

Campina Grande – PB, 07/03/17

Curso: Engenharia de Computação

Disciplina: Probabilidade e estatística aplicada à computação

Professor: Paulo Ribeiro Lins Júnior

Aluno: Filipe Fernandes Ribeiro

Projeto 1

Resumo

Neste relatório estão as interpretações e os resultados dos dados retirados do google com relação a pesquisas sobre gripe no Brasil, com informações sobre o mesmo e os estados do Ceará, Minas Gerais, Paraná e Rio de Janeiro, tendo os dados nacionais no período de 22 de janeiro de 2006 à 09 de agosto de 2015. Os resultados foram feitos com a utilização das bibliotecas Pandas e Matplotlib no Python.

Introdução

Este projeto tem como principal objetivo a aplicação dos conteúdos relacionados a dados estatísticos, com o auxílio de ferramentas computacionais para realizar o cálculo e gerar os gráficos correspondentes.

Medidas

- **Posição**

| Local | Média | Moda | Mediana |
|----------------|--------|-------------|---------|
| Brasil | 199,41 | 149 193 196 | 192 |
| Ceará | 161,83 | 122 148 | 153 |
| Minas Gerais | 218,73 | 128 | 208 |
| Paraná | 196,70 | 181 | 183 |
| Rio de Janeiro | 209,10 | 260 | 204 |

Com esses dados é possível observar que o Brasil e o estado do Ceará se destacam em relação a moda, onde o primeiro é trimodal e o segundo bimodal. Apesar de não se ter uma variação tão grande quanto a média e a mediana, é importante frisar que o cálculo da mediana é bem mais exato do que o da média, pois este não é influenciado por outliers que podem causar uma alteração drástica no valor da média, dessa forma o resultado da mediana mostra exatamente o valor central dos dados, sendo bem mais preciso.

- **Dispersão**

| Local | Amplitude | Desvio médio absoluto | Variância | Desvio padrão |
|----------------|-----------|-----------------------|-----------|---------------|
| Brasil | 343 | 52,75 | 4326,28 | 65,77 |
| Ceará | 258 | 39,65 | 2388,36 | 48,87 |
| Minas Gerais | 377 | 63,87 | 6268,96 | 79,18 |
| Paraná | 494 | 61,36 | 6209,70 | 78,80 |
| Rio de Janeiro | 281 | 49,16 | 3522,18 | 59,35 |

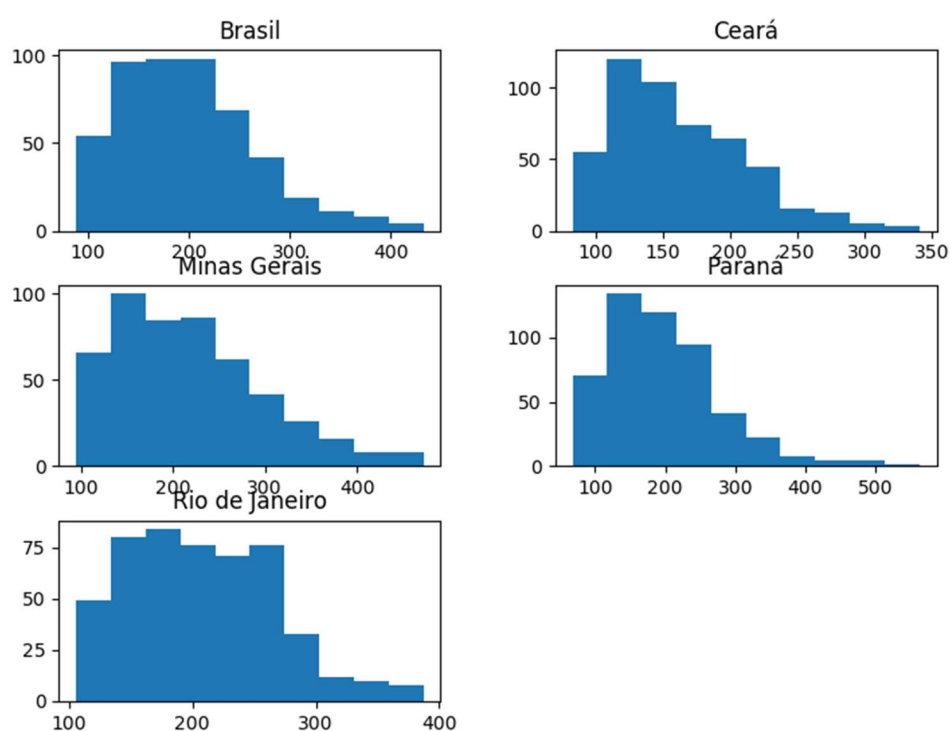
Com essas medidas é possível medir a variação dos dados. Sendo assim, vemos que Paraná e o Ceará foram os estados com os extremos de número de pesquisa nas datas avaliadas, logo, pode-se supor que o Paraná foi o estado com maior número de casos de dengue e o Ceará o menor.

