

ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

Análise de Componentes Principais

- **Pearson (1901)**: Procurava linhas e planos que melhor se adequavam a um conjunto de pontos em um espaço p -dimensional. Criou a Componente Principal (PC)
- **Hotelling (1933)**: Procurava encontrar um pequeno conjunto de variáveis fundamentais que expressa p variáveis. Hotelling procurou maximizar suas 'componentes' no senso da variância das variáveis originais. Chamou de Componentes Principais.

Análise de Componentes Principais

- Pearson (1901) e Hotelling (1933)
 - O objetivo principal é a obtenção de um pequeno número de combinações lineares (componentes principais) de um conjunto de variáveis, que retenham o máximo possível da informação contida nas variáveis originais.
 - Redução dos dados originais
 - Facilitar a interpretação através da descoberta de relacionamentos não suspeitos previamente.

Análise de Componentes Principais

- A análise de componentes principais substitui um conjunto de variáveis correlacionadas por um conjunto de novas variáveis não-correlacionadas, sendo essas combinações lineares das variáveis iniciais e colocadas em ordem decrescente por suas variâncias

$$\text{Var CP}_1 > \text{Var CP}_2 > \dots > \text{Var CP}_p$$

Análise de Componentes Principais

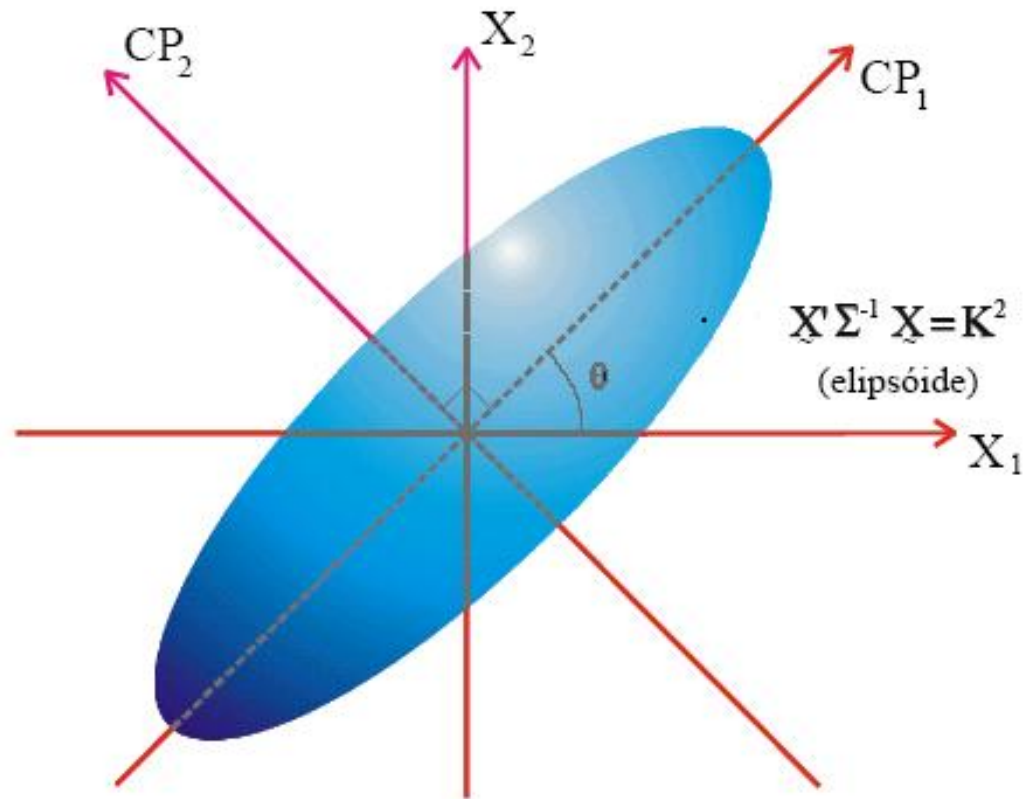
- Algebricamente, componentes principais são combinações lineares particulares das “p” variáveis aleatórias X_1, X_2, \dots, X_p ;
- Geometricamente, essas combinações lineares representam a relação de um novo sistema de coordenadas obtido por deslocamento e rotação do sistema original com X_1, X_2, \dots, X_p como eixos;
- Os novos eixos representam as direções com variabilidade máxima e fornecem uma descrição mais simples e mais parcimoniosa da estrutura de covariância;
- Os componentes principais dependem da matriz de correlação (\mathbf{r}) ou da matriz de covariâncias (Σ) de X_1, X_2, \dots, X_p . O seu desenvolvimento não necessita da suposição de normalidade.

