SENSOMETRIA

Adilson dos Anjos

Departamento de Estatística
Universidade Federal do Paraná
aanjos@ufpr.br

Curitiba, PR
28 de abril de 2015

SENSOMETRIA

– Análise de componentes Principais (ACP)–

ACP

- PCA: Principal Component Analysis (Pearson, 1901)
- Objetivo: explicar a estrutura de variância/covariância por meio de combinações lineares das variáveis originais;
- As combinações lineares são chamadas de Componentes Principais;
- Essas combinações são não correlacionadas;

ACP

- Busca-se uma redução do número de p variáveis para k componentes principais;
- Em geral, a análise de componentes principais serve como um método intermediário de avaliação;
- Por exemplo: construção de agrupamentos em análise de segmentação (cluster).

ACP

- Na análise de componentes principais, busca-se a informação contida em p variáveis por meio de k componentes principais não correlacionadas $(k < p)$;
- A qualidade da informação pode ser medida por meio da proporção de variância total explicada pelas k componentes principais;
- A suposição de normalidade não é necessária para utilização de ACP.

ACP

- Para a obtenção dos componentes principais utiliza-se a matriz de covariâncias dos vetores aleatórios das variáveis originais;
- É comum utilizar-se uma transformação das variáveis;
  Os componentes principais podem ser obtidos a partir da matriz de covariância das variáveis originais padronizadas ou,
- ... de maneira equivalente, a partir da matriz de correlação das variáveis originais
- Os componentes principais dependem apenas da matriz de covariância (ou correlação);

ACP

- Dado o vetor $X' = [X_1, X_2, \ldots, X_p]$ e a matriz de covariâncias Σ com autovalores $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$;
- As combinações lineares são dadas da forma:

$$
\begin{aligned}
Y_1 &= l_1' X &&= l_{11} X_1 + l_{21} X_2 + \ldots + l_{p1} X_p \\
Y_2 &= l_2' X &&= l_{12} X_1 + l_{22} X_2 + \ldots + l_{p2} X_p \\
\vdots \quad & \quad \vdots && \quad \vdots \\
Y_p &= l_p' X &&= l_{1p} X_1 + l_{2p} X_2 + \ldots + l_{pp} X_p
\end{aligned}
$$

- A primeira combinação linear é a que possui a maior variância

ACP

- A variância total é:

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

- Assim, a proporção da variância total explicada pela k-ésima componente principal é

$$\text{Var devida a k-ésima CP} = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \ldots + \lambda_p}$$

ACP

- Se Σ é a matriz de covariância associada com o vetor $X' = [X_1, X_2, \ldots, X_p]$ e Σ tem os pares de autovalores e autovetores (λ_p, e_p) com $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$, o i-ésimo componente principal é dado por:

$$Y_i = e_{1i}X_1 + e_{2i}X_2 + \ldots + e_{pi}X_p$$

- Com a condição:

$$Var(Y_i) = e_i'\Sigma e_i = \lambda_i \qquad e \qquad Cov(Y_i, Y_k) = e_i'\Sigma e_k = 0 (i \neq k)$$

ACP

- Ver exemplo 8.1, p. 360: Johnson e Wichern. 1992.

ACP

- Em geral, entre 80 e 90% da variabilidade pode ser explicada por até 3 componentes principais (considerando um grande número de variáveis originais);
- A variabilidade depende do fenômeno em estudo;
- Em alguns casos, pode-se considerar uma variabilidade menor, em geral, em torno de 70%.

ACP

- Há diferenças entre os componentes principais obtidos a partir da matriz de variância e covariância em comparação com a matriz de correlação;
- Recomenda-se a padronização quando as variáveis possuem escalas diferentes (inclusive em magnitude)

Representação Gráfica

ACP

- Utiliza-se um gráfico de dispersão entre as duas componentes principais e o escores de cada componente;

ACP

Definição dos Componentes

- Percentual: depende do fenômeno (scree-plot);
- Interpretação prática do componente;
- Aproximação da matriz de variâncias/covariâncias (original).

ACP

- Quando as variáveis possuem Dxistribuição Normal (normal multivariada) podem ser realizadas algumas inferências;
- Exemplo: elipses de confiança.

SENSOMETRIA

– ACP com FactoMineR–
Exemplos:

ACP

- Coxinha (MINGOTI, 2007)
- Suco de Laranja (HUSON; LÊ; PAGÈS; 2011)

ACP

Algumas referências da web:

- ACP
- ACP
- ACP
- ACP