

Programação para Engenharia II

Análise Exploratória e Testes Estatísticos em MATLAB

Sumário

- 1 Importação e Pré-processamento de Dados
- 2 Exploração de Dados
- 3 Testes Estatísticos
- 4 Modelagem e Previsão
- 5 Boas Práticas e Fontes de Dados
- 6 Bibliografia

Importação e Pré- processamento de Dados

Information

Antes de realizar qualquer análise, é fundamental **importar e inspecionar os dados**. Isso garante que entendemos seu formato, número de linhas e presença de valores faltantes.

Objetivo: carregar dados externos no **MATLAB** para posterior análise.

```
1 % Lista de arquivos CSV na pasta atual
2 files = dir('*.csv'); disp('Arquivos CSV disponíveis:'); disp({files.name}');
```

Código 1: Listagem de arquivos CSV disponíveis.

```
1 data = readtable('temperaturas.csv');  
2 plot(data.Dia, data.Temperatura, '-o');  
3 xlabel('Dia');  
4 ylabel('Temperatura (°C)');  
5 title('Temperatura Diária');  
6 grid on;
```

Código 2: Leitura de dados de um CSV e visualização inicial.

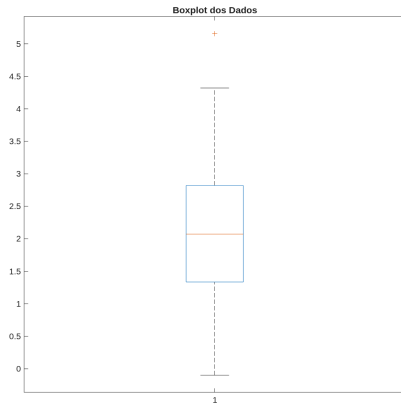
Exploração de Dados

Nota:

Estatísticas descritivas são o **primeiro passo** para entender o comportamento dos dados: média, mediana e desvio padrão resumem a distribuição.

```
1 % Estatísticas básicas
2 data = randn(100,1);
3 media = mean(data); mediana = median(data); desvio = std(data);
4 fprintf('Média=%.2f, Mediana=%.2f, Desvio=%.2f\\n',media,mediana,desvio);
```

Código 3: Cálculo de média, mediana e desvio padrão.



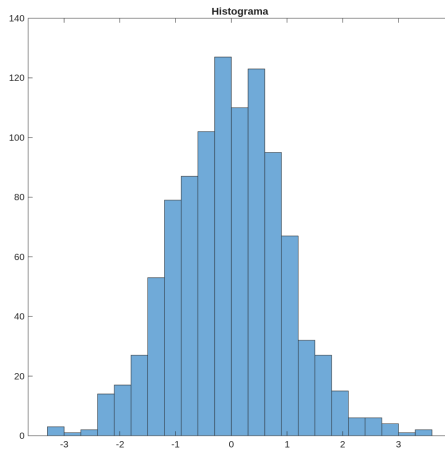
0

histograma é usado para verificar a **distribuição de frequência**. Ideal para identificar assimetrias e presença de outliers.

```
1 % Histograma
2 data = randn(1000,1);
3 figure; histogram(data,20); title('Histograma'); xlabel('Valores'); ylabel('Frequência');
4     grid on;
5 saveas(gcf,'figures/plot_histogram.png');
```

Código 4: Distribuição de frequência de uma variável.

Histograma e Distribuição II



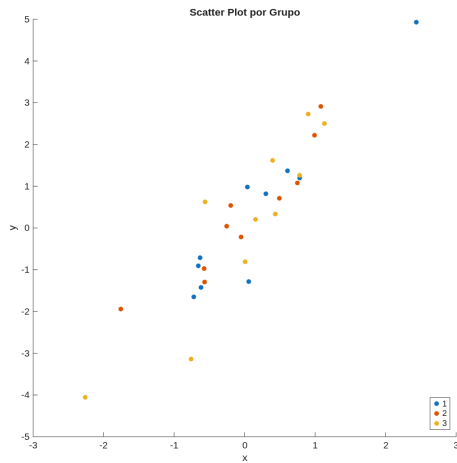
Example

Útil para investigar relações lineares ou não-lineares entre duas variáveis. Aqui usamos cores diferentes para cada grupo.

```
1 % Scatter plot 2D colorido por grupo
2 x = randn(30,1); y = 2*x + randn(30,1);
3 grupo = [ones(10,1); 2*ones(10,1); 3*ones(10,1)];
4 figure;
5 gscatter(x,y,grupo);
6 xlabel('x'); ylabel('y'); title('Scatter Plot por Grupo');
7 saveas(gcf,'figures/plot_scatter2.png');
```

Código 5: Relação entre duas variáveis com cores por grupo.

