

Programa de Disciplina Especialização em Ciência dos Dados

Módulo: II

Disciplina: Estatística Descritiva e Visualização de Informações

Carga Horária: 15 horas (8h teóricas, 7h práticas)

Ofertante: Departamento de Estatística – DEEST/UFOP

Objetivo:

Proporcionar aos alunos o primeiro contato com a ciência Estatística e seus principais conceitos. Instigar a análise crítica própria do pensamento estatístico, essencial para o uso consciente de ferramentas utilizadas aprendizado de máquina. Familiarizar os alunos com softwares para a análise de dados, os quais também são de uso comum em aprendizado de máquina.

Ementa:

Introdução à Estatística; A Estatística e o Trabalho Científico; Noções de Amostragem; Levantamento de dados; Representação Tabular e Gráfica; Distribuição de Frequências; Medidas de Posição (Tendência Central e Separatrizes); Medidas de Dispersão (Absoluta e Relativa); Medidas de Forma (Assimetria e Curtose); Medidas de associação. Outras Estratégias de Análise (Boxplot, gráficos de dispersão, etc).

Conteúdo Programático:

1. Introdução à Estatística:
 - a) História da Estatística e suas aplicações.
 - b) Alguns conceitos: população e amostra, variáveis e classificação de variáveis.
 - c) O método científico e as etapas do método estatístico.
2. Noções de Amostragem:
 - a) Conceitos básicos em amostragem.
 - b) Tipos de amostragem:
 - i. Não-probabilística e Probabilística.
 - ii. Algumas técnicas não-probabilísticas.
3. Levantamento de Dados:
 - a) Mensuração.

- b) Consistência.
 - c) Tabulação de dados.
4. Representação Tabular e Gráfica:
- a) Séries estatísticas.
 - b) Construção de tabelas.
 - c) Construção de gráficos.
 - d) Distribuição de Frequências e sua Representação gráfica.
5. Medidas de Posição:
- a) Médias, mediana, moda, dispersão (desvio padrão amostral, desvio médio absoluto, amplitude, entre outras), separatrizes (percentis) e de forma.
6. Medidas de associação:
- a) Coeficiente de Correlação de Pearson.
7. Outras Estratégias de Análise:
- a) Boxplot, gráficos de dispersão.

Metodologia:

Neste módulo será utilizado o processo de Ensino-Aprendizagem convencional. Será feita a exposição teórica do conteúdo com concomitante utilização do software adotado para a disciplina. O objetivo é a construção do conhecimento baseado na compreensão crítica das técnicas utilizadas na área do conhecimento do módulo. A cada tópico apresentado um trabalho prático (aprendizagem ativa baseada em problema) é proposto para fins de fixação do conteúdo ministrado.

Atividade Prática Proposta:

Após a apresentação teórica de cada um dos métodos os alunos deverão, de forma individual, implementá-lo e testá-lo. Ao fim do módulo os alunos (grupos de 4) deverão escolher um problema da organização para solução.

Softwares:

1. R (<<https://cran.r-project.org/bin/windows/base/>> , última versão);
2. R-Studio (<<https://www.rstudio.com/products/rstudio/download/>> , última versão);
3. Python (<<https://www.python.org/downloads/>> , última versão);

4. Anaconda (<<https://www.anaconda.com/download/#windows>> , última versão);
5. Microsoft Office (2010 ou superior).

Bibliografia:

ADLER, J. *R in a Nutshell*. 2. ed. Sebastopol, Califórnia: O'Reilly Media, 2012. 699 p. ISBN 978-144931-208-4.

BUSSAB, W.; MORETTIN, P. *Estatística Básica*. 9. ed. São Paulo: Saraiva, 2017. 568 p. ISBN 978-854722-022-8.

MAGALHÃES, N. M.; LIMA, A. C. P. *Noções de probabilidade e Estatística*. 7. ed. São Paulo: Edusp, 2007. 428 p. ISBN 978-853140-677-5.

MONTGOMERY, D. C.; RUNGER, G. *Estatística Aplicada e Probabilidade para Engenheiros*. 5. ed. Rio de Janeiro: LTC, 2012. 548 p. ISBN 978-852161-902-4.

SILVA, C. B.; COUTINHO, C. Q. S. O nascimento da estatística e sua relação com a teoria da probabilidade. *Integração*, São Paulo, USJT, n. 41, p. 191–196, Abr. Mai. Jun. 2005.

WIKHAM, H. *Create Elegant Data Visualisations Using the Grammar of Graphics*. 2020. Disponível em: <<https://www.rdocumentation.org/packages/ggplot2/versions/3.1.0>>. Acesso em: 15 abr. 2020.