = \mathbf{a}'_1 \mathbf{X} = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p$$

$$Y_2 = \mathbf{a}'_2 \mathbf{X} = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p$$

$\vdots$

$$Y_p = \mathbf{a}'_p \mathbf{X} = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p$$

$$\Rightarrow E(\mathbf{a}'_i \mathbf{X}) = \mathbf{a}'_i \boldsymbol{\mu}, \quad Var(\mathbf{a}'_i \mathbf{X}) = \mathbf{a}'_i \Sigma \mathbf{a}_i, \quad Cov(Y_i, Y_k) = \mathbf{a}'_i \Sigma \mathbf{a}_k \quad \forall i, k = 1, 2, \dots, p.$$

Os componentes principais são combinações lineares não correlacionadas (ortogonais),  $Y_1, Y_2, \dots, Y_p$ , cujas variâncias são as maiores possíveis. Como  $Var(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\Sigma\mathbf{a}$  pode ser aumentada quando se multiplica por  $\mathbf{a}'\mathbf{a}_i > 1$ , então o vetor  $\mathbf{a}_i$  terá comprimento unitário ( $\mathbf{a}'_i\mathbf{a}_i = 1$ ). E como os componentes são ortogonais, então  $\mathbf{a}'_i\mathbf{a}_k = 0 = Cov(Y_i, Y_k) = \mathbf{a}'_i\Sigma\mathbf{a}_k$ . Assim, pela maximização de forma quadrática com essas restrições (veja seção 2.6.1 da aula 1), define-se:

1º componente:  $Y_1 = \mathbf{a}'_1\mathbf{X} \Rightarrow Max Var(\mathbf{a}'_1\mathbf{X}) \text{ s.a. } \mathbf{a}'_1\mathbf{a}_1 = 1$

$$L = \mathbf{a}'_1\Sigma\mathbf{a}_1 - \lambda_1(\mathbf{a}'_1\mathbf{a}_1 - 1) \Rightarrow C.P.O.: (\Sigma - \lambda_1 I)\mathbf{a}_1 = 0 \Rightarrow \mathbf{a}'_1\Sigma\mathbf{a}_1 = \lambda_1 = Var(Y_1)$$

2º componente:  $Y_2 = \mathbf{a}'_2\mathbf{X} \Rightarrow Max Var(\mathbf{a}'_2\mathbf{X}) \text{ s.a. } \mathbf{a}'_2\mathbf{a}_2 = 1, Cov(Y_1, Y_2) = 0$

$$L = \mathbf{a}'_2\Sigma\mathbf{a}_2 - \lambda_2(\mathbf{a}'_2\mathbf{a}_2 - 1) - \underbrace{\phi(\mathbf{a}'_1\Sigma\mathbf{a}_2)}_{=0} \Rightarrow C.P.O.: \Rightarrow \mathbf{a}'_2\Sigma\mathbf{a}_2 = \lambda_2 = Var(Y_2)$$

í-ésimo componente:  $Y_i = \mathbf{a}'_i\mathbf{X} \Rightarrow Max Var(\mathbf{a}'_i\mathbf{X}) \text{ s.a. } \mathbf{a}'_i\mathbf{a}_i = 1, Cov(Y_i, Y_k) = 0$

$$\mathbf{a}'_i\Sigma\mathbf{a}_i = \lambda_i = Var(Y_i), \quad k < i$$

Ou também, a maximização da forma quadrática se transforma na razão:

$$\max_{\mathbf{a} \neq 0} \frac{\mathbf{a}'\Sigma\mathbf{a}}{\mathbf{a}'\mathbf{a}} = \lambda_i = \frac{\mathbf{e}'_i\Sigma\mathbf{e}_i}{\mathbf{e}'_i\mathbf{e}_i} = \mathbf{e}'_i\Sigma\mathbf{e}_i = Var(Y_i) \quad \because \mathbf{a}_i = \mathbf{e}_i, \mathbf{e}'_i\mathbf{e}_i = 1$$

$$COV(Y_i, Y_k) = \mathbf{e}'_i\Sigma\mathbf{e}_k = \mathbf{e}'_i\lambda_k\mathbf{e}_k = \lambda_k\mathbf{e}'_i\mathbf{e}_k = 0 \quad \because \Sigma\mathbf{e}_k = \lambda_k\mathbf{e}_k, \mathbf{e}'_i\mathbf{e}_k = 0$$

