



UNIVERSIDADE FEDERAL FLUMINENSE
PRÓ-REITORIA DE EXTENSÃO – PROEX
Núcleo de Estudos Empresariais e Sociais - NEES
Programa de Extensão em Ciência de Dados

1. IDENTIFICAÇÃO

➤ **Nome do Curso:**

Especialização em Ciência de Dados

➤ **Forma de Oferta:** Presencial

➤ **Coordenação**

Coordenador: Douglas Rodrigues Pinto

Vice-Coodenador: Marco Aurélio dos Santos Sanfins

2. DISCIPLINAS

ESPECIALIZAÇÃO EM FINANÇAS	
DISCIPLINA	CARGA HORÁRIA
MÓDULO I	
Análise Exploratória de Dados	60 horas
MÓDULO II	
Análise Estatística de Dados	60 horas
MÓDULO III	
Probabilidade e Introdução À Estatística	80 horas
MÓDULO IV	
Modelos Lineares	60 horas
MÓDULO V	
Modelos Lineares Generalizados	40 horas

Estudo Dirigido e Metodologia da Pesquisa Científica	60 horas
--	----------

3. EMENTAS

NOME: Probabilidade e Introdução à Estatística	C.H.: 80 horas
<p>Ementa</p> <ol style="list-style-type: none"> 1. Preliminar: Análise combinatória 2. Probabilidade <ol style="list-style-type: none"> a. Definição e propriedades b. Probabilidade em espaço amostrais finitos c. Probabilidade condicional, teorema de Bayes d. Independência 3. Variáveis aleatórias <ol style="list-style-type: none"> a. Definição b. Esperança e Variância c. Simulação de variáveis aleatórias: uniforme, exponencial, Bernoulli, binomial, normal 4. Leis dos grandes números e Teorema central do limite 5. Obtenção de dados <ol style="list-style-type: none"> a. Amostra e População b. Métodos de amostragem 6. Inferência com R <ol style="list-style-type: none"> a. Estimção pontual b. Intervalos de confiança: média, proporção e variância c. Teste de hipótese: média, proporção e variância 	
<p>Bibliografia</p> <ul style="list-style-type: none"> ● VERZANI J. Using R for introductory statistics, 1st edition. Chapman & Hall/CRC The R Series, 2005. ● PRABHANJAN N. T., SURESH R., B.G. MANJUNATH. A course in statistics with R, 1st edition. Wiley, 2016. ● PENG D. R. Exploratory Data Analysis with R, 1st edition. Leanpub Book, 2016. ● SHELDON R. Probabilidade: um curso moderno com aplicações, 8a edição. Bookman, 2010. 	

NOME: Análise Estatística de Dados	C.H.: 60 horas
<p>Ementa</p> <ol style="list-style-type: none"> 1. Inferência para duas populações 	

- a. Teste para comparação de duas médias
- b. Teste para comparação de duas proporções
- 2. Testes de hipóteses sequenciais**
 - a. Procedimento de testes sequenciais de uma cauda
 - b. Procedimento de testes sequenciais de duas caudas
- 3. Testes de ajustes de dados**
 - a. Testes de independência
 - b. Testes de homogeneidade
 - c. Testes de normalidade
- 4. Clusterização de dados**
- 5. Redução de dimensão**
- 6. Testes não paramétricos para análise de dados**
 - a. Teste de sinais: teste de uma amostra, teste de Posto com sinais de Wilcoxon, teste de duas amostras
 - b. Teste W de Mann-Whitney para duas amostras
 - c. Testes de sequências: Wald-Wolfowitz
 - d. Correlação de Postos de Spearman
- 7. Aplicação: problemas propostos**

Bibliografia

- PENG D. R. Exploratory Data Analysis with R, 1st edition. Leanpub Book, 2016.
- MINGOTI, S. A. Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada. Belo Horizonte, 2005.
- JOHNSON, R. A.; WICHERN, D. W. Applied multivariate statistical analysis. 6th. ed. Upper Saddle River, 2007.
- CASELLA, G. Inferência estatística. Cengage Learning, 2011.
- DALGAARD, P. Introductory Statistics with R. Springer, 2008.
- PRABHANJAN N. T., SURESH R., B.G. MANJUNATH. A course in statistics with R, 1st edition. Wiley, 2016.

