

Unidade Curricular	Análise Exploratória de Dados	Área Científica	Matemática
Licenciatura em	Engenharia Eletrotécnica e de Computadores	Escola	Escola Superior de Tecnologia e de Gestão de Bragança
Ano Letivo	2022/2023	Ano Curricular	2
Nível	1-2	Créditos ECTS	6.0
Tipo	Semestral	Semestre	1
Código	9112-742-2101-00-22		
Horas totais de trabalho	162	Horas de Contacto	T - TP 30 PL 30 TC - S - E - OT - O -

T - Ensino Teórico; TP - Teórico Prático; PL - Prático e Laboratorial; TC - Trabalho de Campo; S - Seminário; E - Estágio; OT - Orientação Tutórica; O - Outra

Nome(s) do(s) docente(s) Maria Clara Rodrigues Bento Vaz Fernandes

Resultados da aprendizagem e competências

No fim da unidade curricular o aluno deve ser capaz de:

1. Compreender o conceito de probabilidade e os seus axiomas e calcular probabilidades simples utilizando as leis básicas da probabilidade e do cálculo combinatório;
2. Compreender os conceitos básicos associados às variáveis aleatórias e efetuar cálculos elementares com base neles;
3. Conhecer e manipular, ao nível básico, as variáveis aleatórias mais comuns e utilizá-las para modelar situações simples;
4. Aplicar e interpretar as formas mais comuns de representar e sintetizar a informação contida num conjunto de dados;
5. Agrupar os dados de acordo com diferentes medidas de similitude (análise grupal);
6. Determinar estimativas pontuais e de intervalo para os parâmetros populacionais mais comuns;
7. Compreender e aplicar a metodologia do teste de hipóteses sobre os parâmetros populacionais mais comuns;
8. Aplicar o teste não paramétrico do Qui-quadrado na análise de dados qualitativos.

Pré-requisitos

Antes da unidade curricular o aluno deve ser capaz de:

1. Manipular conceitos matemáticos básicos
2. Usar uma linguagem de programação de alto nível

Conteúdo da unidade curricular

Introdução; Teoria elementar da probabilidade. Variáveis aleatórias e distribuições de probabilidade. Amostragem aleatória e Distribuições Amostrais. Pré-processamento, organização e representação dos dados. Classificação e agrupamento da informação: Análise grupal (Cluster Analysis). Estimação pontual e estimação por Intervalo. Testes de hipóteses paramétricos e não paramétricos.

Conteúdo da unidade curricular (versão detalhada)

1. Introdução
 - Objeto da estatística;
 - Estatística descritiva e inferência estatística;
 - População e amostra.
2. Teoria Elementar da Probabilidade
 - Experiências aleatórias, espaços amostrais e acontecimentos;
 - Análise combinatória;
 - O conceito de probabilidade;
 - Probabilidade condicional;
 - Acontecimentos independentes;
 - Teorema de Bayes.
3. Variáveis Aleatórias e Distribuições de Probabilidade
 - Definição de variável aleatória;
 - Variáveis discretas: função de probabilidade e função de distribuição;
 - Variáveis contínuas: função densidade de probabilidade e função de distribuição;
 - Parâmetros das distribuições;
 - Variáveis transformadas: transformações lineares e não lineares.
4. Distribuições Conjuntas de Probabilidade
 - Definição de distribuição conjunta;
 - Distribuições marginais;
 - Distribuições condicionais;
 - Independência;
 - Covariância e correlação;
 - Combinação de variáveis;
 - Cálculo do valor esperado e da variância de combinações de variáveis: comb. lineares e não lineares
5. Caracterização de Algumas Distribuições Discretas
 - Distribuição Binomial;
 - Distribuição Binomial Negativa;
 - Distribuição Hipergeométrica;
 - Relação entre as distribuições Binomial e Hipergeométrica;
 - Distribuição de Poisson;
 - Relação entre a distribuição de Poisson e as distribuições Binomial e Hipergeométrica.
6. Caracterização de Algumas Distribuições Contínuas
 - Distribuição Uniforme;
 - Distribuição Exponencial;
 - Relação entre as distribuições de Poisson e Exponencial;
 - Distribuição Normal;
 - Distribuição Normal padronizada;
 - Combinação linear de variáveis normais independentes;
 - Relação entre as distribuições Normal e Binomial;
 - Distribuições do Qui-Quadrado, t de Student e F.
7. Amostragem Aleatória e Distribuições Amostrais
 - Distribuição da média amostral;
 - Valor esperado e variância da média amostral;
 - Forma da distribuição da média amostral quando a população é Normal;
 - Teorema do Limite Central;
 - Implicações na caracterização de distribuições por amostragem;
 - Geração de amostras recorrendo ao Método de Monte Carlo.
8. Estatística Descritiva
 - Classificação dos dados;
 - Dados qualitativos e quantitativos;
 - Caracterização de amostras univariadas;

