

DISCIPLINA: Estatística multivariada

CÓDIGO:

NATUREZA: Optativa

NÚMERO DE CRÉDITOS: 03

CARGA HORÁRIA: 45 horas

NÍVEL: Mestrado e Doutorado

HORÁRIO: 9:00 -12:00 e 15:00 - 17:00 h

DOCENTE: Denis Conrado da Cruz

NÚMERO DE VAGAS: 10

TIPO DE COMPONENTE: (X) DISCIPLINA

(X) TÓPICOS ESPECIAIS

EMENTA

1. Fundamentação básica à análise multivariada; 2. Processo inicial para as análises estatística multivariada; 3. Distribuição Normal Multivariada; 4. Testes de Hipóteses para o Vetor de Médias (Harold Hotelling); 5. Análise da estrutura de covariância (ANCOVA); 6. Análise de componentes principais; 7. Análise da variância multivariada (MANOVA); 8. Análise fatorial (inferência sobre a estrutura das matrizes de covariância); 9. Análise discriminante; 10. Análise de classificação (Método de Fisher); 11. Análise de correspondência; 12. Análise de agrupamentos (Clusters); e 13. Escalonamento Multidimensional.

CONTEÚDO PROGRAMÁTICO

OBJETIVOS

A disciplina tem por objetivo proporcionar ao estudante de pós-graduação conhecimentos e habilidades para a utilização de técnicas multivariadas nas análises e interpretações de dados ecológicos/botânicos de diversas áreas de estudos. Dessa forma, auxiliar os estudantes na análise conjunta de dados multivariados, para que possam responder e concluir a respeito de suas perguntas e hipóteses iniciais de pesquisa. O estudante será capaz de analisar qualitativamente por meio de técnicas multivariadas, variáveis categóricas ou do tipo de medida nominal, ordinal, intervalar ou proporcional, com a utilização das técnicas apropriadas e interpretação dos resultados de forma coerente.

JUSTIFICATIVA:

As ciências não são ou estão consolidadas em afirmações absolutas, mas sim na possibilidade de testar ideias que provam algo, sendo assim as análises estatísticas, quando bem aplicada, são fundamentais para agregar confiabilidade aos resultados dos estudos. As análises estatísticas são executadas após a coleta de dados, para auxiliar nas análises e interpretações, mas antes de iniciar a coleta de dados é necessário delinear o estudo, e pensar quais análises serão executadas. Por isso, não basta ter uma pergunta interessante, é recomendado que antes de iniciar um estudo você deve identificar quais são suas hipóteses em relação a pergunta em questão, para obter respostas confiáveis e coerente condizentes ao seu objetivo em questão. Por essa razão, existem fatores que devam ser chaves antes de iniciar qualquer análise estatística: *i)* Qual é o objeto de seu estudo? (variável dependente); *ii)* Quais variáveis influenciam a variável dependente e as relações entre as variáveis independentes?; *iii)* Quais as variabilidades nos dados e a magnitude dos efeitos?; *iv)* Qual o tamanho, forma, orientação e distribuição de suas unidades amostrais?; *v)* A amostragem está na mesma escala que as questões?; entre outras. Nesse curso será estudado

UNIVERSIDADE FEDERAL RURAL DA AMAZÔNIA
MUSEU PARAENSE EMÍLIO GOELDI

as análises multivariadas nas abordagens ambientais de pesquisas, principalmente ecológicas e botânicas.

A análise multivariada é uma ótima ferramenta estatística aplicável às distintas atividades de pesquisas científicas do conhecimento ambiental. Nos dias atuais, é importante na aplicação desta técnica a utilização de Softwares (Phyton, R, SPSS, SAS, MATLAB, entre outros), para que a pesquisa seja executada de forma rápida, efetiva e satisfatória. Cada vez mais as pesquisas ecológicas requerem profissionais com um bom nível de conhecimento em estatística e programação ligados ao estudo da fauna, flora e áreas afins. Assim, esta disciplina visa contribuir para o desempenho técnico-científico e acadêmico dos Recursos Humanos ligados a investigação científica sobre as mais distintas áreas de conhecimentos ambientais.

