

Lista # 4 – Análise de Componentes Principais

1. Os dados do arquivo *solo.mtw* são observações de uma amostra aleatória constituída de 20 amostras de um determinado solo. Para cada amostra foram medidas as percentagens de areia, sedimentos, argila, quantidade de material orgânico e a acidez do solo (pH). Observe que as porcentagens de areia, sedimentos e argila somam 100%.
 - a) Calcule a matriz de covariância amostral S e a decomponha em seus autovalores. Mostre que a partir daí que esta matriz é não negativa definida.
 - b) Calcule os autovalores normalizados e calcule as componentes principais amostrais. Qual é a consequência de a) nas componentes que você construiu?
 - c) Escolha uma estratégia para tornar a matriz S positiva definida. Explique sua estratégia.
 - d) Faça agora uma análise de componentes principais usando a matriz positiva definida que você encontrou em c). Interprete as componentes e faça uma análise dizendo quantas componentes deveriam ser retidas para explicar aproximadamente a estrutura de variabilidade das variáveis originais. Calcule as correlações das componentes que você decidiu reter no sistema com todas as variáveis originais.
 - e) Para cada componente que você eliminou em d) teste a hipótese de que o autovalor teórico correspondente à componente é igual a zero.
 - f) Escolha a componente que achar mais conveniente para comparar as 20 amostras de solo entre si. Calcule o escore da componente para cada uma das vinte observações. Faça uma análise destes escores e tire alguma conclusão.
 - g) Faça agora uma análise de componentes principais usando a matriz de correlação amostral R calculada a partir da matriz positiva definida do item c). Compare os resultados com aqueles encontrados em d).
 - h) Repita o item f) considerando a análise que você fez em g)

Exercícios do Johnson e Wichern (1998).

2. 8.1 e 8.2 (pág. 503)
3. 8.3 (pág. 503).
4. 8.10 e 8.11 (pág. 506).