NOME: Análise Exploratória de Dados

C.H.: 60 horas

Ementa

- 1. Apresentação e Introdução do R**
- 2. Objetos do R: números, vetores, matrizes, listas e data frame**
- 3. Operações básicas:**
 - a. if, for, while
 - b. Operados lógicos
- 4. Programação e funções em R**
- 5. Importando bancos de dados**
- 6. Manipulação de dados**

- 7. Medidas de posição**
 - a. Média, mediana e moda
- 8. Medidas de dispersão**
 - a. Variância e desvio padrão
- 9. Análise gráfica de dados**
 - a. Histograma
 - b. Gráfico de barras e gráfico de setores
 - c. Boxplots

Bibliografia

- VERZANI J. Using R for introductory statistics, 1st edition. Chapman & Hall/CRC The R Series, 2005.
- PENG D. R. R Programming for data Science, 1st edition. Leanpub Book, 2015.
- PRABHANJAN N. T., SURESH R., B.G. MANJUNATH. A course in statistics wiht R, 1st edition. Wiley, 2016.
- PENG D. R. Exploratory Data Analysis with R, 1st edition. Leanpub Book, 2016.

NOME: Modelos Lineares

C.H.: 60 horas

Ementa

- 1. Motivação: dados de altura de Francis Galton**
- 2. Preliminares:**
 - a. Gráfico de dispersão
 - b. Covariância e Correlação
- 3. Modelos de regressão linear simples:**
 - a. Definição
 - b. Estimação e interpretação dos parâmetros
 - c. Usando o modelo para predições
 - d. Inferência: testes de hipóteses, intervalo de confiança
 - e. Diagnóstico: análise de resíduos, coeficiente de determinação
- 4. Modelos de regressão linear Multivariada:**
 - a. Definição
 - b. Estimação e interpretação dos parâmetros
 - c. Usando o modelo para predições
 - d. Inferência: testes de hipóteses, intervalo de confiança
 - e. Diagnóstico: análise de resíduos, coeficiente de determinação, variável dummy
 - f. Seleção de modelo

Bibliografia

- CAFFO, B. Regression models for data science in R, 1st edition. Leanpub book, 2015.

- FARAWAY J. Linear models with R, second edition. Chapman & Hall/CRC Texts in Statistical Science, 2009.
- MONTGOMERY, D.C. e PECK, E. A., Introduction to Linear Regression Analysis, John Wiley, New York: 1982.

NOME: Modelos Lineares Generalizados	C.H.: 40 horas
Ementa	
<ol style="list-style-type: none"> 1. Modelos Lineares Generalizados: <ol style="list-style-type: none"> a. Definição b. Família Exponencial c. Estimaco e teste de hiptese d. Diagnstico e. Seleo de modelos 2. Anlise de dados binrios: <ol style="list-style-type: none"> a. Representaco do modelo de regresso logstica b. Estimaco e teste de hiptese c. Diagnstico e qualidade de ajuste 3. Anlise de dados de contagem: <ol style="list-style-type: none"> a. Representaco do modelo de regresso de Poisson b. Estimaco e teste de hiptese c. Diagnstico e qualidade de ajuste 	
Bibliografia	
<ul style="list-style-type: none"> ● CAFFO, B. Regression models for data science in R, 1st edition. Leanpub book, 2015. ● FARAWAY J. Extending the Linear Model with R, 1st edition. Chapman & Hall/CRC Texts in Statistical Science, ● 2006. ● DOBSON, A. J. An Introduction to Generalized Linear Models, Chapman and Hall, 2001. 	

NOME: Estudo Dirigido e Metodologia da Pesquisa Cientfica	C.H.: 60 horas
Ementa	
<p>Elaborao de proposta de trabalho cientfico e/ou tecnolgico, envolvendo temas abrangidos pelo curso.</p> <p>Orientaco na elaboraco do projeto de trabalho de estudo dirigido, realizada em conjunto com o professor</p>	

orientador, desde o levantamento e fichamento bibliográfico para fundamentação teórica até o desenvolvimento dos tópicos: introdução, objetivos, materiais e métodos, resultados esperados, cronograma e referências bibliográficas. Orientação da escrita de acordo com as normas de trabalhos acadêmicos da Universidade Federal Fluminense.

Bibliografia

- Individualmente especificada pelo orientador.