- Quartis

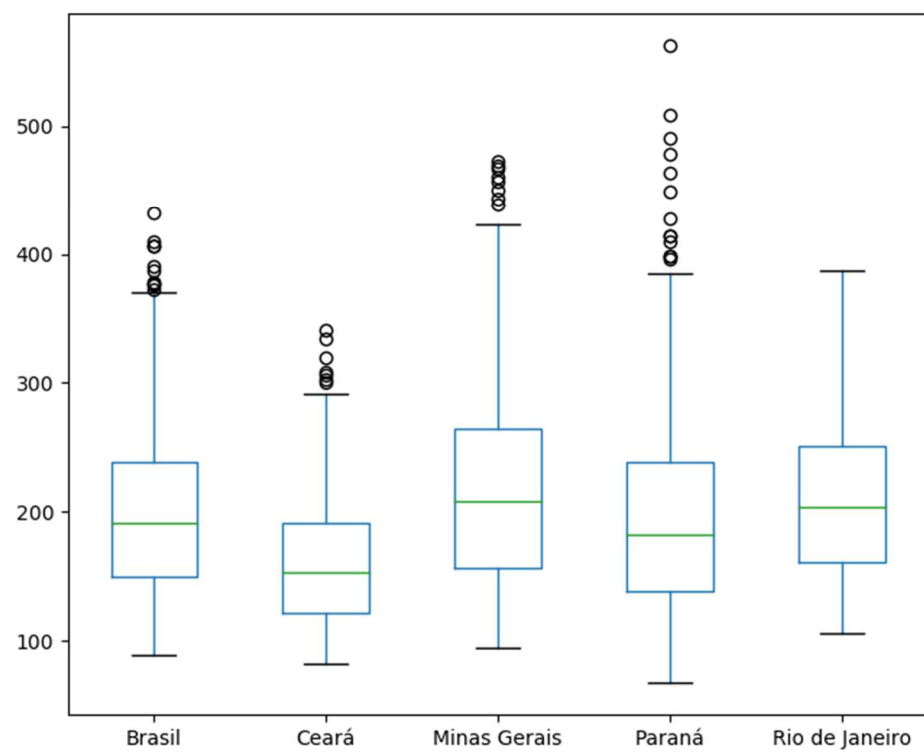
| Local | 1º Quartil | 2º Quartil | 3º Quartil |
|----------------|------------|------------|------------|
| Brasil | 150 | 192 | 239 |
| Ceará | 122 | 153 | 192 |
| Minas Gerais | 157 | 208 | 264 |
| Paraná | 138,50 | 183 | 239 |
| Rio de Janeiro | 161 | 204 | 251,50 |

Com esses valores é possível ter uma média das pesquisas realizadas no período proposto, o segundo quartil é referente ao valor da mediana a média de acessos pode ser analisada pegando o 1º e o 3º Quartil, onde dão 50% dos acessos no tempo.

Gráficos



Histograma



Boxplots

Conclusão

Com este projeto foi possível ter uma maior ideia de como aplicar os conhecimentos a cerca de dados estatísticos na computação e com isso poder-se ter um maior conhecimento sobre os conjuntos de dados aos quais estamos lidando, tendo a aplicação dessas estatísticas extremamente vastas, podendo facilitar e muito na tomada de decisões para realização de uma atividade ou para controle do que está acontecendo.

