Análise de componentes principais

Patrícia de Siqueira Ramos

UNIFAL-MG, *campus* Varginha

12 de Maio de 2016

ACP via matriz de correlações amostral

- Os CPs obtidos a partir de $\mathbf{S}$ são influenciados pelas variáveis de maior variância, sendo pouco úteis quando há grande discrepância entre as variáveis (diferentes escalas)
- Solução: padronizar cada variável pela sua média e desvio padrão, o que equivale a obter CPs a partir da matriz $\mathbf{R}$

ACP via matriz de correlações amostral

- Os CPs são então

$$Y_j = \mathbf{a}_j^T \mathbf{Z} = a_{j1} Z_1 + \cdots + a_{jp} Z_p,$$

em que $Z_i = \dfrac{X_i - \bar{X}_i}{\sqrt{S_{ii}}}$ é a variável padronizada e

$$V(Y_j) = \lambda_j, \qquad Cov(Y_i, Y_j) = 0$$

ACP via matriz de correlações amostral

- A correlação entre Y_j e Z_i é dada por:

$$r_{Y_j,Z_i} = a_{ji}\sqrt{\lambda_j}$$

- e as variáveis Z_i com maiores coeficientes no CP Y_j são as mais correlacionadas com o CP

ACP via matriz de correlações amostral

- A correlação entre Y_j e Z_i é dada por:

$$r_{Y_j,Z_i} = a_{ji}\sqrt{\lambda_j}$$

- e as variáveis Z_i com maiores coeficientes no CP Y_j são as mais correlacionadas com o CP
- A variância total de $\mathbf{Z}$ será

$$\text{Variância total}(\mathbf{Z}) = \sum_{i=1}^{p} \lambda_i = \text{tr}(\mathbf{R}) = p$$

Variação explicada

- A proporção da variância total explicada pelo j-ésimo CP é

$$\frac{V(Y_j)}{\text{V.total de } \mathbf{Z}} = \frac{\lambda_j}{tr(\mathbf{R})} = \frac{\lambda_j}{p}$$

Variação explicada

- A proporção da variância total explicada pelo j-ésimo CP é

$$\frac{V(Y_j)}{\text{V.total de } \mathbf{Z}} = \frac{\lambda_j}{tr(\mathbf{R})} = \frac{\lambda_j}{p}$$

- Geralmente, quando utilizamos **R** no lugar de **S**, precisamos de um número maior de CPs para explicar a mesma quantidade de variação total obtida utilizando-se **S**

Variação explicada

- A proporção da variância total explicada pelo j-ésimo CP é

$$\frac{V(Y_j)}{\text{V.total de } \mathbf{Z}} = \frac{\lambda_j}{tr(\mathbf{R})} = \frac{\lambda_j}{p}$$

- Geralmente, quando utilizamos **R** no lugar de **S**, precisamos de um número maior de CPs para explicar a mesma quantidade de variação total obtida utilizando-se **S**

Obs.: Apesar de estarmos usando a mesma notação, os valores de a_{ij} e λ_j obtidos a partir de **R** não são os mesmos dos obtidos a partir de **S**

Como os CPs predizem $\mathbf{S}$

Seja a matriz $\mathbf{A}$ $p \times p$ formada pelos vetores de coeficientes

$$\mathbf{A} = \left[\begin{array}{cccc} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_p \end{array} \right],$$

que definem os p CPs.

Como os CPs predizem $\mathbf{S}$

Seja a matriz $\mathbf{A}$ $p \times p$ formada pelos vetores de coeficientes

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_p \end{bmatrix},$$

que definem os p CPs.
A matriz de covariâncias amostral das variáveis $X_1, X_2, \dots, X_p$ é dada por (decomposição espectral de $\mathbf{S}$)

$$\mathbf{S} = \mathbf{A}\mathbf{\Lambda}\mathbf{A}^T,$$

em que

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{bmatrix}.$$

Obs.: Vimos que $\mathbf{S} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T$

Como os CPs predizem $\mathbf{S}$

A matriz de covariâncias amostral $\mathbf{S}$ pode ser escrita de forma mais simples

$$\mathbf{S} = \mathbf{A}^* \mathbf{A}^{*T},$$

em que $\mathbf{a}_i^* = \sqrt{\lambda_i}\mathbf{a}_i$

Definir número de CPs

- A variação nas p variáveis originais é totalmente contabilizada por todos os p componentes principais
- Mas, quantos CPs são necessários para resumir adequadamente um conjunto de dados? Há técnicas variadas:

Definir número de CPs

- A variação nas p variáveis originais é totalmente contabilizada por todos os p componentes principais
- Mas, quantos CPs são necessários para resumir adequadamente um conjunto de dados? Há técnicas variadas: (i) reter os k primeiros componentes que expliquem grande parte (70 a 80%) da variação total

