



Disciplina (optativa): Jornada de Análise de Dados de Saúde utilizando Linguagem R

Carga: 15 horas (1 crédito)

Período (descrever): 09, 16, 23 e 30 de julho de 2024

Dia da semana: Terça-feira

Horário: 08h00 as 12h00

Professor responsável: Fernando Fonseca de Almeida e Val e Rubelmar Maia de Azevedo Cruz Neto

Instituição: UEA/PPGMT

E-mail: ffaval@gmail.com e rcruzneto@uea.edu.br

Público-alvo: Alunos de mestrado e doutorado do Programa de Pós-Graduação em Medicina Tropical (PPGMT) da Universidade do Estado do Amazonas (UEA) em convênio com a Fundação de Medicina Tropical Dr. Heitor Vieira Dourado (FMT-HVD). Alunos externos como alunos especiais.

Número de vagas: 40 vagas

Pré-requisitos (se houver):

Não há.

Dia e horário atualizado das atividades: Vide calendário do PPGMT em: <https://ppgmt.uea.edu.br/index.php/ppgmt-calendario-academico/>

Aula 1: Terça-feira - 09/07/24 - 08h00 às 12h00 (4h)

Aula 2: Terça-feira - 16/07/24 - 08h00 às 12h00 (4h)

Aula 3: Terça-feira - 23/07/24 - 08h00 às 12h00 (4h)

Aula 4: Terça-feira - 30/07/24 - 08h00 às 12h00 (4h)

Local das atividades:

Samsung OCEAN, Campus EST/UEA



Av. Darcy Vargas, 1.200 - Parque Dez de Novembro, Manaus - AM, 69050-020, Brasil

Ementa: A disciplina "Jornada de Análise de Dados de Saúde utilizando Linguagem R" é projetada para fornecer aos alunos de Mestrado e Doutorado do Programa de Pós-Graduação em Medicina Tropical (PPGMT) da Universidade do Estado do Amazonas (UEA), bem como alunos externos como alunos especiais, uma compreensão sólida e prática da análise de dados de saúde utilizando a linguagem de programação R. Esta disciplina, oferecida em parceria com a Fundação de Medicina Tropical Dr. Heitor Vieira Dourado (FMT-HVD), será ministrada no Samsung OCEAN, Campus EST/UEA, e terá uma carga horária total de 16 horas, distribuídas em quatro aulas de 4 horas cada.

- Introdução à Linguagem R, RStudio e Google Colab;
- Análise Exploratória de Dados;
- Visualização e Manipulação de Dados;
- Correlação entre Variáveis Numéricas;
- Associação entre Variáveis Categóricas;
- Técnicas de Agrupamento (Clustering);
- Introdução à Modelos de Regressão Linear Simples e Múltipla, Regressão de Dados de Contagem (Variável Dependente Numérica);
- Introdução à Modelos de Regressão Logística Binária e Multinomial (Variável Dependente Categórica);

Objetivos:

1. Introduzir a Linguagem R e Ferramentas Associadas:**
 - Familiarizar os alunos com a linguagem de programação R, a interface do RStudio e a utilização do Google Colab para análise de dados.
2. Desenvolver Habilidades em Análise Exploratória de Dados:**
 - Ensinar técnicas de preparação, importação, manipulação e limpeza de dados, preparando os alunos para uma análise eficaz e precisa.
3. Capacitar na Visualização de Dados:**
 - Introduzir a gramática de gráficos e o pacote ggplot2, permitindo aos alunos criar visualizações claras e informativas dos dados.
4. Explorar Métodos de Correlação e Associação:**
 - Capacitar os alunos para realizar análises de correlação entre variáveis numéricas e associação



entre variáveis categóricas, facilitando a compreensão das relações entre diferentes conjuntos de dados.

5. Introduzir Técnicas de Regressão e Agrupamento:**

- Apresentar modelos de regressão linear simples e múltipla, regressão de dados de contagem, e regressão logística binária e multinomial.
- Ensinar técnicas de agrupamento não hierárquico (k-means) e hierárquico, e a inclusão de grupos em modelos de regressão.

6. Aplicar Métodos de Seleção e Avaliação de Modelos:**

- Ensinar técnicas de seleção de variáveis preditoras, como o método stepwise, e a inclusão de variáveis dummies em modelos de regressão.
- Capacitar os alunos na avaliação de modelos de regressão utilizando diversas métricas.