Scatter Plot 2D II

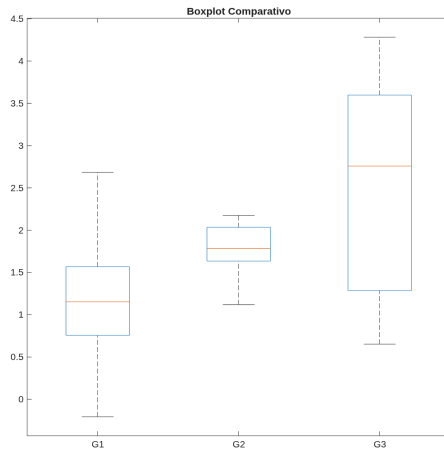


- O boxplot é excelente para **comparar grupos**;
- Exibe medianas, quartis e valores atípicos.

```
1 % Boxplot comparativo entre 3 grupos
2 grupo1 = randn(10,1)+1;
3 grupo2 = randn(10,1)+2;
4 grupo3 = randn(10,1)+3;
5 figure;
6 boxplot([grupo1;grupo2;grupo3], ...
7         [repmat({'G1'},10,1); repmat({'G2'},10,1); repmat({'G3'},10,1)]);
8 title('Boxplot Comparativo');
9 saveas(gcf,'figures/plot_boxplot.png');
```

Código 6: Boxplot comparando grupos.

Boxplot Comparativo II



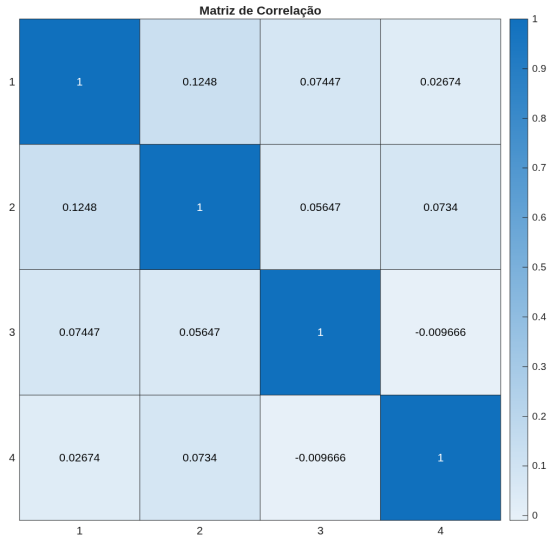
Information

A matriz de correlação ajuda a detectar relações lineares entre variáveis. Valores próximos de 1 ou -1 indicam forte relação.

```
1 % Matriz de correlação e heatmap
2 data = randn(100,4);
3 R = corrcoef(data);
4 figure;
5 heatmap(R, 'Title','Matriz de Correlação');
6 saveas(gcf,'figures/correlacao_matriz.png');
```

Código 7: Correlação entre variáveis numéricas.

Matriz de Correlação II



Testes Estatísticos

T

este usado para verificar se **duas médias** são significativamente diferentes. Assume normalidade dos dados.

```
1 % Teste t de duas amostras
2 data1 = randn(20,1)+1; data2 = randn(20,1);
3 [h,p,ci,stats] = ttest2(data1,data2);
4 fprintf('p-valor=%.3f, t=%.2f\\n',p,stats.tstat);
```

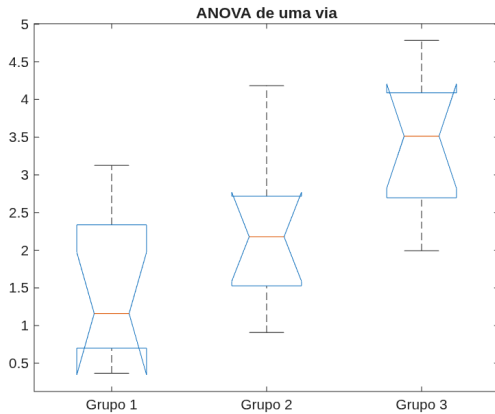
Código 8: Comparação de médias de dois grupos independentes.

