

Relatório 2 da disciplina Probabilidade e Estatística Aplicada à Computação – Demanda e Atendimento

Arthur V. F. Fialho, Eric Guimarães Barbosa, Jardel Brandon de Araujo Regis

Instituto Federal de Educação, Ciência e Tecnologia da Paraíba (IFPB)
Campus de Campina Grande – PB – Brazil

arthurvff3.0@gmail.com, ericguimaraes@msn.com, jardelbrandon@hotmail.com

Resumo. *Este relatório mostra os resultados obtidos na análise de dados do Data Mart Demanda e Atendimento do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Os dados foram minerados com o auxílio linguagem Python e apresentados por meio de gráficos plotados com essa mesma linguagem.*

1. Introdução

O objetivo deste trabalho é plotagem de gráficos do tipo Boxplot e Histogramas com o auxílio da biblioteca Matplotlib. Os gráficos resumem e acrescentam algumas outras informações sobre o Data Mart Demanda e Atendimento, além dos que já foram apresentados no trabalho anterior.

O Data Mart Demanda e Atendimento evidencia o quantitativo bolsas e auxílios demandados e atendidos por ano. Essas informações podem ser filtradas por: Instituição, Região, Área do Conhecimento e Chamada. Também estão separados por Bolsas no País, Bolsas no Exterior, Projetos de Pesquisa, Eventos e Editoração. As informações estão disponíveis desde do ano de 2006, os painéis são atualizados trimestralmente.

2. Apoio a editoração

Considerando que, o gráfico ilustra a quantidade de auxílios à editoração no âmbito nacional, englobando todas as instituições cadastradas entre 2006 e 2015.

Podemos inferir, com base no gráfico abaixo (gráfico 1) que:

- Apesar de apresentar um Outlier acima de R\$ 40.000,00 o valor médio se encontra bem abaixo de R\$ 5.000,00.
- Mesmo com o valor médio de bolsas abaixo de R\$ 5.000,00 percebemos ainda uma grande concentração de bolsas distribuídas até aproximadamente R\$ 18.000,00 representadas pela grande quantidade de discrepâncias acima do limite superior.

- A gráfico apresenta grandes discrepâncias com valores distribuídos em uma larga faixa de valores até aproximadamente R\$ 50.000,00