Análise de Componentes Principais

□ Exemplos:

- Aplicação na confiabilidade de sistemas complexos – redução no conjunto de variáveis originais (peças) para três variáveis (componentes principais) e no final o estudo mostrou que bastava apenas o primeiro deles.
- Análise de componentes principais em imagens multi-temporais de satélites para estudo de vulnerabilidade à perda de solo no semi-árido nordestino. Basicamente foi utilizada a ACP para reunir as informações de duas imagens para compor duas outras, não correlacionadas entre si, com melhores propriedades interpretativas.

Análise de Componentes Principais



Análise de Componentes Principais

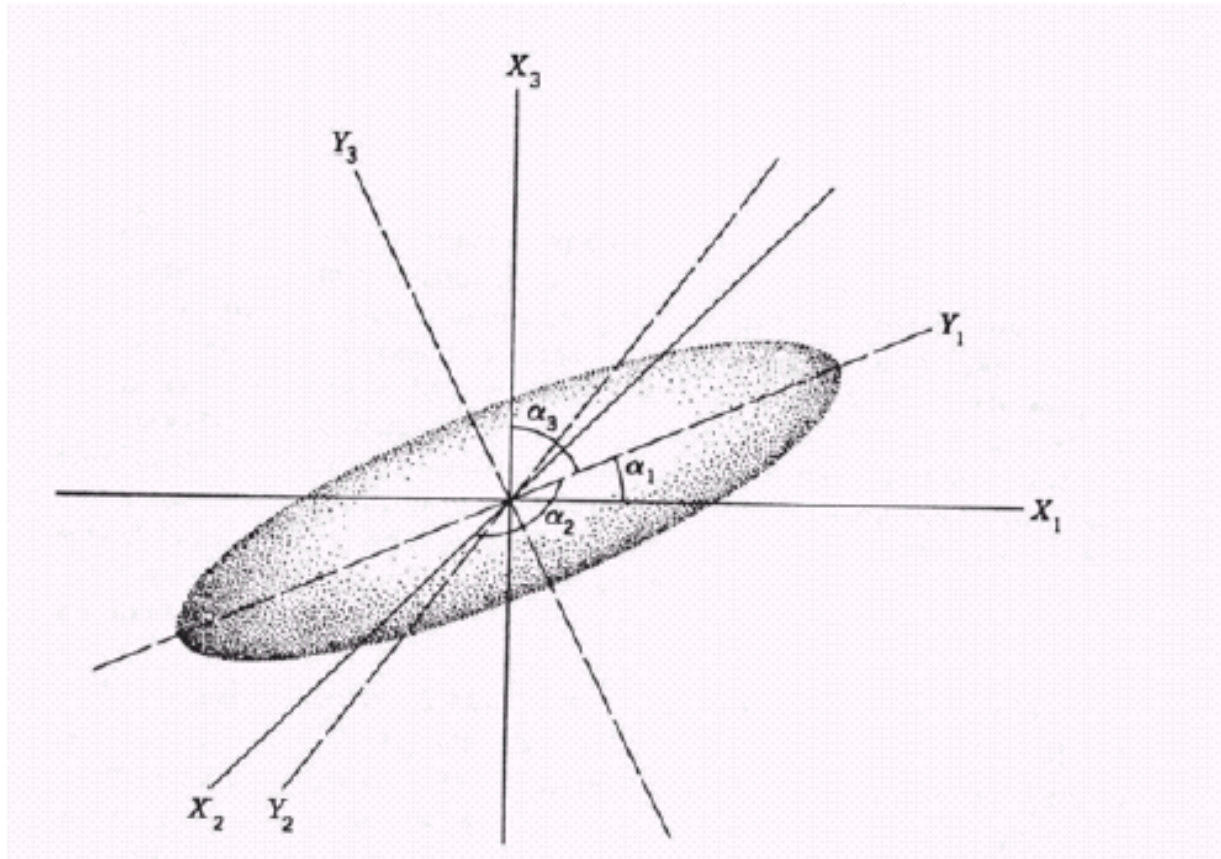


Figura 2.1 Eixos principais com três variáveis

Análise de Componentes Principais

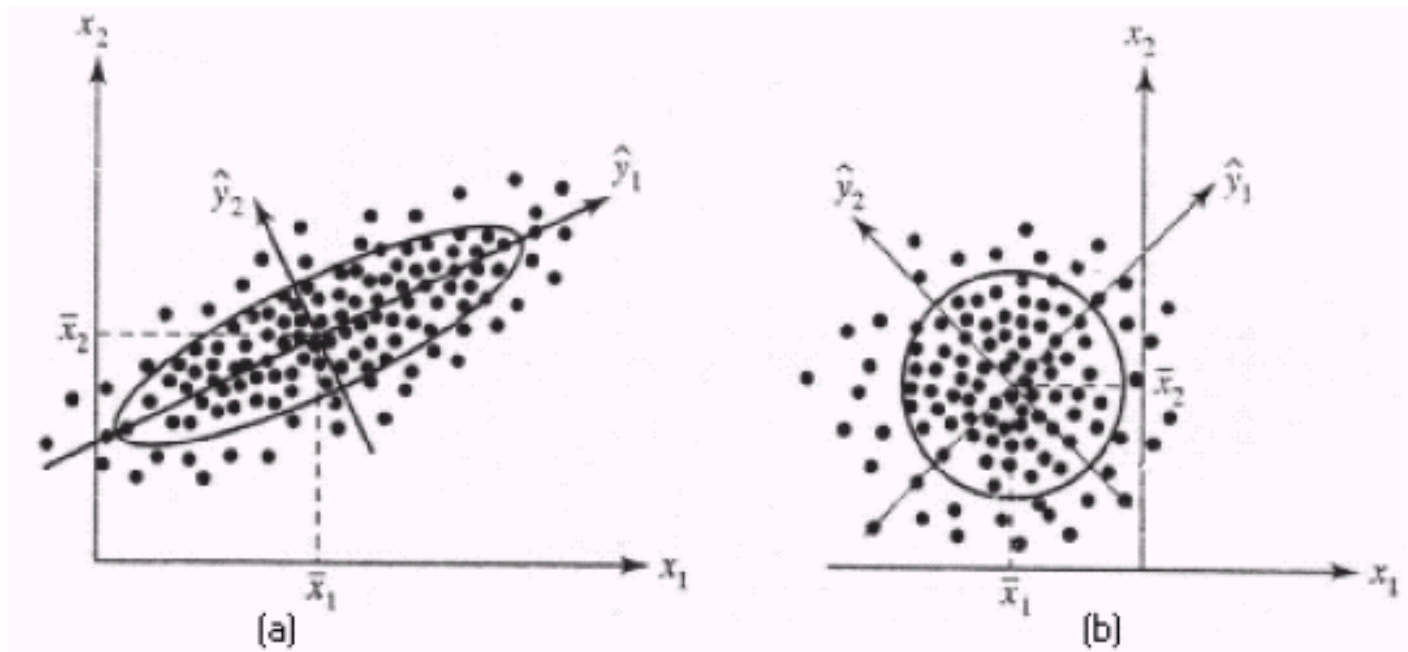
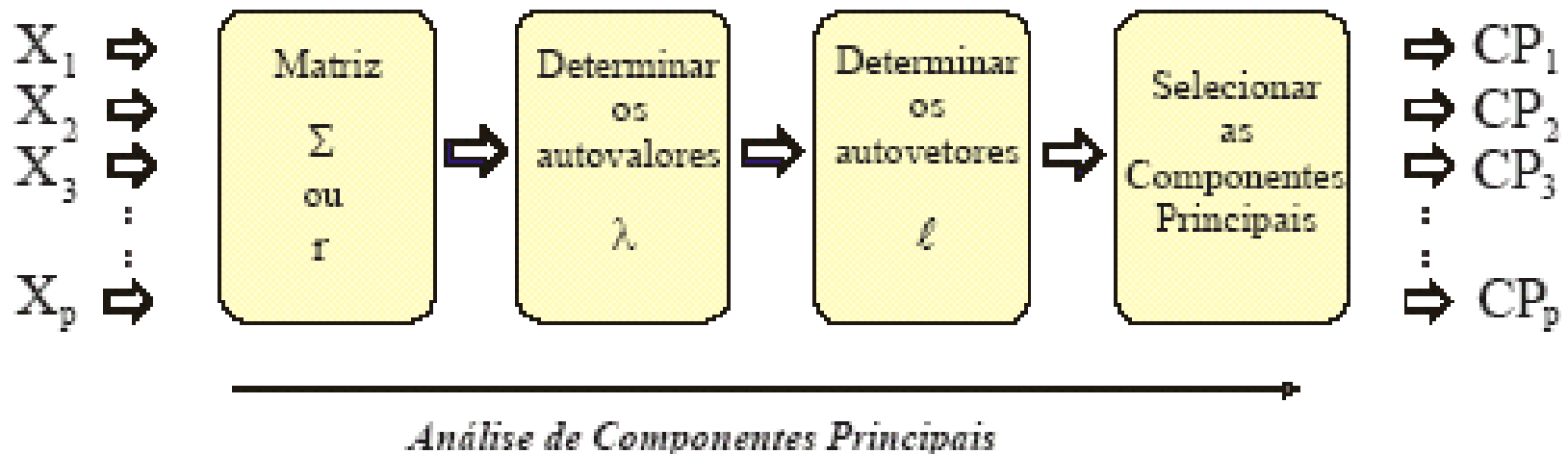