Portanto,  $\mathbf{a}_1 = \mathbf{e}_1, \mathbf{a}_2 = \mathbf{e}_2, \dots, \mathbf{a}_p = \mathbf{e}_p$ , tal que  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ . Isto significa o valor de  $\lambda_1$  é o maior possível de  $\Sigma$ , ou seja, o componente  $Y_1$  retém a maior parcela de variância possível de  $\Sigma$ , seguindo similarmente de  $Y_2$  até  $Y_n$ . Pelo resultado do teorema espectral (aula 1) e pelas soluções ótimas acima, então a variância total é:

$$\sum_{k=1}^p Var(X_k) = traço(\Sigma) = traço(\Lambda) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p Var(Y_i)$$

ou seja, o somatório das variâncias do sistema original de dados é o mesmo que o somatório da variância dos componentes principais. A proporção do *i*-ésimo componente mais representativo é definida como:

$$\%Var(Y_i) = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \times 100 = \frac{\lambda_i}{\text{traço}(\Lambda)} \times 100$$

Por fim, pode-se calcular o coeficiente de correlação entre  $Y_i = \mathbf{e}'_i \mathbf{X}$  e  $X_k$ :

$$\rho_{Y_i, X_k} = \frac{e_{ik} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}$$

Demonstração: fixando  $\mathbf{a}'_k = [0, \dots, 0, 1, 0, \dots, 0]$  de maneira que  $X_k = \mathbf{a}'_k \mathbf{X}$  e

$COV(X_k, Y_i) = COV(\mathbf{a}'_k \mathbf{X}, \mathbf{e}'_i \mathbf{X}) = \mathbf{a}'_k \Sigma \mathbf{e}_i$ . Desde que  $\Sigma \mathbf{e}_i = \lambda_i \mathbf{e}_i$ , então

$COV(X_k, Y_i) = \mathbf{a}'_k \lambda_i \mathbf{e}_i = \lambda_i e_{ik}$ . Assim,

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

Veja o exemplo 8.1 em Johnson e Wichern (2007, p.430-431).

$$\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad \left\{ \begin{array}{l} \lambda_1 = 5.83 \Rightarrow \mathbf{e}'_1 = [0.383, -0.924, 0] \\ \lambda_2 = 2.00 \Rightarrow \mathbf{e}'_2 = [0, 0, 1] \\ \lambda_3 = 0.17 \Rightarrow \mathbf{e}'_3 = [0.924, 0.383, 1] \end{array} \right. \quad \left\{ \begin{array}{l} Y_1 = \mathbf{a}'_1 \mathbf{X} = 0.383X_1 - 0.924X_2 \\ Y_2 = \mathbf{a}'_2 \mathbf{X} = X_3 \\ Y_3 = \mathbf{a}'_3 \mathbf{X} = 0.924X_1 + 0.383X_2 \end{array} \right.$$

$$\%Var(Y_1) = \frac{5.83}{8} \times 100 = 73\%$$

$$\rho_{Y_1, X_1} = \frac{e_{11} \sqrt{\lambda_1}}{\sqrt{\sigma_{11}}} = \frac{0.383 \sqrt{5.83}}{\sqrt{1}} = 0.925, \quad \rho_{Y_2, X_2} = -0.998$$

Isto sugere que relativamente  $X_2$  contribui mais para a determinação de  $Y_1$ , pois a mesma tem o maior peso com maior correlação.

### 2.1.2 Dados amostrais

Uma vez que os dados são amostras coletadas, utiliza-se na álgebra acima a matriz de covariâncias amostral,  $\mathbf{S}_{p \times p}$ , com seus pares de autovalores e autovetores correspondentes,  $(\hat{\lambda}_1, \hat{\mathbf{e}}_1), (\hat{\lambda}_2, \hat{\mathbf{e}}_2), \dots, (\hat{\lambda}_p, \hat{\mathbf{e}}_p)$ , tal que  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p > 0$ . O  $i$ -ésimo componente principal amostra será então:

