

**PROGRAMA DE PÓS GRADUAÇÃO ASSOCIADO EM BIOINFORMÁTICA
PGAB003 - MÉTODOS ESTATÍSTICOS AVANÇADOS EM BIOINFORMÁTICA**

Professoras responsáveis:

Profa. Dra. Elisângela Ap. da Silva Lizzi

Profa. Dra. Gláucia Maria Bressan

Período: 28/04/2026 a 30/06/2026 (Condensada em carga horária)

Carga horária: 60h

EMENTA AMPLIADA

1. Introdução à disciplina

- 1.1. Apresentação geral da área e objetivos
- 1.2. Estratégias de operacionalização dos conteúdos
- 1.3. Integração entre estatística, computação e aplicações biológicas
- 1.4. Visão geral das ferramentas

2. Técnicas de Permutação e Reamostragem

- 2.1. Fundamentos de testes não paramétricos
- 2.2. Testes de Aleatorização e Testes de Permutação
- 2.3. Métodos de reamostragem e estimação de incerteza
- 2.4. Jackknife: princípios e aplicações
- 2.5. Bootstrap: variantes, propriedades e usos em bioinformática
- 2.6. Validação cruzada: estratégias, particionamento e avaliação de modelos
- 2.7. Aplicações computacionais em dados de alta dimensionalidade

3. AMOVA – Análise Molecular de Variância

- 3.1. Conceitos básicos de variância molecular
- 3.2. Estruturas hierárquicas de dados genéticos
- 3.3. Matrizes de distância: construção e escolha do critério
- 3.4. Partição da variância e estatísticas derivadas (Φ -statistics)
- 3.5. Formulação do modelo estatístico da AMOVA
- 3.6. Teste de hipóteses e significância
- 3.7. Métodos de validação e abordagens por permutação
- 3.8. Aplicações em genética de populações e filogeografia

4. Modelos Lineares Generalizados (GLM)

- 4.1. Revisão do conceito de modelos lineares e suas limitações
- 4.2. Família exponencial e função de ligação
- 4.3. Modelo de regressão binomial e a extensão multinomial
- 4.4. Modelo de regressão de Poisson para dados de contagem
- 4.5. Estimação dos parâmetros e interpretação dos coeficientes
- 4.6. Resíduos, medidas de ajuste e diagnóstico do modelo
- 4.7. Aplicações em transcriptômica, metagenômica e dados epidemiológicos

5. Redes Bayesianas – Fundamentos e Algoritmos

- 5.1. Conceitos fundamentais: variáveis, dependências condicionais e grafos direcionados

- 5.2. Representação de conhecimento e inferência em redes estáticas
- 5.3. Algoritmos de Aprendizagem Baseados em Restrições (Constraint-Based)
- 5.4. Algoritmos de Aprendizagem Baseados em Pontuação (Score-Based)
- 5.5. Algoritmos Híbridos e heurísticas combinadas
- 5.6. Critérios de avaliação e validação de modelos
- 5.7. Aplicações em genética, redes regulatórias e integração multi-ômica

6. Aplicações Práticas e Discussão de Literatura Científica

- 6.1. Estudos de caso utilizando dados biológicos reais
- 6.2. Implementação computacional dos métodos estudados
- 6.3. Análise e interpretação crítica de artigos científicos
- 6.4. Boas práticas de reporte, reprodutibilidade e comunicação científica

CONTEÚDO PROGRAMÁTICO

28/04: Tópico 1 – Introdução e Integração

- **Síncrono (4h):** Apresentação da disciplina e visão geral das ferramentas computacionais. Integração Estatística-Computação-Biologia.

30/04: Tópico 1 – Introdução e Integração

- **Síncrono (4h):** Revisão de ambiente R e pacotes específicos para a disciplina.

05/05: Tópico 2 – Permutação e Reamostragem

- **Síncrono (4h):** Fundamentos de testes de aleatorização. Jackknife, Bootstrap e Validação Cruzada. Aplicações em alta dimensionalidade.

07/05: Tópico 2 – Permutação e Reamostragem

- **Síncrono (4h):** Implementação de algoritmos de reamostragem em datasets biológicos.

12/05: Tópico 3 – AMOVA: Análise Molecular de Variância (Dia Único)

- **Síncrono (4h):** Conceitos de variância molecular, matrizes de distância, partição da variância (Φ -statistics) e testes de hipóteses.

14/05: Tópico 3 – AMOVA: Análise Molecular de Variância (Dia Único)

- **Síncrono (4h):** Prática com dados de genética de populações e filogeografia.

19/05: Tópico 4 – Modelos Lineares Generalizados (Parte 1)

- **Síncrono (4h):** Família exponencial e funções de ligação. Regressão Binomial e Multinomial.

21/05: Tópico 4 – Modelos Lineares Generalizados (Parte 1)

- **Presencial (4h):** Discussão guiada de exercícios aplicados de interpretação de coeficientes e odds ratio.

26/05: Tópico 4 – Modelos Lineares Generalizados (Parte 2)

- **Síncrono (4h):** Regressão de Poisson para contagem. Diagnóstico de modelo e resíduos. Aplicações em Transcriptômica e Metagenômica.

28/05: Tópico 4 – Modelos Lineares Generalizados (Parte 2)

- **Síncrono (4h):** Ajuste de modelos GLM e tratamento de superdispersão (overdispersion)

02/06: Tópico 5 – Redes Bayesianas (Parte 1)

- **Síncrono (4h):** Conceitos fundamentais: DAGs, dependências condicionais e inferência em redes estáticas.

04/06: Tópico 5 – Redes Bayesianas (Parte 1)

- **Síncrono (4h):** Estudo de grafos e representação de conhecimento.

09/06: Tópico 5 – Redes Bayesianas (Parte 2)

- **Síncrono (2h):** Aprendizagem de estrutura: Algoritmos baseados em Restrição (Constraint-based) e Pontuação (Score-based).

11/06: Tópico 5 – Redes Bayesianas (Parte 2)

- **Síncrono (2h):** Algoritmos Híbridos e heurísticas.

16/06: Tópico 5 – Redes Bayesianas (Parte 3)

- **Síncrono (4h):** Validação de modelos e aplicações em bioinformática

18/06: Tópico 5 – Redes Bayesianas (Parte 3)

- **Síncrono (4h):** Prática avançada com o pacote bnlearn.

23/06: Tópico 6 – Aplicações Práticas e Estudos de Caso

- **Síncrono (4h):** Implementação computacional integrada com dados reais (estudos de caso).

30/06: Tópico 6 – Discussão de Literatura e Encerramento

- **Síncrono (4h):** Preparação para a análise crítica de artigos científicos.

02/07: Tópico 6 – Discussão de Literatura e Encerramento

- **Presencial (4h):** Análise crítica de artigos, discussão sobre reprodutibilidade e apresentação dos resultados finais.

RESUMO DA CARGA HORÁRIA (72H/aula):

- **Encontros Síncronos/Presenciais:** 10 encontros x 4h = 40h
- **Atividades Síncronas/Práticas:** 32h
- **Total:** 72h