

Aplicação de Estatística Descritiva em Análise de Dados

Para analisar o conjunto de dados serão necessárias algumas técnicas de estatística descritiva. A estatística descritiva é a etapa inicial da análise estatística, que tem como objetivo descrever e sumarizar os dados.

Neste notebook, vamos explorar um conjunto de dados de preços de combustíveis em 3 cidades fictícias. Nele teremos as informações de preço de gasolina, etanol e diesel em 5 postos de gasolina de cada cidade num total de 15 postos.

```
# Importar as bibliotecas necessárias
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
import numpy as np

# Carregar o arquivo CSV
df = pd.read_csv('files/COMBUSTIVEIS.csv')

# Visualizar os primeiros registros do dataframe
print(df.head())

# Calcular estatísticas descritivas para cada cidade e tipo de combustível
statistics = df.groupby(['CIDADE', 'COMBUSTIVEL']).agg(
    media_preco=('PRECO', 'mean'),
    mediana_preco=('PRECO', 'median'),
    desvio_padrao=('PRECO', 'std'),
    variancia=('PRECO', 'var')
).reset_index()

print(statistics)
```

Análise dos resultados

1 - A cidade A apresenta uma variabilidade moderada nos preços, com o etanol apresentando a maior variação (desvio padrão de 0.499249).

2 - A cidade B mostra a maior variabilidade nos preços de diesel e gasolina, com desvios padrão de 1.044811 e 0.926418, respectivamente, sugerindo uma maior competitividade entre os postos.

3 - A cidade C tem uma baixa variabilidade nos preços de etanol e gasolina, mas uma variabilidade maior para o diesel.

4 - A cidade B parece ter a maior autonomia e competitividade entre os postos de combustível, dado os altos desvios padrão para diesel e gasolina.

5 - As cidades A e C mostram mais uniformidade nos preços, especialmente para etanol e gasolina na cidade C.

6 - O combustível com maior discrepância de preço é o diesel na cidade B, com um desvio padrão de 1.044811 e uma variância de 1.09163.

Conformidade/Variabilidade:

A cidade B mostra maior variabilidade nos preços, indicando uma política de autonomia e competitividade entre os postos. As cidades A e C apresentam mais uniformidade, especialmente para etanol e gasolina.

Geração de gráficos de boxplot para visualização dos resultados

```
# Configurar o estilo do seaborn
sns.set_theme(style="whitegrid")

# Criar visualizações para cada tipo de combustível comparando entre as cidades
def plot_fuel_prices_by_fuel(df):
    unique_fuels = df['COMBUSTIVEL'].unique()
    fig_paths = []

    for fuel in unique_fuels:
        fuel_data = df[df['COMBUSTIVEL'] == fuel]
        plt.figure(figsize=(12, 6))
        palettes = sns.color_palette("husl",
len(df['CIDADE'].unique()))
        sns.boxplot(x='CIDADE', y='PRECO', hue='CIDADE',
data=fuel_data, palette=palettes, legend=False)
        plt.title(f'Preços do {fuel.capitalize()} nas Cidades')
        plt.xlabel('Cidade')
        plt.ylabel('Preço (R$)')

        # Definir a escala do eixo y a cada 20 centavos, específico para cada combustível
        min_price = fuel_data['PRECO'].min()
        max_price = fuel_data['PRECO'].max()
        plt.yticks([round(x, 2) for x in np.arange(min_price,
max_price + 0.2, 0.2)])

        # Salvar a figura
        fig_path = f'precos_{fuel.lower()}_cidades.png'
        plt.savefig(fig_path)
        plt.close()
        fig_paths.append(fig_path)

    return fig_paths
```

```
# Plotar e salvar as visualizações
fig_paths = plot_fuel_prices_by_fuel(df)
for path in fig_paths:
    print(f'Figura salva em: {path}')
```

Devido a um problema com a instalação do jupyter notebook, não foi possível mostrar os gráficos de boxplot para mostrar aqui, portanto é necessário rodar o código para gerar os gráficos em formato de imagem e acessar os mesmos no diretório de trabalho.