

Prova Prática  
Pontuação total: 10  
**Prazo: 23/10/2025 - A ser marcado em aula**

Nome:	Matrícula:
Nome:	Matrícula:

A prova deve ser feita por no máximo dois alunos.

Na resolução da prova use as funções geradoras de dados (`gerar_dados`, `gerar_dados_rl` e `gerar_tdf`), todas disponíveis no arquivo [gerar\\_dados\\_v11.R](#) na [página da disciplina](#).

A função `gerar_dados` foi programada para gerar uma amostra aleatória estratificada de pessoas do município X. As variáveis aleatórias de interesse são: `Y1` (medido em `un`), `Y2` (medido em `un`) e `Sexo`. Adicionalmente, assuma que `Y1` e `Y2` não podem assumir valores reais negativos.

A função `gerar_tdf` foi programada para gerar uma tabela de distribuição de frequências do tipo comum, dessas que se encontra em publicações. Considere que esta tabela descreve uma análise de interesse (já publicada) e que é necessário determinar as medidas estatísticas básicas com finalidades de entendimento e comparações.

A função `gerar_dados_rl` gera uma amostra de um estudo sobre influência de uma variável fixa `X` (medido em `un`) sobre uma variável aleatória `Y` (medido em `un.dia-1`).

Todos os dados gerados são fictícios e tem finalidades exclusivamente didáticas para fins de avaliação prática em análise de dados.

Realizar a análise dos dados com respostas às questões propostas.

## 1 AED: Apresentações tabulares e gráficas (2.0)

Considere os dados gerados pela função `gerar_dados` para a questão.

### 1.1 Diagrama de caixa (boxplot) para `Y1` e `Y2` (1.0)

1. (0.5) Antes e após a eliminação de possíveis outliers<sup>1</sup>;
2. (0.5) Após a eliminação de possíveis outliers<sup>2</sup>.

### 1.2 Para `Y1` (1.0)

1. (0.5) Uma apresentação tabular contendo apenas as freqüências: absoluta (`Fi`), relativa (`Fr, %`) e acumulada (`Fac, %`), nessa ordem<sup>2</sup>;
2. (0.5) Histograma e o polígono de freqüência acumulada dos dados<sup>2</sup>.

## 2 AED: Medidas estatísticas básicas (3.0)

### 2.1 AED: Medidas determinadas a partir dos vetores (1.5)

Considere os dados gerados pela função `gerar_dados` para a questão.

Para as variáveis `Y1` e `Y2` elaborar apresentações tabulares<sup>2</sup> contendo as seguintes estimativas:

1. (0.5) Tendência central: média, mediana e moda;
2. (0.5) Posição: quartis e decis;
3. (0.5) Dispersão: amplitude total, variância, desvio padrão e coeficiente de variação.

<sup>1</sup>Não distinguindo sexo

<sup>2</sup>Para cada sexo: M seguido de F

## 2.2 AED: Medidas determinadas a partir de apresentações tabulares (1.5)

Considere os dados gerados pela função `gerar_dados_tdf` para a questão.

Elabore uma apresentação tabular contendo:

1. (0.5) Tendência central: média, mediana e moda;
2. (0.5) Posição: quartis e decis;
3. (0.5) Dispersão: amplitude total, variância, desvio padrão e coeficiente de variação.

## 3 AED: Medidas estatísticas de associação (1.5)

Considere os dados gerados pela função `gerar_dados` para a questão.

1. (0.5) Estimativas: covariância e correlação linear simples<sup>2</sup>;
2. (0.5) Diagramas de dispersão dos dados<sup>2,3</sup>;
3. (0.5) Um estudo semelhante foi realizado em um outro município, por outras pessoas. Contudo, as unidades de medida usadas foram: Y1 (100 \* un) e Y2 (100 \* un).

Para comparar associações entre as variáveis de ambos os estudos, qual seria a medida estatística recomendada? Justifique.

## 4 AED: Regressão linear (2.5)

Considere os dados gerados pela função `gerar_dados_rl` para a questão.

1. (1.0) Ajuste aos dados dois modelos de regressão linear: polinômios de grau I e II (ambos não forçado para a origem) e apresente as análises dos modelos;
2. (1.0) Apresente diagramas de dispersão dos dados<sup>4</sup> com os modelos ajustados.
3. (0.5) Com base nos gráficos e na análise da regressão, qual modelo melhor explica o fenômeno em estudo? Justifique com fundamentação estatística.

## 5 Contextualização (1.0)

Localize um artigo científico (periódico Qualis A ou B) - em área de seu interesse - no qual a análise exploratória de dados (AED - possivelmente com medidas de associação e uso de regressão linear como modelo explicativo) teve papel preponderante. Discuta o artigo com ênfase nos recursos da AED usados e também na adequação das normas básicas das apresentações gráficas e tabulares adotada pelo periódico.

## 6 Bônus (1.0)

Pelos critérios de ajustamento e escolha de modelos vistos em aula, os coeficientes de determinação ( $r^2$ ) de modelos lineares ajustados (forçados e não forçados para a origem) são comparáveis? Justifique com fundamentação estatística.

**Observações:**

- Informe com clareza na prova os números de matrícula usados para possibilitar a adequada correção computacional;
- As normas para apresentações gráficas e tabulares são obrigatórias, serão observadas e corrigidas;
- Sugere-se (mas não é obrigatório) o uso do ambiente R na resolução das questões propostas;
- Cada hora de atraso na entrega da avaliação implica na perda de 25%. Portanto, após 4 horas não entregue.

<sup>3</sup>Considere Y2 no eixo das ordenadas e Y1 no eixo das abscissas

<sup>4</sup>Considere Y no eixo das ordenadas e X no eixo das abscissas