

Curso de Especialização em *Data Science* e Estatística Aplicada

Módulo IV - **Análise de Dados Categorizados**

Prof. Dr. Márcio Augusto Ferreira Rodrigues

21/05/2025

Instruções

- O desenvolvimento desta atividade deve ser realizada de forma individual ou em dupla.
- Deve-se completar o arquivo Rmd enviado na atividade.
- É necessário devolver o arquivo em Rmd e em pdf.
- Valor da atividade: 10 pontos.

Descrição da atividade

Um estudo desenvolvido pela Profa. Denise Gonçalves, do Departamento de Otorrinolaringologia da UFMG, teve como interesse a ocorrência de manifestações otorrinolaringológicas em pacientes HIV positivos. Neste estudo, 112 pacientes foram acompanhados no período de março de 1993 a fevereiro de 1995, sendo 91 HIV positivo e 21 HIV negativo. A classificação quanto à infecção pelo HIV seguiu os critérios Do *Center for Disease Control* (CDC, 1987), sendo ela: HIV soronegativo (não possui o HIV), HIV soropositivo assintomático (possui o vírus mas não desenvolveu o quadro clínico de AIDS), com ARC (*Aids Related Complex*: apresenta baixa imunidade e outros indicadores clínicos que antecedem o quadro clínico de AIDS), ou AIDS (apresenta infecções oportunistas que definem AIDS).

As Covariáveis medidas no estudo são:

- **id** : Idade do Paciente (medida em anos);
- **sex**: Sexo do Paciente (0 se Masculino e 1 se Feminino)
- **grp**: Grupo de Risco (1 se HIV Soronegativo, 2 se HIV Soropositivo Assintomático, 3 se ARC e 4 se AIDS)
- **ats**: Atividade Sexual (1 se Homossexual, 2 se Bissexual e 3 se Heterossexual)
- **ud**: Uso de Droga Injetável (1 se Sim e 2 se Não)
- **ac**: Uso de Cocaína por Aspiração (1 se Sim e 2 se Não)

Questão 1 Faça a leitura do conjunto de dados *aids.txt* e formate as variáveis **sex**, **grp**, **ats**, **ud** e **ac** para fator.

Questão 2 Faça as seguintes análises:

- (i) Construa a tabela de contigência entre as variáveis **sex** e **ud**.

- (ii) Verifique se existe associação entre as variáveis **sex** e **ud**.
 - (iii) Determine a razão de chances (Odds ratio) entre as variáveis **sex** e **ud**. Interprete os resultados.
-

Questão 3 Faça as seguintes análises:

- (i) Construa a tabela de contingência entre as variáveis **ud** e **grp**.
 - (ii) Verifique se existe associação entre as variáveis **ud** e **grp**.
 - (iii) Determine a razão de chances (Odds ratio) entre as variáveis **sex** e **ud**. Interprete os resultados.
-

Questão 4 Utilizando as covariáveis **id**, **sex**, **grp**, **ats**, **ud** e **ac**, e o modelo de regressão logística, responda:

- (i) Faça o ajuste do modelo de regressão logística, considerando a variável **cens** como variável resposta e todos os efeitos principais de todas as covariáveis.
 - (ii) Avalie a qualidade do modelo ajustado
 - (iii) Avalie a predição do modelo ajustado utilizando a curva ROC.
 - (iv) interprete os resultados do modelo ajustado.
-

Gabarito

Questão 1

Questão 2

- (i)
 - (ii)
 - (iii)
-

Questão 3

- (i)
 - (ii)
 - (iii)
-

Questão 4

(i)

(ii)

(iii)

(iv)