

Análise Multivariada de Dados - 2022

Lista de Exercícios

Kaique Matias de Andrade Roberto

19 de outubro de 2022

Estes são os Exercícios recomendados para a disciplina. Afim de que você possa extrair o maior proveito possível destes exercícios tenha em mente as seguintes observações:

- esta é a **única** lista de exercícios da disciplina toda;
- esta lista **contém** os exercícios que resolveremos em aula;
- as Seções estão nomeadas de acordo com as aulas (por exemplo, na Seção 10 estão os exercícios recomendados para a Aula 10);
- os exercícios que aparecem em aula estão marcados com (A);
- os exercícios com (*) ou (**) são exercícios que consideramos mais desafiadores.

0 Tipos de Variáveis, Estatística Descritiva, Principais Variáveis Aleatórias, Amostragem

Exercício 0.1. Qual é a diferença entre variáveis qualitativas e quantitativas?

Exercício 0.2. Classificar as variáveis a seguir:

- | | |
|--|--|
| a - Faturamento da empresa; | e - Distância percorrida em km; |
| b - Ranking de desempenho: bom, médio, ruim; | f - Casado: sim ou não; |
| c - Número de carros vendidos; | g - Localização: Barueri ou Santana de Parnaíba; |
| d - Faixa de renda; | h - Municípios do Grande ABC. |

Exercício 0.3 (A). Considere a tabela abaixo:

Nome	Idade (anos)	Peso (kg)	Altura (m)
Mariana	48	62	1,60
Luiz	54	84	1,76
Roberta	41	56	1,62
Leonardo	30	82	1,90
Melissa	28	54	1,68
Sandro	50	70	1,72

Calcule a média, variância e desvio-padrão das variáveis Idade, Peso e Altura.

Exercício 0.4 (Resumo Estatística Descritiva). Faça um Resumo das Seções 2.1-2.4 do Livro [2].

Exercício 0.5 (Resumo Probabilidade). Faça um Resumo das Seções 4.1-4.7 do Livro [2] ou do Capítulo 5 do Livro [7].

Exercício 0.6. Defina o conceito de variável aleatória. Você consegue descrever ao menos três variáveis aleatórias dentro do escopo da sua área de atuação?

Exercício 0.7 (Resumo Amostragem). Faça um Resumo das Seções 6.1-6.4 do Livro [2].

Exercício 0.8. Quais são as principais diferenças entre os modelos Bernoulli, Binomial e Poisson?

Exercício 0.9. Aponte as diferenças entre uma variável aleatória discreta e uma variável aleatória contínua.

Exercício 0.10. Descreva o que é a "falta de memória" da Distribuição Geométrica.

Exercício 0.11 (Resumo Variáveis Aleatórias). Faça um Resumo do Capítulo 5 do Livro [2].

Exercício 0.12. Faça uma tabela com a esperança, variância e distribuição acumulada das principais distribuições (vide slides 68 e 69 da Aula-00).

Exercício 0.13 (A). Inspirado nos dados da planilha motocicletas.xlsx, realize as seguintes tarefas:

- a - Crie uma base de dados (uma planilha por exemplo) com ao menos 15 entradas, 3 variáveis quantitativas e 3 variáveis qualitativas.
- b - Realize a análise descritiva da sua base de dados.

1 O que é Análise Multivariada? O que é Análise de Conglomerados?

Exercício 1.1 (A). Descreva em suas palavras os objetivos da análise multivariada. Qual é a principal diferença entre a análise multivariada e a estatística descritiva?

Exercício 1.2 (A). Descreva em suas palavras os objetivos da análise de conglomerados.

Exercício 1.3 (Resumo Matrizes). Faça um Resumo do Capítulo 2 do Livro [6].

Exercício 1.4 (Resumo Gráficos Multivariados). Faça um Resumo do Capítulo 3 do Livro [6].

Exercício 1.5 (A). Usando os dados na planilha "dados-multivariados", quais insights você consegue obter para os dados no spreadsheet "empregados-europa" à partir de gráficos univariados/bivariados?

Exercício 1.6 (**). Uma Análise Estatística bem-executada está fortemente apoiada em um procedimento consistente de coleta de dados. Faça um resumo dos Capítulos 1 e 2 do Livro [3] para ter um primeiro contato com os vários cuidados ao lidar com a coleta de dados.

2 Análise de Conglomerados I: Medidas de Distância e Similaridade

Exercício 2.1. Por quê existe a necessidade de eventualmente considerar expressões diferentes de distância para realizar uma análise de conglomerados?

Exercício 2.2. Faça um quadro com o principal uso de cada distância/medida de similaridade definida na Aula-02.

Exercício 2.3. Calcule as medidas de distância/similaridade para os dados na planilha "aula-02-exemplos.xlsx".