Figura 2.2: (a) eixos de máxima variabilidade $\hat{\lambda}_1 \neq \hat{\lambda}_2$ e
(b) eixos com variação isotrópica $\hat{\lambda}_1 \doteq \hat{\lambda}_2$

Análise de Componentes Principais

FIGURA 7: Esquema para obtenção das “p” componentes principais



Análise de Componentes Principais

- Seleção do número de componentes
 - Kaiser ou Método da Raiz Latente: autovalores maiores que 1 (Johnson = 0,7);
 - *Scree test* – gráfico dos autovalores.

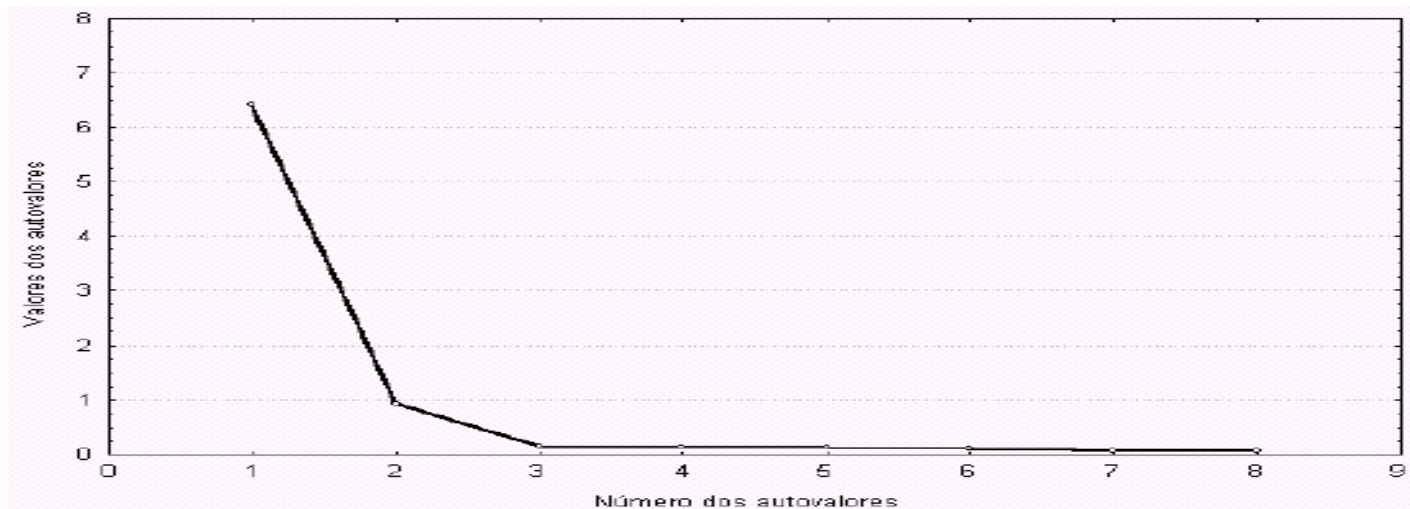


Figura 2.4: Seleção do número de componentes principais pelo método *scree test*. No caso, selecionam-se duas componentes

Análise de Componentes Principais

- Após a seleção dos componentes, a elaboração do gráficos dos dois primeiros componentes pode ser muito útil para entender a relação entre as variáveis e os componentes;
- Se apenas os dois primeiros componentes “explicarem” mais de 80 ou 90% da variabilidade total das variáveis originais isto significa que o fenômeno sob estudo pode ser muito simplificado;
- No primeiro exemplo, de 80 variáveis originais apenas três componentes “explicam” quase tanto quanto os 80 e, no final, apenas o primeiro componente já foi suficiente para o objetivo do trabalho.