

Lista nº 6 – Análise de Componentes Principais e Análise Fatorial

1. O conjunto de dados “*pendigit*”. (<http://archive.ics.uci.edu/ml/datasets/Pen-Based+Recognition+of+Handwritten+Digits>) pertence ao UC Irvine Machine Learning Repository. Leia a descrição do conjunto de dados na página da Web citada acima. O arquivo `pendigit3.txt` contém um subconjunto daqueles dados, com $n = 1055$ observações de $p = 16$ variáveis. A última coluna do conjunto de dados (17ª. coluna) contém 3's, indicando que a observação multivariada ($p=16$) representa uma particular caligrafia do dígito 3. As 16 variáveis referem-se às posições da caneta no papel em 8 instantes igualmente espaçados no tempo, estando dispostos como $(x_1, y_1), (x_2, y_2), \dots, (x_8, y_8)$.
 - a) Calcule a matriz de covariância amostral S e a decomponha em seus autovalores. A partir da decomposição espectral, verifique se esta matriz é não negativa definida.
 - b) 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 aproximar 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) Agora abra os dados em `pendigit8.txt`. Combine dois conjuntos de dados, formando uma matriz de dados com $n = 2110$ observações. Execute o PCA neste conjunto de dados e apresente visualmente o resultado por meio de um `scatterplot matrix` dos componentes principais. Você identifica uma boa separação entre as observações correspondente ao dígito 3 e aquelas correspondentes ao dígito 8?
2. Sejam Y_i , a i -ésima componente principal, X_k , a k -ésima variável e, $(\lambda_i, \mathbf{e}_i)$ o i -ésimo par de autovalores/autovetores da matriz de covariâncias Σ . Mostre:

$$\text{Corr}(Y_i, X_k) = \frac{e_{ik} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}, \quad i, k = 1, 2, \dots, p.$$

3. Abra o arquivo `home.csv`. Ele se refere a uma pesquisa que foi feita para estudar o comportamento das pessoas em relação ao hábito de fazer compras. As variáveis são assim definidas:
 - Q1: Eu prefiro passar uma noite calma e tranquila em casa do que sair para ir a uma festa;
 - Q2: Eu sempre observo cuidadosamente os preços, mesmo em produtos pequenos;
 - Q3: A internet é mais interessante que cinemas
 - Q4: Eu não compro produtos que são anunciados na televisão pois não acredito em propagandas;
 - Q5: Eu gosto mesmo é de ficar em casa;

- Q6: Eu sempre compro produtos que estão vendidos com desconto;
- Q7: As empresas desperdiçam muito dinheiro em propaganda.

Foram entrevistadas 25 pessoas. Cada pessoa teve que expressar sua concordância com a frase em questão. A escala medida foi de 1 a 7, onde 1 representa a discordância total e 7 representa a concordância total.

- a) Faça uma análise fatorial e interprete os fatores. Justifique sua escolha pelo método de eliminação de matrizes \mathbf{L} e Ψ que tiver utilizado.
- b) Faça uma rotação ortogonal dos fatores encontrados em (a) e compare as soluções. Qual lhe parece melhor?
- c) Dê uma matriz residual relativa à análise fatorial antes da rotação e depois da rotação ortogonal. O que você pode observar?
- d) Dê as comunalidades e variâncias específicas estimadas antes da rotação e depois da rotação ortogonal.
- e) O que você poderia dizer sobre o hábito de compras da população da qual esta amostra de pessoas foi obtida? (Para responder esta questão você precisa obter escores dos fatores usando a solução que achou mais conveniente em (b).
- f) Agora faça uma rotação não ortogonal nos seus fatores encontrados em (a). Compare as soluções encontradas com as anteriores

Exercícios do Johnson e Wichern (2007).

4. 8.1 e 8.2 (pág. 470)
5. 8.3 (pág. 471).
6. 8.4 e 8.5 (pág. 471). Ler “Principal components with special structures” (pág. 439).
7. 8.10 e 8.11 (pág. 473).
8. 9.1, 9.2 e 9.3, pág. 530.
9. 9.7, pág. 531.
10. 9.8, pág. 531.
11. 9.26 e 9.27, pág. 537.
12. 9.28, 9.29, 9.30 e 9.31, pág. 537.