$$\hat{Y}_i = \hat{\mathbf{a}}_i' \mathbf{X} = \hat{a}_{i1} X_1 + \hat{a}_{i2} X_2 + \dots + \hat{a}_{ip} X_p$$

$$\Rightarrow E(\mathbf{a}_i' \mathbf{X}) = \mathbf{a}_i' \bar{X}, \quad Var(\hat{\mathbf{a}}_i' \mathbf{X}) = \hat{\mathbf{a}}_i' \hat{\mathbf{S}} \hat{\mathbf{a}}_i, \quad Cov(\hat{Y}_i, \hat{Y}_k) = \hat{\mathbf{a}}_i' \hat{\mathbf{S}} \hat{\mathbf{a}}_k \quad \forall i, k = 1, 2, \dots, p.$$

em que  $\mathbf{a}_i = \mathbf{e}_i, \mathbf{e}_i' \mathbf{e}_i = 1, \mathbf{e}_i' \mathbf{e}_k = 0$ . Da mesma forma, o ordenamento dos componentes é definido pela maximização das variâncias das  $p$  combinações lineares, com os seguintes resultados:

$$Var(\hat{Y}_i) = \hat{\lambda}_i, \quad Cov(\hat{Y}_i, \hat{Y}_k) = 0 \quad (i \neq k)$$

$$\sum_{i=1}^p Var(\hat{Y}_i) = traço(\tilde{\mathbf{\Lambda}}) = traço(S) = \sum_{k=1}^p Var(X_k), \quad r_{Y_i, X_k} = \frac{\hat{e}_{ik} \sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}}$$

### 2.2 Variáveis padronizadas

Os componentes são influenciados pela escala das variáveis, justamente porque as matrizes de covariâncias,  $\mathbf{\Sigma}_{p \times p}$  ou  $\mathbf{S}_{p \times p}$ , são sensíveis à escala de um par de variáveis. Esse problema pode ser amenizado calculando-se os autovalores e autovetores das matrizes de correlação:  $\mathbf{P}_{p \times p}$  (populacional) e  $\mathbf{R}_{p \times p}$  (amostral). Na prática, basta padronizar as variáveis originais antes de computar os componentes principais, isso porque a própria matriz de covariâncias das variáveis padronizadas é a matriz de correlação das variáveis originais.

Desse modo, por exemplo, o  $i$ -ésimo componente principal da matriz  $\mathbf{P}_{p \times p}$  é definido por:

$$Y_i = \mathbf{a}'_i \mathbf{Z} = a_{i1}Z_1 + a_{i2}Z_2 + \dots + a_{ip}Z_p$$

sendo

$$Var(Y_i) = \lambda_i, \quad Cov(Y_i, Y_k) = 0 \quad (i \neq k)$$

$$\sum_{i=1}^p Var(Y_i) = \text{traço}(\mathbf{\Lambda}) = \text{traço}(\mathbf{P}) = \sum_{k=1}^p Var(Z_k), \quad r_{Y_i, Z_k} = e_{ik} \sqrt{\lambda_i}$$

$$\text{em que } Z_k = \frac{(X_k - \mu_k)}{\sqrt{\sigma_{kk}}}$$

Exemplo 8.2 em Johnson e Wichern (2002, p.433-434):

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 4 \\ 4 & 100 \end{bmatrix} \quad \lambda_1 = 100.16 \Rightarrow \mathbf{e}'_1 = [0.40, 0.999] \quad \begin{cases} Y_1 = 0.40X_1 + 0.999X_2 \\ Y_2 = 0.999X_1 - 0.040X_2 \end{cases}$$

$$\mathbf{P} = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix} \quad \lambda_1 = 1 + \rho = 1.40 \Rightarrow \mathbf{e}'_1 = [0.707, 0.707] \quad \begin{cases} Y_1 = 0.707Z_1 + 0.707Z_2 \\ Y_2 = 0.707Z_1 - 0.707Z_2 \end{cases}$$

sendo:

$$Y_1 = 0.707 \left( \frac{X_1 - \mu_1}{1} \right) + 0.707 \left( \frac{X_2 - \mu_2}{10} \right) = 0.707(X_1 - \mu_1) + 0.0707(X_2 - \mu_2)$$

Conclusões:

- os componentes de  $\mathbf{P}$  são mais equilibrados que os de  $\mathbf{\Sigma}$ ;
- com a padronização, todas as variáveis tem variância unitária, não havendo, portanto, dominância direta de nenhuma delas;
- os pesos são diferentes.