PROGRAMAÇÃO DA DISCIPLINA:

Data	Tema	Professor
04/10	MANHÃ - 9:00 -12:00 h ✓ Apresentação da disciplina; ✓ Fundamentação básica à análise multivariada; ✓ Aplicabilidade da análise multivariada;	Denis Conrado da Cruz
	TARDE- 15:00 - 17:00 h ✓ Processo inicial para as análises estatística multivariada; ✓ Distribuição Normal Multivariada;	
05/10	MANHÃ - 9:00 -12:00 h ✓ Testes de Hipóteses para o Vetor de Médias (Harold Hotelling); ✓ Análise da estrutura de covariância (ANCOVA); ✓ Exercícios - casos de estudos (Prático – Software R);	Denis Conrado da Cruz
	TARDE- 15:00 - 17:00 h ✓ Análise de componentes principais ✓ Análise da variância multivariada (MANOVA); ✓ Exercícios - casos de estudos (Prático – Software R);	
06/10	MANHÃ - 9:00 -12:00 h ✓ Análise fatorial (inferência sobre a estrutura das matrizes de covariância); ✓ Análise discriminante; ✓ Exercícios - casos de estudos (Prático – Software R);	Denis Conrado da Cruz
	TARDE- 15:00 - 17:00 h ✓ Análise de classificação (Método de Fisher); ✓ Exercícios - casos de estudos (Prático – Software R);	
07/10	MANHÃ - 9:00 -12:00 h ✓ Análise de correspondência; ✓ Exercícios - casos de estudos (Prático – Software R);	Denis Conrado da Cruz
	TARDE- 15:00 - 17:00 h ✓ Análise de agrupamentos (Clusters); ✓ Exercícios - casos de estudos (Prático – Software R);	
08/10	MANHÃ - 9:00 -12:00 h ✓ Escalonamento Multidimensional. ✓ Exercícios - casos de estudos (Prático – Software R);	Denis Conrado da Cruz

UNIVERSIDADE FEDERAL RURAL DA AMAZÔNIA
MUSEU PARAENSE EMÍLIO GOELDI

	TARDE ✓ Revisão geral e aplicações de exercícios para auxílio na avaliação final;	
11/10	MANHÃ /TARDE ✓ Avaliação final - Parte 1 – apresentação de exemplo prático;	-
18/10	Entrega da atividade proposta – Estudo de caso ecológico e/ou botânico utilizando as mais diversas análises multivariadas estudadas (Parte 2).	-

AVALIAÇÃO:

A avaliação dos estudantes será pontuada de acordo com os seguintes critérios: *i)* Presença nas aulas (Peso – 10%); *ii)* Participação do estudante nas aulas, com respeito aos temas abordados em classe (Peso – 20%); *iii)* Apresentação de um exercício em aula – Parte 1 (Peso – 30%); *iv)* Elaboração de um estudo de caso abordando as análises multivariadas, entregue após uma semana do termino da disciplina – Parte 2 (Peso – 40%).

MATERIAL DIDÁTICO NECESSÁRIO

- ✓ **Google Meet.**
- ✓ **Software solicitado para a disciplina:** Instalação do RStudio Desktop 1.4.1103, disponível: <https://rstudio.com/products/rstudio/download/#download>
- ✓ **Instalação dos seguintes pacotes:**
 - > install.packages("vegan");
 - > install.packages("mcmcPack");
 - > install.packages("ggplot2");
 - > install.packages("betapart");
 - > install.packages("Simpsons");
 - > install.packages("mclust");
 - > install.packages("diveRsity");
 - > install.packages("dplyr");
 - > install.packages("mclust");
 - > install.packages("coronabr");
 - > install.packages("graphics");
 - > install.packages("ggfortify");
 - > install.packages("lifecycle");
 - > install.packages("car");
 - > install.packages("carData");
 - > install.packages("Rcpp");
 - > install.packages("RcppEigen");
 - > install.packages("heplots");
 - > install.packages("psych");
 - > install.packages("biotools");
 - > install.packages("SpatialEpi");
 - > install.packages("MASS");
 - > install.packages("factoextra");
 - > install.packages("ade4TkGUI");
 - > install.packages("scales");
 - > install.packages("ISLR");

OBS: outros pacotes R poderão ser solicitados no decorrer da disciplina.