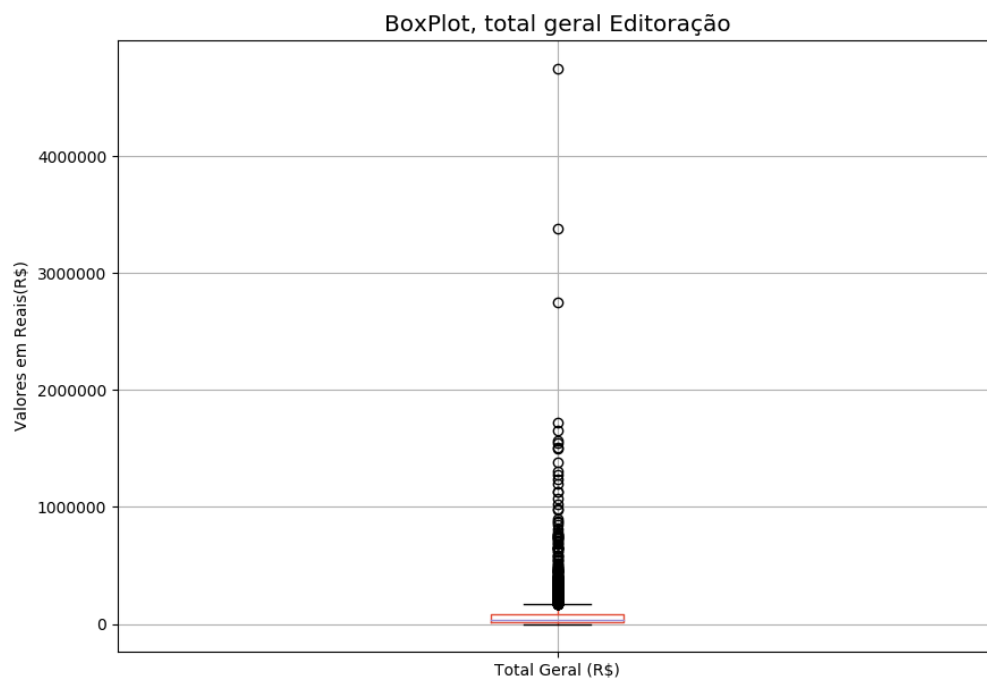


Gráfico 1. Apoio à editoração, a nível nacional, entre 2006 e 2015

Podemos inferir, com base no gráfico abaixo (gráfico 2) que:

- Assim com no gráfico 1, o gráfico 2 traz uma grande discrepância de valores que partem próximo de zero até R\$ 200.000,00
- Percebe-se um Outliers à R\$ 200.000,00, porém sendo uma ocorrência bem isolada.
- Percebe-se também uma grande concentração de discrepâncias ao longo do limite superior até aproximadamente o valor de R\$ 100.00,00.

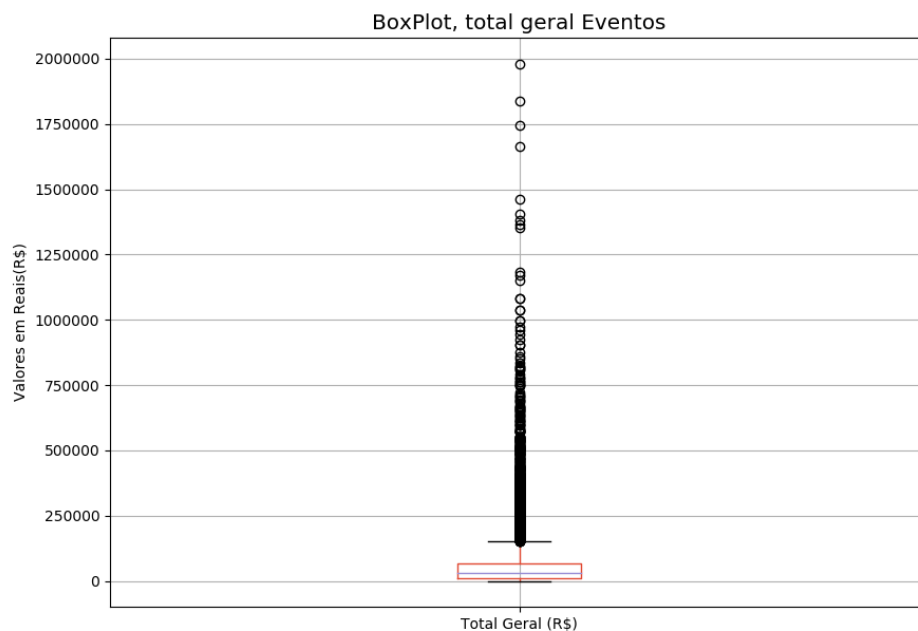


Gráfico 2. Total Geral em R\$ das Propostas de Bolsas para Eventos

3. Histograma de Apoio a Editoração

Podemos inferir, com base no gráfico abaixo (gráfico 3) que:

- A quantidade de auxílios diminui de forma visivelmente exponencial, de modo que se faz necessário dividir os valores em dois gráficos com escalas reajustadas.
- Percebe que a maior quantidade de auxílios ocorre em bolsas de até 1.000,00, apresentando uma frequência de aproximadamente 2.000 ocorrências.
- A maior quantidade de auxílios, neste caso, é inversamente proporcional aos valores das bolsas.