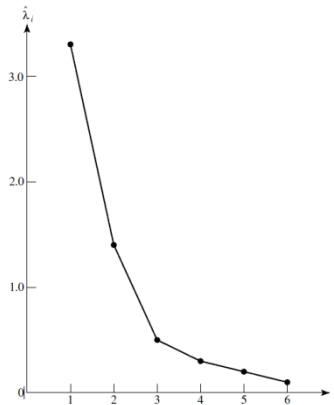
### 3 NÚMERO DE COMPONENTES

Existe sempre a questão de quantos componentes reter. Não existe uma resposta definitiva. Entretanto, existem alguns critérios para a seleção do número de componentes.

#### I. Representatividade:

- Screeplot (cotovelos): com uma porcentagem de 70% ou mais para os primeiros componentes, seleciona-os sem perda demasiada de informação;

- b) **Kaiser**: o último componente selecionado é aquele de igual ou superior à média dos demais autovalores.
- c) **Jolliffe ( $\Delta$ Kaiser)**: o menor autovalor deve ser 0.70 vezes à média dos demais.



## II. Análise prática

- a) **Sejam passíveis de interpretação**: como se pode utilizar um componente como variável resposta numa regressão, se não sabe o que ela significa?
- b) **Situação ideal**: é aquela em que os componentes principais de maior interesse do pesquisador são os de maior expressão relativa à variância total, que represente um valor pequeno de  $k$ .

## III. Representatividade versus interpretação

- a) Se poucos componentes são selecionados, a estrutura não é revelada e dimensões importantes podem ser omitidas.
- b) Se muitos componentes são mantidos, a interpretação se torna difícil.

O pesquisador deve ter um conjunto de componentes mais representativos e parcimonioso possível.

*Comentário*: um valor pequeno para os últimos autovalores de uma covariância amostral ou matriz de correlação pode indicar a suspeita de dependência linear do conjunto de dados. Se isto ocorre, uma (ou mais) variável é redundante e a mesma deve ser excluída. Considere, por exemplo,  $x_1, x_2$  e  $x_3$  como sendo subescores e  $x_4$  o total

dos escores:  $x_1 + x_2 + x_3$ . Assim, embora a combinação linear  $\mathbf{e}'\mathbf{x} = [1,1,1,-1]\mathbf{x} = x_1 + x_2 + x_3 - x_4$  é sempre zero, erro de arredondamento no cálculo dos autovalores no cálculo de autovalores pode resultar em pequenos valores. Se a expressão linear relacionando  $x_4$  com  $(x_1, x_2 \text{ e } x_3)$  passou despercebida, os menores pares de autovalores-autovetores fornecem um pista de sua existência. Desse modo, embora os “grandes” autovalores e os correspondentes autovetores sejam importantes na análise de componentes principais, autovalores próximo de zero não devem ser ignorados. Autovalores associados com os últimos autovalores pode destacar dependências lineares nos dados e causar problemas de interpretação nas análises subsequentes.

## 4 INTERPRETAÇÃO

### 4.1 Pesos das variáveis (*loading*)

A interpretação dos componentes formados pode ser feita com base nos coeficientes do vetor  $\mathbf{a}_k$  atribuídos a cada variável, ou melhor, sobre os pesos das variáveis. Os pesos das variáveis (*loading*) correspondem à carga ou importância de cada variável para o valor de cada componente principal. As variáveis mais importantes são as de maiores pesos, negativos ou positivos. O sinal indica se a correlação é positiva ou negativa.