7. Introduzir Modelos Avançados de Regressão:**

- Apresentar modelos de regressão para dados de contagem, como a regressão de Poisson e a regressão binomial negativa, além de modelos inflacionados de zeros e outros modelos avançados.

Ao final do curso, espera-se que os alunos estejam aptos a aplicar técnicas avançadas de análise de dados em suas pesquisas, contribuindo para a geração de conhecimento na área de saúde tropical. A disciplina enfatiza uma abordagem prática, com exemplos e exercícios aplicados que permitirão aos alunos consolidar seu aprendizado e desenvolver competências essenciais para suas carreiras acadêmicas e profissionais.

Método das atividades:

Os encontros serão organizados por meio de aulas expositivas presenciais e realização de atividades utilizando o software.

Frequência:

Os alunos deverão ter no mínimo 75% de frequência e aqueles que ultrapassarem esse limite de faltas estarão automaticamente reprovados na disciplina.

Método de avaliação:

A nota final será atribuída pela participação nas atividades propostas.

Conteúdo programático e cronograma:

Data	Horário	Assunto	Professores
09/07/24	08:00 12:00	<ul style="list-style-type: none"> - Introdução à Linguagem R, RStudio e Google Colab; - Análise Exploratória de Dados: Preparação e importação de dados, Manipulação e limpeza de dados; - Visualização Dados: Introdução à Gramática de Gráficos (Grammar of Graphics), plotagem de gráficos com o pacote ggplot2; - Correlação entre Variáveis Numéricas; - Comparação de Médias; - Associação entre Variáveis Categóricas; 	Fernando Val e Rubelmar Neto
16/07/24	08:00 12:00	<ul style="list-style-type: none"> - Regressão Linear Simples e Múltipla; - Métricas de avaliação de modelos de regressão; - Seleção de variáveis preditoras (Stepwise); - Inclusão de variáveis Dummies em modelos de regressão; - Regressão de Box-Cox; 	Fernando Val e Rubelmar Neto
23/07/24	08:00 12:00	<ul style="list-style-type: none"> - Análise de Agrupamentos Não Hierárquica (kmeans); - Análise de Agrupamentos Hierárquica; - Inclusão de grupos em modelos de regressão; - Análise dos Componentes Principais; - Regressão por Componentes Principais (PCR, Principal Component Regression); 	Fernando Val e Rubelmar Neto
30/07/24	08:00 12:00	<ul style="list-style-type: none"> - Regressão de Dados de Contagem: Poisson Regression e Negative Binomial Regression; - Regressão Logística e Multinomial; - Zero-Inflated Regression; - Introdução a outros modelos de regressão; 	Fernando Val e Rubelmar Neto



Bibliografia:

FÁVERO, L.; BELFIORE, P. **Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®**. GEN LTC, 2017.

FIELD, A. **Descobrendo a estatística com SPSS**. SP: Armed, 2009.

FIELD, A.; MILES, J.; FIELD, Z. **Discovering statistics using R**, 2012.

MARIO F. TRIOLA. **Introdução à Estatística**. 12ed. Rio de Janeiro: LTC, 2023.

MAUGERI, A. et al. Applying a hierarchical clustering on principal components approach to identify different patterns of the SARS-CoV-2 epidemic across Italian regions. **Scientific Reports 2021 11:1**, v. 11, n. 1, p. 1-9, 29 mar. 2021.

RICO-HESSE, R. Molecular evolution and distribution of dengue viruses type 1 and 2 in nature. **Virology**, v. 174, n. 2, p. 479-493, 1 fev. 1990.

WANG, C. et al. A study of the dengue epidemic and meteorological factors in Guangzhou, China, by using a zero-inflated poisson regression model. **Asia-Pacific Journal of Public Health**, v. 26, n. 1, p. 48-57, 11 jun. 2014.

YOUNESPOUR, S. et al. Evaluating Related Factors to the Number of Involved Lymph Nodes in Patients with Breast Cancer Using Zero-Inflated Negative Binomial Regression Model. **Journal of Biostatistics and Epidemiology**, v. 6, n. 4, p. 259-266, 9 mar. 2021.