

# Trabalho da disciplina de Probabilidade e Estatística

*Estatística Descritiva: medidas de tendência central e dispersão; correlação linear; regressão linear*

**Individual ou em dupla**

1. Cada grupo deverá selecionar um **conjunto de dados** para análise.

O conjunto de dados escolhido deve seguir os seguintes critérios:

No mínimo, **100 observações (linhas)**.

Incluir pelo menos **5 variáveis** (colunas), das quais:

**3 variáveis contínuas** (exemplo: idade, altura, renda, temperatura).

**2 variáveis categóricas** (exemplo: gênero, tipo de produto, setor econômico, região geográfica).

*Exemplos de fontes para encontrar conjuntos de dados nos links postados no classroom:*

<https://classroom.google.com/c/Njg1NTYwMzg1MTA1/m/Njg1OTI4OTk0NzZz/details>

2. A partir do conjunto de dados escolhido, o grupo deverá realizar as seguintes análises:

a) **Estatísticas Descritivas:**

- **Média:** Calcular a média para todas as variáveis contínuas.
- **Mediana:** Calcular a mediana para as variáveis contínuas.
- **Moda:** Identificar a moda, principalmente para as variáveis categóricas e, se aplicável, para variáveis contínuas.
- **Desvio Padrão e Variância:** Calcular o desvio padrão e a variância das variáveis contínuas.
- Para cada variável analisada, apresentar no relatório seus dados obtidos.

b) **Distribuição dos Dados:**

- **Histograma:** Criar um histograma para pelo menos uma das variáveis contínuas, demonstrando a distribuição dos valores.
- **Boxplot:** Gerar um boxplot para uma das variáveis contínuas, mostrando a dispersão dos dados e possíveis *outliers*.
- Dê títulos aos gráficos, informações relevantes e legendas

c) **Correlação entre Variáveis:**

- Calcular a **correlação** entre duas variáveis contínuas, utilizando o **coeficiente de correlação de Pearson**.
- Criar um **gráfico de dispersão (scatter plot)** para mostrar a relação entre essas duas variáveis.

- Interpretar os resultados (escrever o que o grupo entende que os gráficos estão indicando)

d) **Regressão Linear:**

- Aplicar um modelo de **regressão linear simples** utilizando duas variáveis contínuas, sendo uma a variável independente (X) e a outra a variável dependente (Y).
- Interpretar o coeficiente angular (beta) e o coeficiente de determinação ( $R^2$ ) para avaliar o ajuste do modelo.
- Gerar um **gráfico de dispersão com a reta ajustada** da regressão linear.

3. Após a realização das análises, o grupo deve organizar suas conclusões em um relatório.

a) **Descrição do Conjunto de Dados:**

- Breve **descrição das variáveis** (quais são numéricas [quantitativas] e quais são categóricas [qualitativas]) e do **contexto** (se são dados de saúde, econômicos, esportivos, etc.).
- Indicação da **fonte** dos dados.

b) **Resultados das Análises:**

- **Tabelas** com os resultados das estatísticas descritivas (média, mediana, moda, desvio padrão e variância).
- **Gráficos** gerados (histograma, boxplot, gráfico de dispersão e gráfico da reta de regressão linear).
- **Interpretação dos resultados**, incluindo o que os gráficos e as estatísticas descritivas revelam sobre os dados.

c) **Discussão da Correlação e Regressão:**

- Discussão dos resultados da correlação: a relação entre as variáveis é **positiva**, **negativa** ou **nula**? **Quão forte** é a correlação?
- Interpretação do modelo de regressão linear: a variável independente (X) consegue **explicar** a variável dependente (Y)? O que o **valor de  $R^2$**  indica sobre o ajuste do modelo?

d) **Conclusão:**

- Resumo final dos resultados obtidos a partir da análise.
- Quais padrões ou tendências foram identificados? Como esses resultados podem ser úteis para decisões futuras?