Definir número de CPs

- A variação nas p variáveis originais é totalmente contabilizada por todos os p componentes principais
- Mas, quantos CPs são necessários para resumir adequadamente um conjunto de dados? Há técnicas variadas:
(i) reter os k primeiros componentes que expliquem grande parte (70 a 80%) da variação total
(ii) se a matriz **R** for usada, ficar apenas com os componentes com valores de $\lambda > 1$ (KAISER, 1958). Outros autores optam por $\lambda > 0,7$ (JOLLIFE, 1972)

Definir número de CPs

- A variação nas p variáveis originais é totalmente contabilizada por todos os p componentes principais
- Mas, quantos CPs são necessários para resumir adequadamente um conjunto de dados? Há técnicas variadas:
  (i) reter os k primeiros componentes que expliquem grande parte (70 a 80%) da variação total
  (ii) se a matriz **R** for usada, ficar apenas com os componentes com valores de $\lambda > 1$ (KAISER, 1958). Outros autores optam por $\lambda > 0,7$ (JOLLIFE, 1972)
  (iii) analisar um *scree plot* (gráfico de $\lambda \times$ número de componentes). Devemos procurar pelo "cotovelo", que é o ponto depois do qual os λ_i diminuem mais lentamente

Definir número de CPs

- A variação nas p variáveis originais é totalmente contabilizada por todos os p componentes principais
- Mas, quantos CPs são necessários para resumir adequadamente um conjunto de dados? Há técnicas variadas:
  (i) reter os k primeiros componentes que expliquem grande parte (70 a 80%) da variação total
  (ii) se a matriz **R** for usada, ficar apenas com os componentes com valores de $\lambda > 1$ (KAISER, 1958). Outros autores optam por $\lambda > 0,7$ (JOLLIFE, 1972)
  (iii) analisar um *scree plot* (gráfico de $\lambda \times$ número de componentes). Devemos procurar pelo "cotovelo", que é o ponto depois do qual os λ_i diminuem mais lentamente
  (iv) verificar se o componente tem interpretação razoável e útil

Calcular os escores - a partir da matriz $\mathbf{S}$

- Há algumas formas de calcular o escore de um componente em uma observação (a partir da matriz $\mathbf{S}$). Uma forma muito comum é:

  - Fazer o produto coeficiente$\cdot$(valor da variável - média), somando com os valores de todas as p variáveis (parecido com o que o R faz)

Ex.: Calcular os escores

Ex.: a) Obter os escores da análise de componentes principais a partir de **S** usando o produto coeficiente·(valor da variável - média) (que é parecido com a forma que o R calcula) para os dados das 12 empresas.

$$\mathbf{X} = \begin{bmatrix} 9893 & 564 & 17689 \\ 8776 & 389 & 17359 \\ 13572 & 1103 & 18597 \\ 6455 & 743 & 8745 \\ 5129 & 203 & 14397 \\ 5432 & 215 & 3467 \\ 3807 & 385 & 4679 \\ 3423 & 187 & 6754 \\ 3708 & 127 & 2275 \\ 3294 & 297 & 6754 \\ 5433 & 432 & 5589 \\ 6287 & 451 & 8972 \end{bmatrix}, \bar{\mathbf{X}} = \begin{bmatrix} 6267,42 \\ 424,67 \\ 9606,42 \end{bmatrix}, \mathbf{A} = \begin{bmatrix} -0,425 & -0,900 & -0,099 \\ -0,028 & -0,097 & 0,995 \\ -0,905 & 0,426 & 0,016 \end{bmatrix}.$$

(continua)

Ex.: Verificar que $\mathbf{S} = \mathbf{A\Lambda A}^T$

b) Verificar que $\mathbf{S} = \mathbf{A\Lambda A}^T$, sabendo que a ACP resultou: $V(Y_1) = 41474391$, $V(Y_2) = 2539507$, $V(Y_3) = 21092,53$ e que

$$\mathbf{S} = \begin{bmatrix} 9550608,6 & 706121,06 & 14978232,5 \\ 706121,1 & 76269,52 & 933915,1 \\ 14978232,5 & 933915,06 & 34408113,0 \end{bmatrix}.$$

Ex.: ACP para a matriz $\mathbf{R}$

c) Obter os CPs a partir da matriz $\mathbf{R}$:

$$\mathbf{R} = \begin{bmatrix} 1 & 0,83 & 0,83 \\ 0,83 & 1 & 0,58 \\ 0,83 & 0,58 & 1 \end{bmatrix}.$$

Sabendo que a decomposição espectral resultou em

$$\mathbf{\Lambda} = \begin{bmatrix} 2,49 & 0 & 0 \\ 0 & 0,42 & 0 \\ 0 & 0 & 0,08 \end{bmatrix},$$

$$\mathbf{P} = \begin{bmatrix} -0,62 & -0,001 & 0,79 \\ -0,56 & -0,71 & -0,44 \\ -0,56 & 0,71 & -0,43 \end{bmatrix}.$$