Embora as correlações das variáveis com os componentes principais ajudem na interpretação dos componentes, cada uma somente mede a contribuição univariada de um  $X$  individual em um componente. Ou melhor, eles não indicam a importância de  $X$  em  $Y$  na presença de outros  $X$ 's. Por essa razão, alguns estatísticos (e.g., Rensvold (1992)), recomendam somente o uso dos coeficientes  $\mathbf{e}_{ik}$  e não as correlações para serem usados nas interpretações dos componentes. Embora os coeficientes e as correlações possam levar para diferentes *rankings* em termos de importância das variáveis para cada componente, a experiência aponta que esses rankings geralmente não são diferentes. Na prática, variâncias com coeficientes ( $\mathbf{e}_{ik}$ ) relativamente grandes (em valor absoluto) tende a ter correlações relativamente grandes. Assim, coeficientes



são úteis para um contexto multivariado e as correlações para uma situação univariada, cujos valores fornecem resultados similares. Recomenda-se que os coeficientes e as correlações sejam examinados para ajudar na interpretação dos componentes.

## 4.2 Interpretação geométrica

Supondo que  $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , por exemplo, os componentes principais têm uma atrativa interpretação. Sabe-se que a densidade de  $\mathbf{X}$  é constante da elipsóide centrada em  $\boldsymbol{\mu}$ :

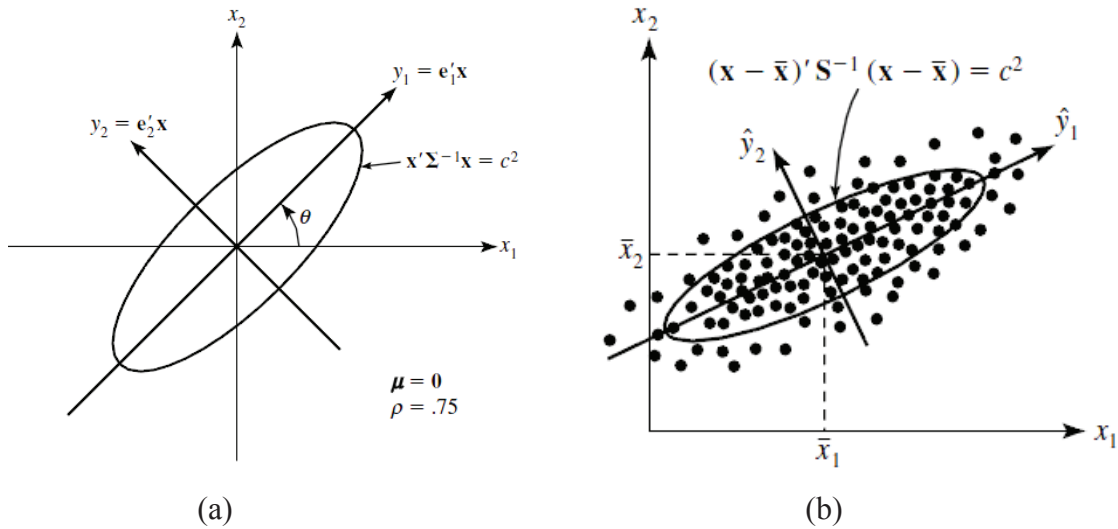
$$c^2 = (\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) = \chi_p^2(\alpha)$$

em que os eixos são dados por  $\pm c \lambda_k^{1/2} \mathbf{e}_k$ . Considerando o teorema espectral e tomando  $\boldsymbol{\mu} = \mathbf{0}$ , tem-se que:

$$\begin{aligned} \boldsymbol{\Sigma}^{-1} &= \lambda_1^{-1} \mathbf{e}_1 \mathbf{e}_1' + \lambda_2^{-1} \mathbf{e}_2 \mathbf{e}_2' + \dots + \lambda_p^{-1} \mathbf{e}_p \mathbf{e}_p' \Rightarrow \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} = \mathbf{X}' (\lambda_1^{-1} \mathbf{e}_1 \mathbf{e}_1' + \lambda_2^{-1} \mathbf{e}_2 \mathbf{e}_2' + \dots + \lambda_p^{-1} \mathbf{e}_p \mathbf{e}_p') \mathbf{X} \\ &= \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} = \lambda_1^{-1} (\mathbf{e}_1' \mathbf{X})^2 + \lambda_2^{-1} (\mathbf{e}_2' \mathbf{X})^2 + \dots + \lambda_p^{-1} (\mathbf{e}_p' \mathbf{X})^2 \end{aligned}$$

$$c^2 = \lambda_1^{-1} (Y_1)^2 + \lambda_2^{-1} (Y_2)^2 + \dots + \lambda_p^{-1} (Y_p)^2 \quad \therefore Y_k = \mathbf{e}_k' \mathbf{X} = c \lambda_k^{1/2}$$

