



UNIVERSIDADE PRESBITERIANA MACKENZIE
Faculdade de Computação e Informática



Componente Curricular: exclusivo de curso (X) Eixo Comum () Eixo Universal ()		
Curso: CIÊNCIA DA COMPUTAÇÃO		Núcleo Temático: FUNDAMENTOS DE COMPUTAÇÃO
Disciplina: ANÁLISE DE DADOS		Código da Componente: ENEX50016
Carga horária: 06 ha	(04) Sala de Aula (02) Laboratório	Etapas: 2ª.
Professores: Alcides Teixeira Barboza Junior Ismar Frango Silveira Orlando Bisacchi Coelho Péricles do Prado Turnes Junior		DRTs: 1123792
Carga horária: 76 Sala de Aula, 38 Laboratório, 0 EaD		Semestre Letivo: 2ºSEM/2018
Ementa: Análise Exploratória de Dados: variáveis quantitativas e categóricas; gráficos para variáveis quantitativas; gráficos para variáveis categóricas; medidas de posição; medidas de dispersão; curvas de densidade; distribuições normais; correlações entre variáveis; regressão; distribuições marginais. Inferência: populações e amostras; técnicas de obtenção de amostras aleatórias; probabilidade; distribuições amostrais; testes de hipótese; inferência sobre a média populacional; inferência sobre proporções populacionais; comparação de proporções. Prática: uso de módulos Python para análise exploratória de dados e inferência.		



Objetivos:		
Fatos e Conceitos	Procedimentos e Habilidades	Atitudes, Normas e Valores
Conhecer as principais técnicas de Estatística e Probabilidade. Travar contato com uma série de aplicações dessas técnicas a casos de aplicação concretos. Ter contato com o ecossistema computacional contemporâneo, baseado em Python, para Análise de Dados.	Desenvolver a capacidade de reconhecer oportunidades de aplicação das técnicas de Estatística e Probabilidade a situações reais. Ser capaz de modelar um dado problema de forma a torná-lo tratável através de métodos e técnicas de Probabilidade e Estatística, identificando as abordagens que podem eventualmente ser usadas na sua resolução. Saber usar o ecossistema computacional de Análise de Dados. Estar apto a implementar protótipos de soluções baseadas nessas técnicas computacionais.	Estar atento para identificar oportunidades de resolução de problemas de pesquisa e desenvolvimento e do dia a dia das organizações por meio de técnicas de Estatística e Probabilidade. Perceber a importância da Análise Exploratória.



Conteúdo Programático:

- 1. O ecossistema Python para Análise de Dados.**
- 2. Análise Exploratória de Dados.**
 - 2.1. Variáveis quantitativas e categóricas.
 - 2.2. Gráficos para variáveis categóricas.
 - 2.3. Gráficos para variáveis quantitativas.
 - 2.4. Medidas de posição.
 - 2.5. Medidas de dispersão.
 - 2.6. Correlações entre variáveis.
 - 2.7. Curvas de densidade.
 - 2.8. Distribuições normais.
 - 2.9. Regressão.
 - 2.10. Distribuições marginais.
- 3. Inferência.**
 - 3.1. Populações e amostras.
 - 3.2. Técnicas de obtenção de amostras aleatórias;.
 - 3.3. Probabilidade.
 - 3.4. Modelos probabilísticos discretos e contínuos.
 - 3.5. Distribuições amostrais.
 - 3.6. Testes de hipótese.
 - 3.7. Inferência sobre a média populacional.
 - 3.8. Inferência sobre proporções populacionais.
 - 3.9. Comparação de proporções.

Metodologia:

- Aulas expositivas introduzindo os conceitos, associadas a mecanismos de aprendizagem ativa em que os alunos aprofundam o entendimento dos conceitos.
- Utilização de material multimídia ilustrando conceitos e aplicações concretas da Estatística.
- Utilização de estudos de caso e de arquivos de dados associados a esses casos.
- Atividades em classe, individuais ou em pequenos grupos.
- Desenvolvimento de projetos computacionais, em pequenos grupos de alunos.
- Utilização da linguagem de programação Python e das bibliotecas abertas NumPy, Matplotlib e Pandas.
- Utilização do ambiente Mackenzie Virtual.



Critério de Avaliação:

O cálculo da avaliação final é feito por meio da seguinte fórmula:

$$N1 = 0,7 * P1 + 0,15 * ATIV1 + 0,15 * PROJ1$$

$$N2 = 0,7 * P2 + 0,15 * ATIV2 + 0,15 * PROJ2$$

$$MI = (N1 + N2) / 2 + NP$$

Se $FREQ < 65\%$, então o aluno está **REPROVADO**.

Se $65\% \leq FREQ < 75\%$ e $MI < 8.5$, então o aluno está **REPROVADO**.

Se $65\% \leq FREQ < 75\%$ e $MI \geq 8.5$, então o aluno está **APROVADO** e a Média Final (MF) é obtida por $MF = MI$.

