

Plano de Ensino

01. Dados de Identificação da Disciplina:

Semestre:	2025.2	Curso:	Estatística
Turma:	A	Código Componente:	IME0473
Componente:	TÓPICOS EM ESTATÍSTICA III	UA Responsável:	IME
Carga Horária:	96	UA Solicitante:	IME
Teórica/Prática:	64/32	EAD/PCC:	-/-
Horários:	246T56	Docente:	Prof(a) David Henriques Da Matta

02. Ementa:

Fundamentos de aprendizado estatístico: diferença entre aprendizagem supervisionada e não supervisionada; relação entre precisão e interpretabilidade de modelos. Regressão linear: simples, múltipla e extensões. Técnicas de classificação: regressão logística, análise discriminante e comparação entre métodos. Métodos de reamostragem: validação cruzada e trade-off entre viés e variância. Estratégias de seleção e regularização de modelos: subset selection, ridge e lasso. Modelos baseados em árvores: árvores de decisão, bagging, random forests e boosting. Support Vector Machine (SVM).

03. Programa:

1- Fundamentos do aprendizado estatístico: conceitos básicos de aprendizado estatístico; diferença entre aprendizado supervisionado e não supervisionado; relação entre precisão e interpretabilidade de modelos; introdução aos tipos de variáveis e estrutura dos dados; métricas de avaliação de modelos.

2-Regressão linear: regressão linear simples; regressão linear múltipla: seleção de variáveis e diagnóstico de modelo. Extensões: regressão por splines.

3-Técnicas de classificação: regressão logística; análise discriminante linear (LDA) e quadrática (QDA); comparação entre métodos de classificação (curvas ROC); implementação prática e interpretação dos resultados.

4- Métodos de reamostragem: validação cruzada k-fold; bootstrap; trade-off entre viés e variância; aplicação prática de métodos de reamostragem em seleção de modelos.

5- Seleção e regularização de modelos: técnicas de subset selection: forward, backward, stepwise; regularização: ridge regression, lasso regression; comparação entre métodos de seleção e regularização; implementação e interpretação dos resultados.

6- Modelos baseados em árvores: árvores de decisão: construção, poda, interpretação; bagging e random forests: conceitos e implementação; boosting: Gradient Boosted Trees; comparação entre modelos baseados em árvores e outros métodos.

7- Support Vector Machines (SVM): conceito de margem ótima e vetores de suporte; kernel trick: kernels lineares e não lineares; parâmetros de ajuste e interpretação; aplicações práticas em dados reais; comparação entre SVM e outros classificadores.

04. Cronograma:

1. Fundamentos do aprendizado estatístico (2 aulas).
2. Regressão linear (16 aulas).
3. Técnicas de classificação (18 aulas).
4. Métodos de reamostragem (12 aulas).
5. Seleção e regularização de modelos (16 aulas).
6. Modelos baseados em árvores (14 aulas).
7. Support Vector Machines (10 aulas).
8. Avaliações (8 aulas).

05. Objetivos Gerais:

Desenvolver uma compreensão teórico-prática das técnicas estatísticas aplicadas ao aprendizado de máquina (Machine Learning) e promover o domínio das ferramentas computacionais, capacitando os alunos a resolver problemas reais em diferentes contextos profissionais e de pesquisa.

06. Objetivos Específicos:

Ao final do curso, espera-se que os alunos sejam capazes de:

- Utilizar o software R para importar, organizar, limpar e processar dados;
- Aplicar métodos estatísticos de aprendizado de máquina para análise de dados, realizando inferências, previsões e classificações;
- Compreender as vantagens e limitações dos diferentes métodos abordados;
- Interpretar e comunicar os resultados de forma crítica e contextualizada;
- Desenvolver autonomia na busca e aplicação de novas técnicas e abordagens estatísticas.