Esta última equação define um elipsóide com os eixos,  $Y_1, Y_2, \dots, Y_p$ , dispostos na direção de  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p$ , respectivamente. Como  $\lambda_1$  é o maior autovalor, o maior eixo tem a direção definida por  $\mathbf{e}_1$ , similarmente aos demais autovalores de menor representatividade. Por exemplo, a representação gráfica em vetor aleatório de distribuição bivariada:



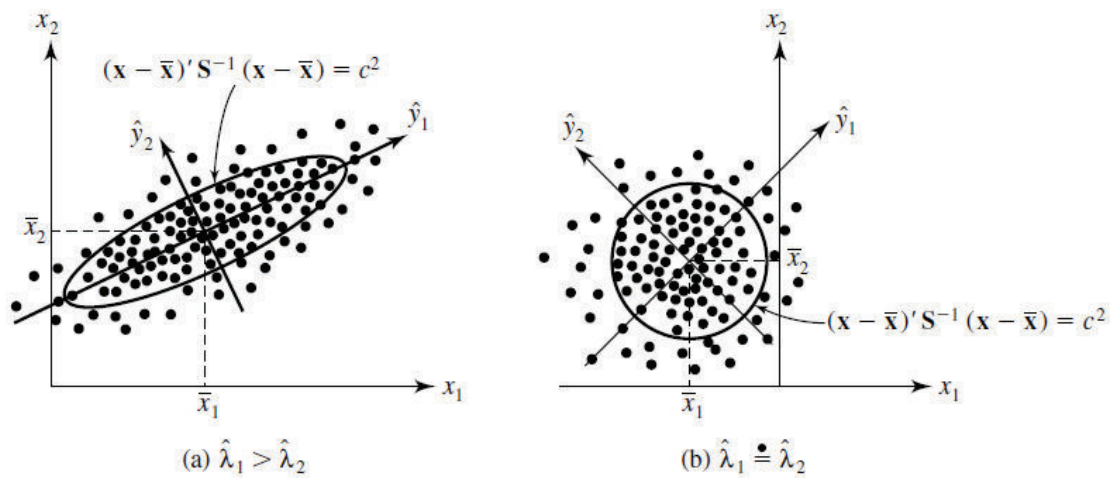
Quando  $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , cada ponto do sistema de coordenadas  $\mathbf{X}$  é projetado ortogonalmente no novo sistema de coordenada  $Y_1, Y_2, \dots, Y_p$ . Se  $\boldsymbol{\mu} > 0$ , como no painel (b), com componentes amostrais, as interpretações definidas são válidas da mesma forma, apenas sendo necessário definir o  $k$ -ésimo componente principal centrado na média:  $y_k = \mathbf{e}'_k(\mathbf{X} - \boldsymbol{\mu})$ .

Resumidamente:

- a) autovetores: determinam a rotação a ser realizada nos eixos nos sentidos de maior variabilidade;
- b) autovalores: determinam as variâncias desses novos eixos coordenados.

Assim, os dados podem ser expressos em novas coordenadas, que coincidem com os eixos de contorno  $c^2 = (\mathbf{X} - \boldsymbol{\mu})' \mathbf{S}^{-1} (\mathbf{X} - \boldsymbol{\mu})$ . Ademais, quando os contornos de distância constante são próximos de um círculo ou, equivalentemente, quando os autovalores de  $\mathbf{S}$  são próximos (ou iguais), a variação amostral é homogênea em todas as direções.

Veja as figuras abaixo.