Código

```
'''Projeto de Probabilidade e estatística aplicada a programação.
Aluno: Filipe Fernandes Ribeiro
Curso: Engenharia de Computação'''

'''Imports'''
from pandas import *
from matplotlib.pyplot import *
'''Funções'''
def linhas_f(linhas):
    for i in range(172):
        linhas.append(i)
'''Variáveis'''
linhas_extras = []
linhas_f(linhas_extras)
c_label = ['Data', 'Brasil', 'Ceará', 'Minas Gerais', 'Paraná', 'Rio
de Janeiro']
dados = read_csv('data.csv', header=None, skiprows=linhas_extras,
usecols=[0, 1, 2, 4, 5, 6], names=c_label)
'''Código'''
while True:
    print('Selecione uma opção: \n'
          'A - Brasil\n'
          'B- Ceará\n'
          'C- Minas Gerais\n'
          'D- Paraná\n'
          'E- Rio de Janeiro\n'
          'F- Histograma\n'
          'G- Boxplots\n'
          'H- Finalizar Programa')
    opcao = input('Escolha: ').upper()
    if opcao == 'A':
        print('BRASIL\n'
              'Média: ' + str(dados['Brasil'].mean()) + '\n'
              'Moda: \n' + str(dados['Brasil'].mode()) + '\n'
              'Mediana: ' + str(dados['Brasil'].median()) + '\n'
              'Amplitude: ' + str(dados['Brasil'].max() -
dados['Brasil'].min()) + '\n'
              'Desvio Absoluto: ' + str(dados['Brasil'].mad()) + '\n'
              'Variância: ', dados['Brasil'].var(), '\n'
              'Desvio Padrão: ', dados['Brasil'].std(), '\n'
              'Quartis: \n'
              'Q1 = ', dados['Brasil'].quantile(q=0.25), '\n'
              'Q2 = ', dados['Brasil'].quantile(), '\n'
              'Q3 = ', dados['Brasil'].quantile(q=0.75), '\n')
    elif opcao == 'B':
        print('CEARÁ\n'
              'Média: ' + str(dados['Ceará'].mean()) + '\n'
              'Moda: \n' + str(dados['Ceará'].mode()) + '\n'
              'Mediana: ' + str(dados['Ceará'].median()) + '\n'
              'Amplitude: ' + str(dados['Ceará'].max() -
dados['Ceará'].min()) + '\n'
              'Desvio Absoluto: ' + str(dados['Ceará'].mad()) + '\n'
              'Variância: ', dados['Ceará'].var(), '\n'
              'Desvio Padrão: ', dados['Ceará'].std(), '\n'
              'Quartis: \n'
              'Q1 = ', dados['Ceará'].quantile(q=0.25), '\n'
              'Q2 = ', dados['Ceará'].quantile(), '\n'
              'Q3 = ', dados['Ceará'].quantile(q=0.75), '\n')
    elif opcao == 'C':
```

```

print('MINAS GERAIS\n'
      ' Média: ' + str(dados['Minas Gerais'].mean()) + '\n'
      ' Moda: \n' + str(dados['Minas Gerais'].mode()) + '\n'
      ' Mediana: ' + str(dados['Minas Gerais'].median()) +
'\n'
      ' Amplitude: ' + str(dados['Minas Gerais'].max() -
dados['Minas Gerais'].min()) + '\n'
      ' Desvio Absoluto: ' + str(dados['Minas Gerais'].mad())
+ '\n'
      ' Variância: ', dados['Minas Gerais'].var(), '\n'
      ' Desvio Padrão: ', dados['Minas Gerais'].std(), '\n'
      ' Quartis: \n'
      ' Q1 = ', dados['Minas Gerais'].quantile(q=0.25), '\n'
      ' Q2 = ', dados['Minas Gerais'].quantile(), '\n'
      ' Q3 = ', dados['Minas Gerais'].quantile(q=0.75), '\n')
elif opcao == 'D':
    print('PARANÁ\n'
          ' Média: ' + str(dados['Paraná'].mean()) + '\n'
          ' Moda: \n' + str(dados['Paraná'].mode()) + '\n'
          ' Mediana: ' + str(dados['Paraná'].median()) + '\n'
          ' Amplitude: ' + str(dados['Paraná'].max() -
dados['Paraná'].min()) + '\n'
          ' Desvio Absoluto: ' + str(dados['Paraná'].mad()) + '\n'
          ' Variância: ', dados['Paraná'].var(), '\n'
          ' Desvio Padrão: ', dados['Paraná'].std(), '\n'
          ' Quartis: \n'
          ' Q1 = ', dados['Paraná'].quantile(q=0.25), '\n'
          ' Q2 = ', dados['Paraná'].quantile(), '\n'
          ' Q3 = ', dados['Paraná'].quantile(q=0.75), '\n')
elif opcao == 'E':
    print('RIO DE JANEIRO\n'
          ' Média: ' + str(dados['Rio de Janeiro'].mean()) + '\n'
          ' Moda: \n' + str(dados['Rio de Janeiro'].mode()) + '\n'
          ' Mediana: ' + str(dados['Rio de Janeiro'].median()) +
'\n'
          ' Amplitude: ' + str(dados['Rio de Janeiro'].max() -
dados['Rio de Janeiro'].min()) + '\n'
          ' Desvio Absoluto: ' + str(dados['Rio de
Janeiro'].mad()) + '\n'
          ' Variância: ', dados['Rio de Janeiro'].var(), '\n'
          ' Desvio Padrão: ', dados['Rio de Janeiro'].std(), '\n'
          ' Quartis: \n'
          ' Q1 = ', dados['Rio de Janeiro'].quantile(q=0.25),
'\n'
          ' Q2 = ', dados['Rio de Janeiro'].quantile(), '\n'
          ' Q3 = ', dados['Rio de Janeiro'].quantile(q=0.75),
'\n')
elif opcao == 'F':
    dados.hist(grid=False)
    show()
elif opcao == 'G':
    dados.boxplot(grid = False)
    show()
elif opcao == 'H':
    break

```