Conteúdo da unidade curricular (versão detalhada)

- Estatísticas de localização (média, mediana e moda) e dispersão (variância);
 - Coeficiente de assimetria e coeficiente de kurtose;
 - Caracterização de amostras bivariadas;
 - Ajuste de uma relação linear recorrendo ao método dos mínimos quadrados;
 - Coeficiente de correlação e coeficiente de determinação.
9. Análise Grupal (Cluster Analysis)
- Objetivo da análise grupal;
 - Medidas de similitude;
 - Classificação hierárquica;
 - Classificação não hierárquica.
10. Estimação Pontual
- Estimadores e estimativas;
 - Propriedades desejáveis dos estimadores pontuais: não enviesamento, eficiência e consistência;
 - Métodos de estimação: método da máxima verosimilhança e método dos mínimos quadrados;
 - Seleção de um estimador.
11. Estimação por Intervalo
- O conceito de intervalo de confiança;
 - Especificação de intervalos de confiança;
 - Intervalo de confiança para o valor esperado de uma população contínua;
 - Intervalo de confiança da proporção binomial: amostras de grande e de pequena dimensão;
 - Intervalo de confiança para a variância de uma população Normal;
 - Intervalo de confiança para a diferença entre os valores esperados de duas populações contínuas;
 - Intervalo de confiança para a diferença entre proporções binomiais com amostras de grande dimensão;
 - Intervalo de confiança para a razão entre variâncias de populações Normais;
 - Dimensionamento de amostras.
12. Testes de Hipóteses
- Análise do procedimento básico envolvido no teste de hipóteses;
 - Definição das hipóteses;
 - Identificação da estatística de teste e caracterização da sua distribuição;
 - Definição da regra de decisão;
 - Erro do tipo I e especificação do nível de significância;
 - Cálculo da estatística de teste e tomada de decisão;
 - Valor de Prova;
 - Erro do tipo II e potência do teste;
 - Relação entre testes de hipóteses e intervalos de confiança;
 - Referência aos testes de hipóteses mais comuns
13. Testes Não Paramétricos
- Classificação dos testes de utilização mais frequente;
 - Teste do Qui-quadrado para dados qualitativos.

Bibliografia recomendada

1. Rui Guimarães, Sarsfield Cabral - Estatística - 2ª edição, Verlag Dashofer, 2010.
2. William Navidi - Statistics for Engineers & Scientists - 5th edition, McGraw-Hill Education, 2020.
3. Alberto Leon-Garcia - Probability, Statistics, and Random Processes for Electrical Engineering - 3rd edition, Pearson Prentice Hall, 2008.
4. Ronald K. Pearson - Exploratory Data Analysis Using R - Chapman and Hall/CRC, 2020.
5. Subhash Sharma - Applied Multivariate Techniques - John Wiley & Sons, 1996.

Métodos de ensino e de aprendizagem

Os conteúdos apresentados serão abordados em ambiente presencial, em regime teórico-prático, acompanhados da resolução de exercícios. Em horário não presencial os tópicos serão explorados por meio de exercícios de aplicação e recurso a ferramentas informáticas a desenvolver em R. Realizar-se-ão sessões tutoriais em horário não presencial, sempre que necessário, individuais ou de grupo.

Alternativas de avaliação

1. Exame Final - (Ordinário, Trabalhador) (Final, Recurso, Especial)
 - Exame Final Escrito - 100%
2. Alternativa 2 - (Ordinário, Trabalhador) (Final)
 - Prova Intercalar Escrita - 40% (A realizar no período de aulas.)
 - Exame Final Escrito - 40% (A realizar no final do semestre)
 - Trabalhos Práticos - 20%

Língua em que é ministrada

Português, com apoio em inglês para alunos estrangeiros

Validação Eletrónica

Maria Clara Rodrigues Bento Vaz Fernandes	António Jorge da Silva Trindade Duarte	Orlando Manuel de Castro Ferreira Soares	Paulo Alexandre Vara Alves
07-10-2022	11-10-2022	21-10-2022	03-11-2022