### 4.3 Escores e gráficos dos componentes principais

São medidas compostas de cada componente para cada elemento amostral. Representa, pois, o grau que cada observação tem no grupo de variáveis com cargas significativas no componente. A definição do escore  $k$ -ésimo componente principal, para a  $j$ -ésima observação amostral é:

$$\hat{Y}_{jk} = \hat{\mathbf{e}}_k' \mathbf{X} = \hat{e}_{k1} X_{j1} + \hat{e}_{k2} X_{j2} + \dots + \hat{e}_{kp} X_{jp} \quad k = 1, 2, \dots, p, j = 1, 2, \dots, n.$$

Se  $\hat{\mathbf{Y}}_j' = [\hat{Y}_{j1}, \hat{Y}_{j2}, \dots, \hat{Y}_{jp}]$ ,  $\mathbf{X}_j' = [X_{j1}, X_{j2}, \dots, X_{jp}]$ ,  $\hat{\mathbf{O}} = [\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \dots, \hat{\mathbf{e}}_p]$ , então:

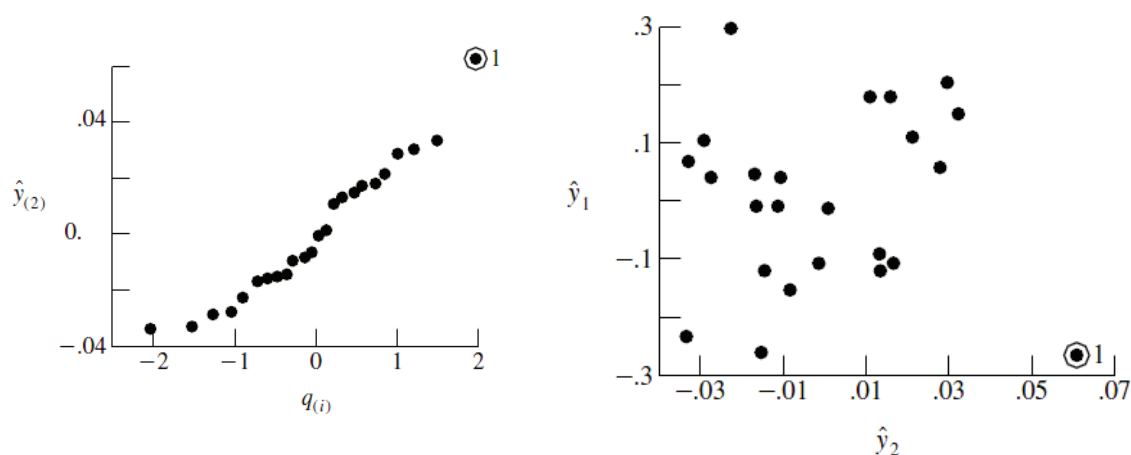
$$\hat{\mathbf{Y}}_j = \hat{\mathbf{O}}' \mathbf{X}_j \quad \text{ou} \quad \mathbf{X}_j = \hat{\mathbf{O}} \hat{\mathbf{Y}}_j$$

Ou seja, a observação amostral multivariada,  $\mathbf{X}_j$ , pode ser recuperada dos escores dos componentes principais correspondentes. Desse modo, as magnitudes dos últimos componentes determinam quão bem os primeiros componentes ajustariam as observações. Essa medida é feita tomando-se o quadrado desse desvio.

Muitas vezes, esses escores são plotados com o intuito de agrupar as observações, simplificar a representação para uma ou duas dimensões, bem como detectar outliers e averiguar a suposição de normalidade. Desde que os componentes principais sejam combinações lineares das variáveis originais, é razoável esperar que sejam normalmente distribuídos. Isto é frequentemente necessário para verificar que os primeiros poucos

componentes principais atendam essa suposição de normalidade, principalmente quando os seus escores são usados como “dados de insumos” para análises adicionais (regressão).

Desse modo, pode-se construir um diagrama de dispersão para um par de componentes e um gráfico Q-Q para cada componente. Dessa maneira, é possível verificar se os componentes aproximam-se de uma distribuição normal e se existem observações discrepantes poderão se aquelas cujo desvio ao quadrado é grande. Para maiores detalhes, veja exemplo 8.7 em Johnson e Wichern (2007, p.450-451).



## 5 INFERÊNCIAS PARA GRANDES AMOSTRAS

Conforme mencionado, os autovalores e autovetores da matriz de covariância e (correlação) são a essência de uma análise de componentes principais. Os autovetores definem a direção da máxima variabilidade, enquanto que os autovalores especifica a variância. Quando os primeiros autovalores são maiores que os demais, então, a variância total pode ser “expressa” em poucas  $p$  dimensões.