REFERENCIAS BIBLIOGRÁFICAS

- ANDERSON, T. W. **An Introduction to Multivariate Statistical Analysis**. 2ª ed. New York: John Wiley & Sons, 1984.
- ARBOLINO, Roberta et al. Towards a sustainable industrial ecology: Implementation of a novel approach in the performance evaluation of Italian regions. **Journal of Cleaner Production**, v. 178, p. 220-236, 2018.
- BAROUCHE, J; SAPORTA G. **Análise de dados**. Rio de Janeiro: Zahar editores. 1980.
- BUSSAB, W.; O. MIAZAKI, E. S.; ANDRADE, D. F. **Introdução à Análise de Agrupamentos**. 9º SINAPÉ. São Paulo. 1990.
- CARDOSO, ANA. As Estratégias de Amostragem e a Inferência Estatística. **ANAIS do Instituto Hidrográfico**, p. 117, 2005.
- DOABI, SHAHAB AHMADI; AFYUNI, MAJID; KARAMI, MAHIN. Multivariate statistical analysis of heavy metals contamination in atmospheric dust of Kermanshah province, western Iran, during the spring and summer 2013. **Journal of Geochemical Exploration**, v. 180, p. 61-70, 2017.
- EVERITT, B. S. **Graphical Techniques for Multivariate Data**. London: Heinemann Educational Books, 1978.
- GREENACRE, Michael; BLASIUS, Jorg (Ed.). **Multiple correspondence analysis and related methods**. CRC press, 2006.
- GREENACRE, M. J. Theory and Applications of Correspondence Analysis. New York: Academic Press, 1984.
- JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. New Jersey: Englewood Cliffs, 1998. MORRISON, D. F. Multivariate Statistical Methods. McGraw-Hill, 1976.
- HAIR, A. et al. **Análise Multivariada de dados**. Porto Alegre: Artmed. 2005.
- HERRANDO PÉREZ, SALVADOR. **Manual de ecología matemática un enfoque práctico al análisis multivariado (PCA, CLUSTER y MDS) para detectar patrones en ecología**. 2002.
- HOU, DEYI et al. Integrated GIS and multivariate statistical analysis for regional scale assessment of heavy metal soil contamination: A critical review. **Environmental Pollution**, v. 231, p. 1188-1200, 2017.
- JOHNSON, Richard Arnold et al. **Applied multivariate statistical analysis**. Upper Saddle River, NJ: Prentice hall, 2002.
- JOHNSON, RICHARD A.; WICHERN, Dean V. **Applied Multivariate Statistical Analysis**. 2014.
- LANDIM, P. M. B. **Análise estatística de dados geológicos multivariados**. Oficina de Textos, 2011.
- KHATTREE R.; NAIK, D. **Multivariate data reduction and discrimination – with SAS software**. SAS USA: Institute Inc. North Caroline. 2000.
- MARDIA, K.V.; KENT, J.T.; BIBBY, J.M. **Multivariate analysis**. London: Academic,. 1979.
- PLA, LAURA E. **Análisis multivariado: método de componentes principales**. OEA, Washington, DC (EUA). Secretaría General. Programa Regional de Desarrollo Científico y Tecnológico, 1986.
- PEREIRA, J. C. R. **Análise de dados qualitativos: estratégias metodológicas para as ciências da saúde humanas e sociais**. Edusp, 1999.
- REIS, ELIZABETH. **Estatística multivariada aplicada**. Edições Sílabo, 2001.
- SEBER, GEORGE AF. **Multivariate observations**. John Wiley & Sons, 2009.
- TEFAS, Anastasios; PITAS, Ioannis. **Principal Component Analysis**. Intelligent Systems, 2018.
- ZOU, HUI; HASTIE, TREVOR; TIBSHIRANI, Robert. Sparse principal component analysis. **Journal of computational and graphical statistics**, v. 15, n. 2, p. 265-286, 2006.