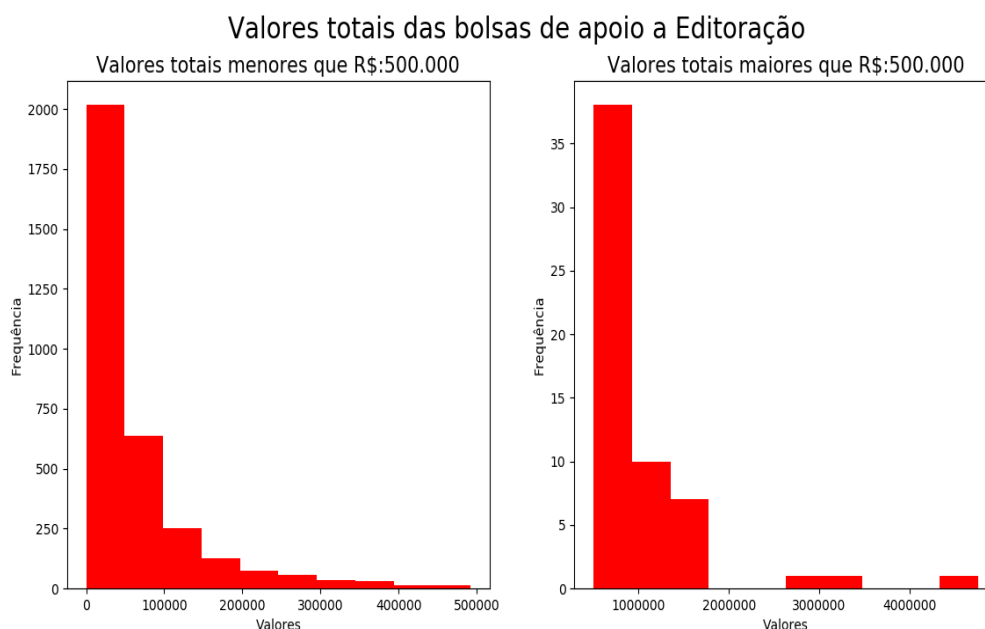


Gráfico 3. Bolsas por ano ofertadas no exterior

4. Apoio a Eventos

Podemos inferir, com base no gráfico abaixo (gráfico 4) que:

- Assim como no gráfico 3, ocorre a diminuição drástica de apoios quando comparadas ao valor de cada bolsa, sendo necessário mais uma vez um reajuste de escala para que possamos visualizar mais detalhadamente os dados.
- Mesmo dividindo os gráficos, ambos são semelhantes, o que nos leva a concluir que o comportamento é praticamente constante, decaindo da mesma forma que na primeira parte do gráfico.
- Analisando este tipo de gráfico, o valor das bolsas parece ser bem impactante na frequência com que são deferidas, isso pode inferir que apoios de menor valor apresentam uma probabilidade de deferimento bem maior que bolsas de valor mais elevado. Isso complementa as informações dos gráficos de Boxplot, que por apresentar Outliers, podem transmitir a ideia de que são muito significativos diante do montante gasto com as demais faixas de valores, embora saibamos que a presença de alguns outliers são de grande importância para uma completa análise.