Example

A ANOVA é uma generalização do teste t para 3 ou mais grupos. Se o resultado for significativo, é necessário realizar testes post-hoc.

```
1 % Exemplo de ANOVA de uma via (one-way ANOVA)
2 % Gerando três grupos de dados com médias diferentes
3 grupo1 = randn(10,1) + 1;
4 grupo2 = randn(10,1) + 2;
5 grupo3 = randn(10,1) + 3;
6 dados = [grupo1; grupo2; grupo3];
7 grupos = [repmat({'Grupo 1'},10,1); repmat({'Grupo 2'},10,1); repmat({'Grupo 3'},10,1)];
8
9 % Realizando ANOVA de uma via
10 [p,tbl,stats] = anova1(dados, grupos);
11 title('ANOVA de uma via');
12
13 % Salvando a figura
14 saveas(gcf,'figures/ml_anova.png');
```

Código 9: ANOVA de uma via.



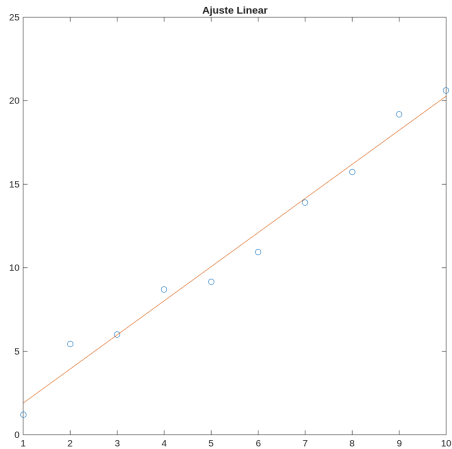
Modelagem e Previsão

Objetivo: ajustar uma reta que melhor representa os dados.

```
1 % Regressão linear simples
2 x = (1:10)'; y = 2*x + randn(10,1);
3 coeffs = polyfit(x,y,1); yfit = polyval(coeffs,x);
4 figure; plot(x,y,'o'); hold on; plot(x,yfit,'r-');
5 title('Regressão Linear'); legend('Dados','Ajuste'); grid on;
6 saveas(gcf,'figures/regressao_linear.png');
```

Código 10: Ajuste de regressão linear simples.

Regressão Linear Simples II

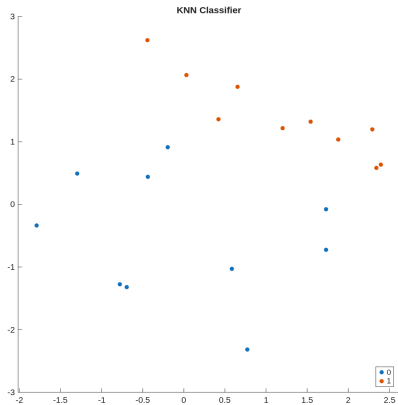


- O algoritmo *K-Nearest Neighbors* classifica novos pontos
- com base na proximidade de vizinhos já rotulados.

```
1 % Exemplo de KNN simples
2 X = [randn(10,2)+1; randn(10,2)-1]; Y = [ones(10,1); zeros(10,1)];
3 mdl = fitcknn(X,Y,'NumNeighbors',3);
4 figure; gscatter(X(:,1),X(:,2),Y); title('KNN Classifier');
5 saveas(gcf,'figures/ml_knn.png');
```

Código 11: Classificação simples usando KNN.

Classificação com KNN II



Boas Práticas e Fontes de Dados

- › Documentar cada passo da análise
- › Garantir reprodutibilidade (scripts)
- › Validar resultados com dados de teste
- › Usar versionamento (*Git*)

- › dados.gov.br (Brasil)
- › data.gov (EUA)
- › Kaggle Datasets
- › Our World in Data
- › UCI ML Repository

(Chapman, 2016)



Bibliografia



CHAPMAN, Stephen J. **Programação em MATLAB para Engenheiros**. 5. ed. São Paulo: Thomson, 2016.



MORAIS, V.; VIEIRA, C. **MATLAB Curso Completo**. [S. l.]: FCA, 2013. 644 p.



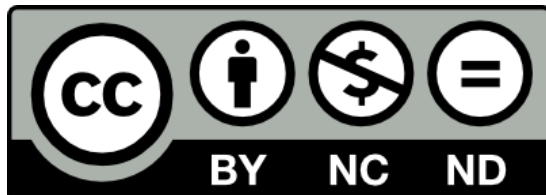
PALM, William J. **Introdução ao MATLAB para engenheiros**. 3. ed. Porto Alegre: AMGH, 2013. Tradução de Tales Argolo Jesus. xiv, 562 p.



MATHWORKS, INC. **MATLAB Help Center**. [*S. l.: s. n.*], 2025.

<https://www.mathworks.com/help/index.html>. Acesso em: 14 ago. 2025.

Estes slides estão protegidos por uma licença Creative Commons



Este modelo foi adaptado de Maxime Chupin.

Programação para Engenharia II

Análise Exploratória e Testes Estatísticos em MATLAB