07. Metodologia:

O curso será conduzido de modo que o aluno assuma papel ativo em seu processo de aprendizado, sendo responsável por construir e aprofundar o conhecimento sob a tutoria do docente. Além de introduzir conceitos e métodos fundamentais, o professor proporrá desafios que incentivem o desenvolvimento de competências essenciais para o enfrentamento de situações práticas e problemas que os alunos encontrarão após a graduação. As atividades incluirão aulas expositivas e dialogadas, uso de ferramentas computacionais (principalmente R), resolução de exercícios práticos, estudo de casos e discussão de artigos científicos relevantes. Será estimulada a autonomia dos alunos na exploração de conteúdos e no desenvolvimento contínuo de habilidades, promovendo uma integração entre teoria e prática.

08. Avaliações:

A avaliação do curso será contínua, baseada em apresentações desenvolvidas pelos próprios alunos ao longo do semestre. Cada apresentação terá nota atribuída de zero a dez e deverá contemplar tanto os aspectos conceituais quanto a aplicação prática das técnicas estudadas.

Os alunos serão organizados em grupos de, no máximo, três integrantes, sendo que a atribuição de notas será individual, levando em consideração a participação, o domínio do conteúdo e a qualidade da apresentação.

As datas das avaliações serão definidas em consenso com os alunos, de modo a respeitar o ritmo e o progresso das turmas. A nota final do curso corresponderá à média aritmética das notas obtidas nas apresentações realizadas ao longo do semestre.

Observações:

1. Essas datas poderão sofrer alterações.
2. Haverá prova substitutiva para o aluno que perder qualquer das atividades avaliativas, com ausência justificada e tenha solicitado uma segunda chamada em até 7 (sete) dias após a data de realização da avaliação.
3. O estudante poderá solicitar segunda chamada de avaliação de componentes curriculares à unidade acadêmica ou à unidade acadêmica especial responsável pelo componente curricular, de forma remota, até 7 (sete) dias após a data da realização da avaliação.
4. O aluno será aprovado se a média final (MF) for igual ou superior a 6 (seis) pontos e se tiver frequência igual ou superior a 0.75 da carga horária total da disciplina. Caso contrário será reprovado.
5. Os resultados serão disponibilizados via sistema SIGAA UFG.
6. As atividades supervisionadas mencionadas no Art. 16 do RGCG serão apresentadas pelo professor em sala de aula e supervisionadas no horário de atendimento da disciplina.

09. Bibliografia:

[1]: FARAWAY J. J., Linear Models With R, Chapman & Hall, 2004.

[2]: HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. The Elements of. Statistical Learning: Data Mining, Inference, and Prediction. Second Edition. February 2009.

[3]: LITTELL R., STROUP W. W., FREUND R., SAS for Linear Models, Wiley-SAS, 2002.

10. Bibliografia Complementar:

[1]: BERK, A. R. Statistical Learning from a Regression Perspective. Springer, 2008.

[2]: WICKHAM H., GROLEMUND, G. R for Data Science. O'Really, 2016.

[3]: SEARLE, S.R. Linear Models, 1997.

[4]: DALGAARD, P. Introductory Statistics with R. Springer, 2nd Edition 2008.

11. Livros Texto:

[1]: BERK, A. R. Statistical Learning from a Regression Perspective. Springer, 2008. (C1)

[2]: WICKHAM H., GROLEMUND, G. R for Data Science. O'Really, 2016. (C2)

[3]: HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. The Elements of. Statistical Learning: Data Mining, Inference, and Prediction. Second Edition. February 2009. (B2)

12. Horários:

Dia	Horário	Sala Distribuída
2 ^a	T5	105, CAB (24)
2 ^a	T6	105, CAB (24)
4 ^a	T5	105, CAB (24)
4 ^a	T6	105, CAB (24)
6 ^a	T5	105, CAB (24)
6 ^a	T6	105, CAB (24)

13. Horário de Atendimento do(a)s Professor(a):

1. Terça-feira das 17h às 18h.; sala 230; IME-UFG

2. Quinta-feira das 17h às 18h; Sala 230; IME-UFG

14. Professor(a):

David Henriques Da Matta. Email: dhmatta@ufg.br, IME

Prof(a) David Henriques Da Matta