

Nome da Disciplina: Ciência de Dados

Código: ECN XXX

Carga horária: 30 horas/aula

Créditos: 2

Curso: Ciências Econômicas

Pré-requisito: Não há

Período Letivo: 2º Semestre de 2023

Horários: 15:50 – 17:30 (quarta-feira)

Professor: Emerson G. Dos Santos (emerson.gomes@unifesp.br)

Rafael S. M. Ribeiro (rsmribeiro@cedeplar.ufmg.br)

EMENTA

1-Introdução a Ciência de Dados em R e Python e aplicações: Conceitos gerais (Business Intelligence, Business Analytics, Big Data, entre outros tópicos).

2- Uso de R e Python para análise de dados: Planejamento (Abordagens para fundamentar as escolhas sobre os dados e técnicas a serem utilizadas), Coleta e Tratamento de dados (Organização e lógica para lidar com os dados), Técnicas de Análise de dados (formulação e interpretação de: Análise de Regressão e Classificação, Análise de Componentes Principais e Fatorial, Análise de Agrupamento, Simulação, Previsão, entre outras técnicas), Comunicação e apresentação (Abordagens para divulgação das informações e resultados das análises).

PROGRAMA

- 1) Introdução a Ciência de Dados em R e Python;
- 2) Planejamento das análises, Coleta e Tratamento de dados;
- 3) Uso de técnicas de Análise de dados em aplicações diversas;
- 4) Comunicação e apresentação;

BIBLIOGRAFIA

MORETTIN, P. A.; BUSSAB, W. O. Estatística básica. 9 ed. São Paulo: Saraiva, 2017. 554 p.

DEVORE, J. L. Probabilidade e estatística para engenharia e ciências. 9 ed. Cengage Learning Brasil, 2019. 630p.

Hair Jr., J. F., Anderson, R. E., Tatham, R. C., Black, W. C. (2005). Análise Multivariada de Dados. 5ª. edição. Porto Alegre: Bookman.

Johnson, R. A., Wichern, D. W. (2002). Applied Multivariate Statistical Analysis. 5th edition. Upper-Saddle River: Prentice Hall.

Manly, B. J. F. (2008). Métodos Estatísticos Multivariados. 3ª. edição. Porto Alegre: Bookman.

Corrar, L. J., Filho, J. M. D., Paulo, E. (2014). Análise Multivariada: para os Cursos de Administração, Ciências Contábeis e Economia. 1ª. edição. São Paulo: Atlas.

Fávero, L. P.; Belfiore, P. Silva, F.L.; Chan, B.L. (2009). Análise de dados – modelagem multivariada para tomada de decisões. Rio de Janeiro: Elsevier.

Assunção, R. (2017). Fundamentos Estatísticos de Ciência dos Dados. Disponível em: <https://homepages.dcc.ufmg.br/~assuncao/EstatCC/FECD.pdf>

R for Data Science (<https://r4ds.had.co.nz/>)

MCKINNEY, W. Python for Data Analysis: Data Wrangling with Pandas, Numpy, and IPython. O'Reilly, 2017

Severance, C. R.; Python para Todos: Explorando Dados com Python 3, IEEE-UFCG.

Grus, J.; Data Science do zero: Primeiras regras com o Python; Alta Books (2016)

Outras referências serão apresentadas conforme o desenvolvimento dos tópicos e serão apresentadas nos materiais extras durante o curso

AVALIAÇÃO

A avaliação do curso consistirá em:

1. Exercícios e participação diária: 20 pontos
2. Avaliação/prova: 40 pontos
3. Trabalho final com análise de dados: 40 pontos

CRONOGRAMA

Horas	Data	Dia	Tema
2	17/ago	qui	Apresentação geral da disciplina e aula introdutória
4	24/ago	qui	Introdução ao R (nivelamento)
6	31/ago	qui	Introdução a linguagem Python (nivelamento)
8	14/set	qui	Análise de Regressão com R e Python
10	21/set	qui	Análise de Classificação com R e Python
12	28/set	qui	Análise de Componentes Principais e Fatorial com R e Python
14	05/out	qui	Análise de Agrupamentos com R e Python
16	19/out	qui	Previsão (Séries temporais) com R e Python
18	26/out	qui	Introdução à Simulação com R e Python
20	09/nov	qui	Apresentação parcial dos trabalhos (20 pontos)
22	16/jnov	qui	Discussões para melhorias das análises
24	23/nov	qui	Introdução a outras técnicas de análise de interesse
26	30/nov	qui	Revisão e plantão de dúvidas para prova e apresentação
28	07/dez	qui	Prova (40 pontos)
30	14/dez	qui	Apresentação final dos trabalhos (20 pontos)