Na prática, as decisões sobre a qualidade do componente principal deve ser feita sobre as bases dos pares de autovalores-autovetores  $(\hat{\lambda}_1, \hat{e}_1), (\hat{\lambda}_2, \hat{e}_2), \dots, (\hat{\lambda}_p, \hat{e}_p)$  extraídos de

$\mathbf{R}_{p \times p}$ . Esses autovalores-autovetores em alguma medida da população subjacentes, o que requer alguns testes estatísticos.

### 5.1 Intervalos de confiança para os autovalores

Procura-se calcular intervalos de confiança e testa se  $\hat{\lambda}_i = 0$  na suposição que as observações são amostras aleatórias de uma população normal. Deve-se também assumir que os (desconhecidos) autovalores de  $\Sigma_{p \times p}$  são distintos e positivos:  $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$ . Anderson (1983) e Girshik (1939), Johnson e Wichern (2002) demonstraram que  $\hat{\lambda}_i$  são independentemente distribuídos e se aproximam  $N(\lambda_i, 2\lambda_i^2/n)$ . Usando essa distribuição normal, obtém-se:

$$\Pr\left(\left|\hat{\lambda}_i - \lambda_i\right| \leq z_{(\alpha/2)}\lambda_i\sqrt{2/n}\right) = 1 - \alpha$$

O intervalo de confiança  $100(1 - \alpha)\%$  de uma grande amostra para  $\hat{\lambda}_i$  pode ser obtido:

$$\frac{\hat{\lambda}_i}{\left(1 + z_{(\alpha/2)}\sqrt{2/n}\right)} \leq \lambda_i \leq \frac{\hat{\lambda}_i}{\left(1 - z_{(\alpha/2)}\sqrt{2/n}\right)}$$

Pelo exemplo 8.8. em Johnson e Wichern (2007, p.469), tem-se:

$$n = 103, \hat{\lambda}_i = 0.0014, z_{0.05/2} = 1.96 \rightarrow 95\%IC$$

$$\frac{0.0014}{\left(1 + 1.96\sqrt{2/103}\right)} \leq \lambda_i \leq \frac{0.0014}{\left(1 - 1.96\sqrt{2/103}\right)} \quad \text{ou} \quad 0.0011 \leq \lambda_i \leq 0.0019$$

Rejeita-se a hipótese nula que  $\hat{\lambda}_i = 0$ . No entanto, o limite inferior está bem próximo de zero, que é um indicativo que este componente principal representa pouco na variância total. O intervalo de confiança pode aumentar quando  $n$  aumentar. Logo, deve-se ter cuidado com esse teste, o que acaba sendo de pouca utilidade prática.

### 5.2 Teste da estrutura de correlação (esfericidade de Bartlett)

Para que a análise de componentes principais tenha algum sentido, é necessário que as variáveis sejam correlacionadas. Se as matrizes  $\mathbf{S}_{p \times p}$  e  $\mathbf{R}_{p \times p}$  forem diagonais, tal técnica devolverá, em alguma ordem, as próprias variáveis originais. Sob a suposição e

normalidade multivariada, é possível testar se  $\mathbf{S}$  e  $\mathbf{R}$  são matrizes diagonais ( $I$ , esfericidade), o que indica que as variáveis são independentes:

$$H_0 : \mathbf{R} = I$$

$$H_1 : \mathbf{R} \neq I$$

Para testar a hipótese nula, tem-se que calcular a estatística  $T$  :

$$T = - \left[ N - \frac{1}{6}(2P + 11) \right] \sum_{i=1}^p \ln(\hat{\lambda}_i)$$

Sob a hipótese nula, a estatística  $T$  tem  $\chi^2_{(1/2)p(p-1)}$  com  $(1/2)p(p-1)$  graus de liberdade.

Se  $T \geq T_c$ , para um nível de significância ( $\alpha$ ), rejeita-se a hipótese nula.

Aplicando o exemplo 3.5 para 12 empresas, em Mingoti (2005, p.94):

$$T = - \left[ 12 - \frac{1}{6}(2(3) + 11) \right] [\ln(2.4925) + \ln(0.4235) + \ln(0.084)] = 22.21$$

Para um nível de significância de 5%,  $\chi^2_{(1/2)p(p-1)}$  com 3 graus de liberdade,  $T_c = 7.81$ .

Portanto, há indicação de que a matriz de correlação não é diagonal.