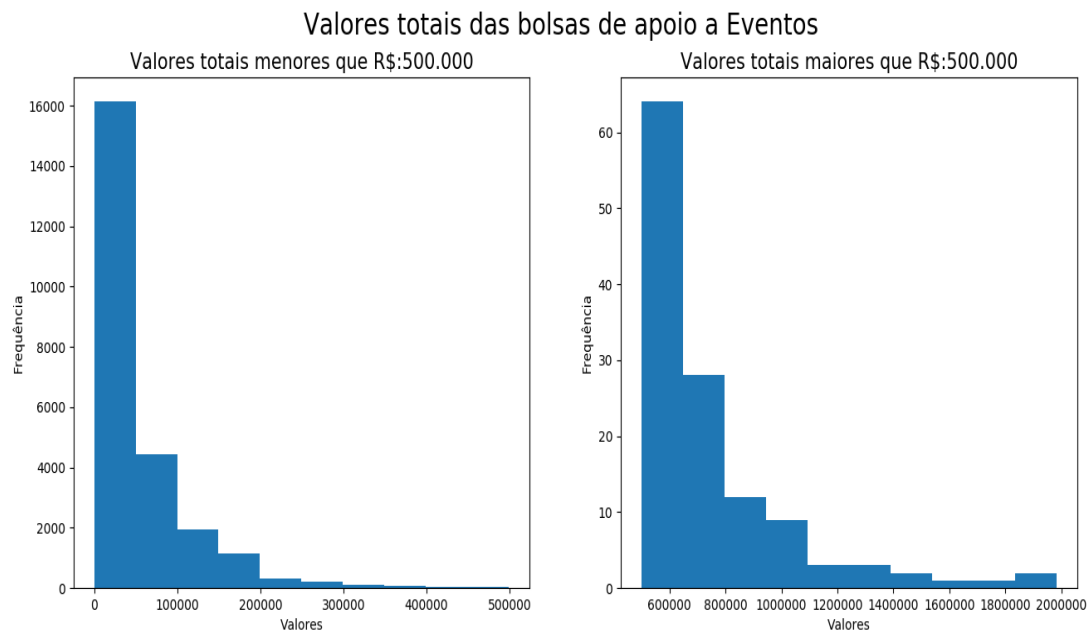


Gráfico 4. Histograma dos valores de Apoio a Eventos

5. Cálculos de Quartis, Assimetria e Curtose

Abaixo seguem as imagens 5, 6 e 7 com os resultados dos cálculos dos Quartis, Assimetria e Curtose obtidas.

```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import matplotlib
4
5 apoio_editoracao = pd.read_csv('Editoracao.csv', sep=';')
6 coluna_total_geral = apoio_editoracao["Total Geral (R$)"]
7
8 print("*****")
9 print("Valores dos Quartis: \n")
10 print("Valor do primeiro quartil:")
11 print(coluna_total_geral.quantile(0.25), "\n")
12 print("Valor do segundo quartil (mediana):")
13 print(coluna_total_geral.quantile(), "\n")
14 print("Valor do terceiro quartil:")
15 print(coluna_total_geral.quantile(0.75))
16 print("*****")
17
18 print("*****")
19 print("Valores de assimetria: \n")
20 print(coluna_total_geral.skew())
21 print("*****")
22
23 print("*****")
24 print("Valores de curtose: \n")
25 print(coluna_total_geral.kurt())
26 print("*****")
27

