

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import pearsonr, norm

# Carregar os dados
file_path = '/content/drive/MyDrive/Python/Frequencia
Igreja/Frequência_Ceia IBCA 2022-2024.xlsx'
data = pd.read_excel(file_path)

# Verificar os dados
print(data.head())
print("Colunas disponíveis:", data.columns)

# Remover colunas desnecessárias (Unnamed)
data = data[['Dias', 'Manhã', 'Noite']]

# Converter coluna 'Dias' para tipo de data
data['Dias'] = pd.to_datetime(data['Dias'], errors='coerce')

# Filtrar apenas datas que caíram em domingo
sundays_data = data[data['Dias'].dt.dayofweek == 6].copy()
sundays_data['Ano'] = sundays_data['Dias'].dt.year

# Salvar os dados de domingos em formato Parquet
sundays_data.to_parquet('frequencias_domingo.parquet', index=False)
print("Arquivo 'frequencias_domingo.parquet' salvo com sucesso.")

# Calcular estatísticas por culto e ano
def calculate_statistics(df, year, column_name):
    stats = {
        'Ano': year,
        'Culto': column_name,
        'Média': df[column_name].mean(),
        'Mediana': df[column_name].median(),
        'Desvio Padrão': df[column_name].std(),
        'Variância': df[column_name].var()
    }
    return stats

# Estatísticas por ano
years = sundays_data['Ano'].unique()
stats_manha = [calculate_statistics(sundays_data[sundays_data['Ano'] == year],
year, 'Manhã') for year in years]
stats_noite = [calculate_statistics(sundays_data[sundays_data['Ano'] == year],
year, 'Noite') for year in years]

```

```

# Criar DataFrame com estatísticas e salvar como Parquet
stats_df = pd.DataFrame(stats_manha + stats_noite)
stats_df.to_parquet('estatisticas_frequencia.parquet', index=False)
print("Arquivo 'estatisticas_frequencia.parquet' salvo com
sucesso.")
print(stats_df)

# Visualização: Boxplot por culto e ano
plt.figure(figsize=(12, 6))
sundays_data_melted = pd.melt(sundays_data, id_vars=['Ano'],
value_vars=['Manhã', 'Noite'], var_name='Culto',
value_name='Frequência')
sns.boxplot(x='Ano', y='Frequência', hue='Culto',
data=sundays_data_melted)
plt.title('Boxplot das Frequências Dominicais por Ano')
plt.xlabel('Ano')
plt.ylabel('Frequência')
plt.show()

# Gráfico de dispersão
plt.figure(figsize=(10, 6))
plt.scatter(sundays_data['Manhã'], sundays_data['Noite'],
alpha=0.7)
plt.title('Gráfico de Dispersão: Culto da Manhã vs. Culto da
Noite')
plt.xlabel('Frequência do Culto da Manhã')
plt.ylabel('Frequência do Culto da Noite')
plt.show()

# Correlação entre cultos
correlation, p_value = pearsonr(sundays_data['Manhã'],
sundays_data['Noite'])
print(f"Correlação entre frequência do culto da manhã e da noite:
{correlation:.2f}, Valor-p: {p_value:.4f}")

# Intervalo de confiança para a amostra
total_population = 1000
sample_size = 82
confidence_level = 0.95

z_score = norm.ppf(1 - (1 - confidence_level) / 2)
sample_proportion = sample_size / total_population
margin_of_error = z_score * np.sqrt((sample_proportion * (1 -
sample_proportion)) / total_population)

confidence_interval = (sample_proportion - margin_of_error,
sample_proportion + margin_of_error)

```

```
print(f"Nível de confiança da amostra ({confidence_level*100}%):  
{confidence_interval}")  
print(f"Tamanho da amostra como proporção da população:  
{sample_proportion:.2%}")  
print(f"Margem de erro: ±{margin_of_error:.2%}")
```