

IBI 5086: Introdução a Métodos Estatísticos para a Bioinformática
PROGRAMA

2º Semestre/2020

Professora: Júlia Maria Pavan Soler (pavan@ime.usp.br)

Período: 27/08 a 19/11/2019 **Horário:** Quintas (8:00 - 12:00h)

Plataforma de Apoio: e-Disciplinas (Moodle da USP)

Monitor: Lucas Ramalho Anderson

Objetivo: Introduzir conceitos e métodos de Estatística na análise de dados com aplicações em Bioinformática.

Conteúdo:

1. Álgebra linear básica: cálculo matricial, sistemas lineares, espaços vetoriais, produto interno, norma, ortogonalidade, determinantes, decomposição em valores singulares.
2. Modelos estatísticos: estrutura de dados, modelos lineares (ANOVA e regressão), comparação de médias (correções para múltiplos testes), modelos lineares generalizados (tabelas de contingência, testes qui-quadrado, modelos de regressão logística).
3. Análise Multivariada de Dados: componentes principais, análise discriminante, agrupamento, MANOVA, correlação canônica.
4. Intervalos de Confiança e Testes baseados em Bootstrap e Permutação
7. Simulação de Monte Carlo

Recursos:

Aulas não presenciais, síncronas, com apoio do Google Meet e e-Disciplinas.

Forma de avaliação:

Listas de exercícios em grupos de até 5 alunos (30%), projeto individual (35%) e prova (35%).

Bibliografia Básica:

1. Estatística básica:

Bussab, W.O.; Morettin, P.A. (2017). Estatística Básica. Editora Saraiva.

Magalhães, M.N.; de Lima, A.C.P. (2013). Noções de Probabilidade e Estatística. EDUSP.

2. Álgebra Linear

Strang, G. (2010) Álgebra linear e suas aplicações - Ed. Cengage Learning.

Coelho, F.U.; Lourenço, M.L. (2001). Um curso de álgebra linear - EDUSP.

3. Planejamento de Experimentos, Modelos ANOVA, Dados categorizados

Box, GEP; Hunter, WG; Hunter, JS. (1978) Statistics for Experimenters. John Wiley & Sons (ver 2th ed., 2006).

Neter, J; Kutner, MH; Li, W; Nachtsheim, JC. (2005) Applied Linear Statistical Models. 5th ed., Boston: McGraw-Hill.

Oehlert, GW. (2010) A first course in Design and Analysis of Experiments. Univ. of Minnesota, Licensed by Creative Commons.

Giolo, S. R. (2017). *Introdução à Análise de Dados Categóricos com Aplicações*. Blucher, ABE-Projeto Fisher.

4. Análise Multivariada

Everitt, B.S. (2005). *An R and S-Plus Companion to Multivariate Analysis*. Springer.

Johnson, R.A. e Wichern, D.W. (2002). *Applied Multivariate Statistical Analysis*. Prentice Hall.

Manly, B.F.J. (2005). *Multivariate Statistical Methods. A primer*. Chapman & Hall/CRC.

5. Métodos computacionalmente intensivos

Liu, J.S. (2004). *Monte Carlo methods in scientific computing*. Springer.

Manly, B.F.J. (2007). *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Chapman & Hall/CRC.

Ross, S. (1997). *Simulation*. 2nd Ed., New York. Academic Press.

6. Conteúdos adicionais:

James, G; Witten, D; Hastie, T; Tibshirani, R. (2015). *An Introduction to Statistical Learning*. Springer.

Irizarry, R.A; Love, M.I. (2015). *Data Analysis for the Life Sciences*.

<http://leanpub.com/dataanalysisforthelifesciences>.

7. Recursos computacionais

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.Rproject.org/>.