Se $FREQ \geq 75\%$ e $MI \geq 7.5$, então o aluno está **APROVADO** e a Média Final (MF) é obtida por $MF = MI$.

Se $FREQ \geq 75\%$ e $MI < 7.5$, então o aluno poderá fazer a **PROVA FINAL (PF)**; nesse caso, a Média Final (MF) é obtida por $MF = (MI + PF) / 2 + NP$.

Se $MF \geq 6.0$, aluno está **APROVADO**. Caso contrário, está **REPROVADO**.

Onde:

- **P1** (nota A no Sistema de Notas) é uma nota obtida a partir de prova escrita, individual e sem consulta;
- **ATIV1** (nota B no Sistema de Notas) é a média aritmética simples obtida a partir das atividades realizadas em sala de aula ao longo da disciplina, individualmente ou em pequenos grupos;
- **LAB1** (nota C no Sistema de Notas) é a nota obtida a partir das atividades computacionais realizadas em laboratório ao longo da disciplina, individualmente ou em pequenos grupos;
- **P2** (nota D no Sistema de Notas) é uma nota obtida a partir de prova escrita, individual e sem consulta;
- **ATIV2** (nota E no Sistema de Notas) é a média aritmética simples obtida a partir das atividades realizadas em sala de aula ao longo da disciplina, individualmente ou em pequenos grupos;
- **LAB2** (nota F no Sistema de Notas) é a nota obtida a partir das atividades computacionais realizadas em laboratório ao longo da disciplina, individualmente ou em pequenos grupos;
- **N1** é a nota intermediária do 1º. bimestre;
- **N2** é a nota intermediária do 2º. bimestre;
- **NP** é a nota de participação, a qual fica a critério de cada professor da disciplina, podendo variar entre 0 (zero) e 1 (um);
- **FREQ** é a frequências nas atividades tal como controlado pelo Sistema de Frequência;
- **MI** é a média intermediária;
- **PF** (nota PF no Sistema de Notas) é a prova final, escrita e individual, cobrindo toda a matéria dada no semestre e
- **NF** é a nota final.

OBS 1.: Ao final do semestre, antes das Provas Finais, o aluno poderá fazer, caso o queira, uma **Prova Substitutiva**. Essa prova será escrita e contemplará todo o assunto coberto ao longo da disciplina. A nota da Prova Substitutiva (**SUB**) substituirá a menor nota intermediária bimestral (N1 ou N2) no cálculo da média intermediária MI; se as notas dos dois bimestres for a mesma, será substituída a nota do bimestre de maior peso.

OBS. 2: O Prof. Orlando não atribui Nota de Participação. Portanto, NP=0.



Bibliografia Básica:

- MOORE, David S.; NOTZ, William I.; FLINGER, Michael A. **A Estatística Básica e sua Prática**. 7 ed. Rio de Janeiro: LTC, 2017.
- PHUONG, Vo. T.H.; CZYGAN, Martin. **Getting Started with Python Data Analysis**. Birmingham: Packt Publishing, 2015.
- VANDERPLAS, Jake. **Python Data Science Handbook**. Sebastopol: O'Reilly, 2017.

Bibliografia Complementar:

- BUSSAB, Wilton O.; MORETTIN, Pedro A. **Estatística básica**. 8 ed. São Paulo: Saraiva, 2013.
- COSTA NETO, Pedro. Luiz O. **Estatística**. 2 ed. São Paulo: E. Blücher, 2002.
- DOWNING, Downing; CLARK, Jeffrey. **Estatística aplicada**. 3 ed. São Paulo: Saraiva, 2011.
- LEVINE, David M.; STEPHAN, David F.; SZABAT, Kathryn A. **Estatística: teoria e aplicações usando Microsoft Excel em português**. 7 ed. Rio de Janeiro: LTC, 2016.
- TRIOLA, Mario F. **Introdução à Estatística: Atualização da Tecnologia**. 11 ed. Rio de Janeiro: LTC, 2013.

Semana	Planejamento de Aula
	Conteúdo
01	Apresentação da disciplina. O ecossistema Python para Análise de Dados. Conceito de Ciência de Dados.
02	Variáveis quantitativas e categóricas. Gráficos para variáveis categóricas. Gráficos para variáveis quantitativas.
03	Medidas de posição. Medidas de dispersão. Correlações entre variáveis.
04	Curvas de densidade. Distribuições normais.
05	Regressão.
06	Distribuições marginais.
07	Prova P1.
08	Populações e amostras. Técnicas de obtenção de amostras aleatórias;.
09	Probabilidade. Modelos probabilísticos discretos e contínuos.
10	Semana da FCI.
11	Independência. Probabilidade Condicional.
12	Distribuições amostrais.
13	Testes de hipótese.
14	Inferência sobre a média populacional.
15	Inferência sobre proporções populacionais.
16	Comparação de proporções.
17	Prova P2.
18	Prova PF.
19	Vista de Notas.