3 Análise de Conglomerados II: Esquemas de Aglomeração

Exercício 3.1. Descreva com as suas palavras o que é um esquema de aglomeração.

Exercício 3.2. Qual é a diferença entre um esquema hierárquico e não-hierárquico?

Exercício 3.3 (A). Para os dados na aba "dados" da planilha "aula-03-exemplos.xlsx" elaborar uma análise de agrupamentos por meio de um esquema de aglomeração hierárquico com método de encadeamento único usando a distância euclidiana.

Exercício 3.4 (*). Para os dados na aba "dados" da planilha "aula-03-exemplos.xlsx" elaborar uma análise de agrupamentos por meio de um esquema de aglomeração hierárquico com método de encadeamento único usando a distância de Minkowski com $m = 5$ e $m = 10$.

4 Análise de Conglomerados III: Implementações

Exercício 4.1 (A). Com base nas implementações dadas em aula, compare os dendogramas obtidos com a aba "empregados-europa" da planilha "aula-04-exemplos-livro-manly.xlsx" para variados métodos hierárquicos e não-hierárquicos (no mínimo dois de cada).

5 Coisas que Precisamos Saber Sobre Matrizes

Para referências e/ou aprofundamentos sobre matrizes e álgebra linear consulte por exemplo [1] (iniciante), [5] (intermediário), [4] (avançado).

Exercício 5.1 (A). Considere as matrizes

$$A = \begin{pmatrix} 3 & 2 \\ -1 & 5 \end{pmatrix}, B = \begin{pmatrix} 2 & -2 & 5 \\ 0 & 2 & 3 \end{pmatrix}, C = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix},$$

$$D = \begin{pmatrix} 0 & -2 \\ -3 & 1 \end{pmatrix}, E = (1 \ 3).$$

Verifique se é possível calcular cada uma das matrizes abaixo, e caso afirmativo, faça o cálculo:

a - $A + 3D$,

d - $B - C^t$,

b - $2D - 3A$,

e - AB ,

c - $B - C$,

f - $DA - AD$.

Exercício 5.2 (A). Calcule o determinante das matrizes abaixo. Também verifique se as matrizes abaixo são inversíveis e caso positivo, calcule a inversa.

a - $A = \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}$

c - $C = \begin{pmatrix} 1 & 2 & 2 \\ 0 & 1 & 2 \\ 1 & 3 & 4 \end{pmatrix}$

e - $E = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 0 & 1 \\ 3 & 1 & -2 \end{pmatrix}$

b - $B = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix}$

d - $D = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 1 & 1 \\ 3 & -1 & 1 \end{pmatrix}$

f - $F = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & -1 \\ 0 & 2 & 0 & 3 \end{pmatrix}$

Exercício 5.3 (A). Decida se cada uma das seguintes matrizes é diagonalizável. Caso afirmativo, realize a diagonalização.

a - $\begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}$

c - $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$

e - $\begin{pmatrix} 1 & 1 & 2 \\ 0 & 0 & -2 \\ 0 & 1 & 3 \end{pmatrix}$

b - $\begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$

d - $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$

f - $\begin{pmatrix} 2 & 1 & 1 \\ 3 & 0 & 1 \\ 6 & 2 & 1 \end{pmatrix}$

6 Análise de Componentes Principais

Exercício 6.1. Descreva os passos necessários para a realização de uma Análise de Componentes Principais.

Exercício 6.2 (A). Realize uma Análise de Componentes Principais considerando os dados na planilha "metodos-multivariados-aula-06-exemplos.xlsx" na aba "passaros".

Exercício 6.3. Realize uma Análise de Componentes Principais considerando os dados na planilha "metodos-multivariados-aula-06-exemplos.xlsx" na aba "empregados-europa".

Referências

- [1] Carlos Alberto Callioli, Hygino Hugueros Domingues, and Roberto Celso Fabrício Costa. *Álgebra linear e aplicações*. Atual, 2000.
- [2] Luiz Paulo Fávero and Patrícia Belfiore. *Manual de análise de dados: estatística e modelagem multivariada com Excel® , SPSS® e Stata®*. Elsevier Brasil, 2017.
- [3] Joseph F Hair, William C Black, Barry J Babin, Rolph E Anderson, and Ronald L Tatham. *Análise multivariada de dados*. Bookman editora, 2009.
- [4] Kenneth Hoffman and Ray Kunze. *Linear algebra*. Englewood Cliffs, New Jersey, 1971.
- [5] Mary Lilian Lourenço and Flávio Ulhoa Coelho. *Um Curso de Álgebra Linear*. EDUSP, 2001.
- [6] Bryan FJ Manly and Jorge A Navarro Alberto. *Métodos estatísticos multivariados: uma introdução*. Bookman Editora, 2008.
- [7] Pedro A Morettin and Wilton O Bussab. *Estatística básica*. Saraiva Educação, 2010.