



CENTRO UNIVERSITÁRIO RITTER DOS REIS

PÓS-GRADUAÇÃO EM DATA SCIENCE E BIG DATA

DISCIPLINA: *Análise de dados com R*

CARGA HORÁRIA: 36

PROFESSOR: Cristofer Weber

ANO/SEM: 2018/1

Horário das Aulas: *Segunda-feira (19h – 22h 30min)*

PLANO DE ENSINO

EMENTA:

Estatística Descritiva e Inferencial; Análise Exploratória de Dados; Testes paramétricos e não paramétricos; Análise de regressão (linear e GLM); Workflow para processos analíticos; transformação e manipulação de dados com R; Uso de diferentes fontes de dados para análise; Programação funcional; Organização de projetos de análise de dados; Reprodutibilidade.

OBJETIVOS:

- Aplicação de técnicas computacionais para análise de dados utilizando a linguagem e ambiente computacional R
- Aplicar melhores práticas para organização de dados e disponibilização de resultados de análises
- Capacitar o aluno para atuação na execução de projetos de análise de dados em empresas através do emprego de processos de análise de dados e princípios de engenharia de software

PROGRAMA:

- Conceitos, introdução ao R (tipos)
- Estatística Descritiva
- *Data Munging e Tidyverse*
- Estatística Inferencial
- *Data Science Workflow*
- Análise exploratória com gráficos
- Fontes de dados (arquivos, web)
- Testes
- Programação funcional, imutabilidade
- Análise de regressão (linear e GLM)
- Organização de código e dados
- Tópicos (a definir) ou complementação de conteúdo anterior
- Reprodutibilidade
- Execução de *scripts*

METODOLOGIA:

- Emprego da linguagem e ambiente computacional R, e suas bibliotecas, na análise de dados em problemas trazidos pelo professor;
- Aulas expositivas e práticas usando os computadores do laboratório de informática, para o uso

da linguagem R;

- Desenvolvimento de projetos de análise de dados de escopo aberto e fechado utilizando dados reais disponíveis na internet.

ATIVIDADES DISCENTES:

Os alunos desenvolverão os exercícios propostos pelo professor, atuarão um ou mais projetos de análise de dados (individuais e/ou em grupo), e contribuirão com os demais colegas na análise de textos trazidos pelo professor e/ou recomendados pelos próprios colegas.

PROCEDIMENTOS, INSTRUMENTOS E CRITÉRIOS DE AVALIAÇÃO:

A avaliação de desempenho do aluno será contínua, considerando um processo gradativo e cumulativo de construção do conhecimento através da participação em aula guiada por leituras, exercícios em aula e trabalhos práticos. A média final será dada da seguinte forma:

Projeto(s): 70%

Exercícios e participação guiada pelas leituras: 30%

BIBLIOGRAFIA RECOMENDADA:

Bibliografia Básica

1. Hadley Wickham; Garrett Grolemund (2017). **R for Data Science**. O'Reilly Media. ISBN 978-1-4919-1039-9. (Disponibilizado online: <http://r4ds.had.co.nz/>)

2. David M Diez; Christopher D Barr; Mine Çetinkaya-Rundel (2015). **OpenIntro Statistics (3rd edition)**. OpenIntro, Inc. ISBN 978-1-94345-003-9. (Disponibilizado online: <https://www.openintro.org/stat/textbook.php>)

Bibliografia Complementar

3. Robert Kabacoff (2015). **R in Action: Data Analysis and Graphics with R (2nd Edition)**. Manning. ISBN 978-1-61729-138-8. (Versão resumida: <http://www.statmethods.net/>)

4. Rob J Hyndman; George Athanasopoulos (2014). **Forecasting: principles and practice**. OTexts. ISBN 978-0-9875071-0-5. (Disponibilizado online: <https://www.otexts.org/fpp>)

5. Alex Reinhart (2015). **Statistics Done Wrong: The Woefully Complete Guide**. No Starch Press. ISBN 978-1-59327-620-1. (Versão resumida: <https://www.statisticsdonewrong.com/>)

6. Peter Bruce; Andrew Bruce (2017). **Practical Statistics for Data Scientists: 50 Essential Concepts**. O'Reilly Media. ISBN 978-1-4919-5293-1.

Bibliografia Sugerida para continuidade

7. Gareth James; Daniela Witten; Trevor Hastie (2014). **An Introduction to Statistical Learning: with Applications in R**. Springer New York. ISBN 978-1-4614-7137-0. (Disponibilizado online: <http://www-bcf.usc.edu/~gareth/ISL/>)