```

```
Programa - [C:\Users\jarde\OneDrive\Documentos\GitHub\Probabilidade_e_Estatistica_Aplicada_
File Edit View Navigate Code Refactor Run Tools VCS Window Help
VariaveisDiscrepantes_Eventos
Programa VariaveisDiscrepantes_Eventos.py
Run VariaveisDiscrepantes_Eventos
C:\Users\jarde\Anaconda3\python.exe C:/Users/jarde/OneDrive/Doc

Cálculos referentes ao Total Geral do arquivo Eventos:

*****
Valores dos Quartis:

Valor do primeiro quartil:
10635.0

Valor do segundo quartil (mediana):
29940.0

Valor do terceiro quartil:
67674.0
*****
Valores de assimetria:

6.12260192407
*****
Valores de curtose:

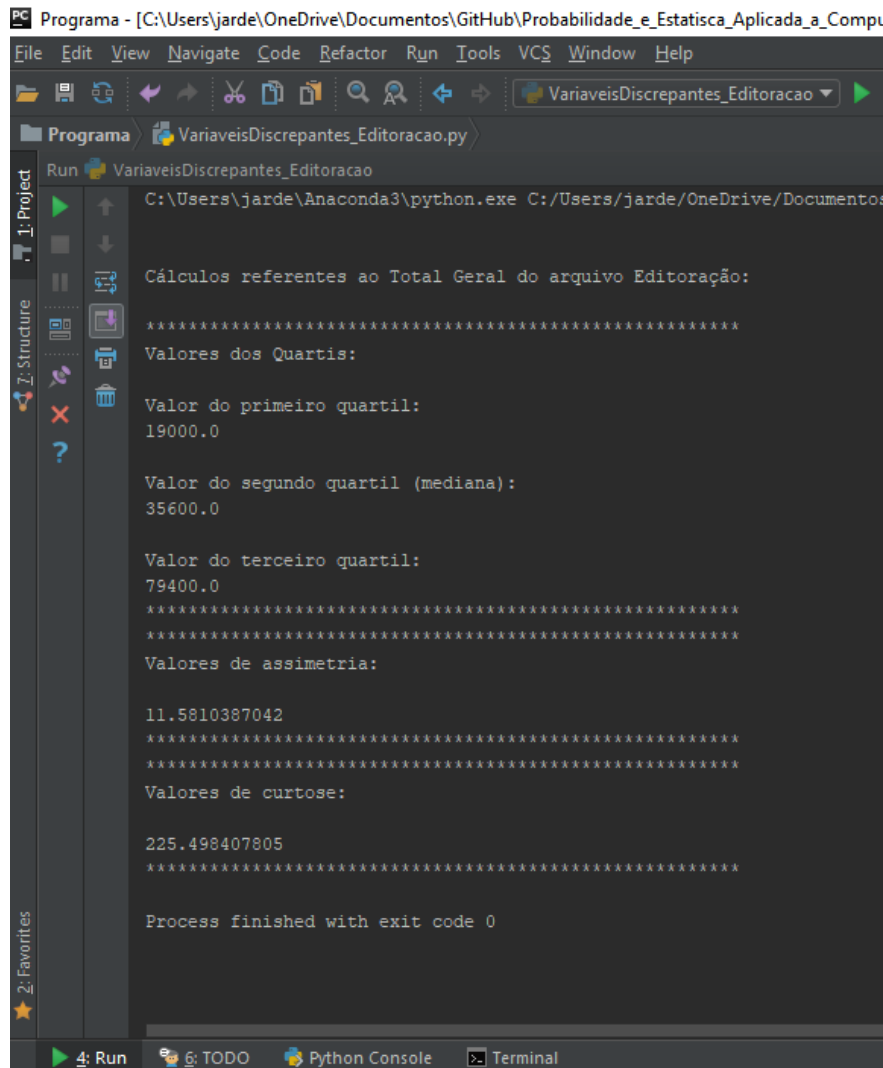
74.9827845774
*****

Process finished with exit code 0
```

Unregistered VCS root detected: The directory C:\Users\jarde\OneDrive\Documentos\GitHub\F

Imagem 6. Cálculos de ‘Eventos’

6.



```
Programa - [C:\Users\jarde\OneDrive\Documentos\GitHub\Probabilidade_e_Estatistica_Aplicada_a_Compu
File Edit View Navigate Code Refactor Run Tools VCS Window Help
VariaveisDiscrepantes_Editoracao
Run VariaveisDiscrepantes_Editoracao
C:\Users\jarde\Anaconda3\python.exe C:/Users/jarde/OneDrive/Documentos/

Cálculos referentes ao Total Geral do arquivo Editoração:

*****
Valores dos Quartis:

Valor do primeiro quartil:
19000.0

Valor do segundo quartil (mediana):
35600.0

Valor do terceiro quartil:
79400.0
*****
Valores de assimetria:

11.5810387042
*****
Valores de curtose:

225.498407805
*****

Process finished with exit code 0
```

Imagem 7. Cálculos de ‘Editoração’

6. Conclusão

Histogramas abordam a distribuição de frequência enquanto os Boxplots são uma alternativa aos Histogramas e modelos de distribuição em folha, ambos permitem uma análise de dados por uma perspectiva que aborda informações muito importantes, embora o Boxplot tenha o diferencial de destacar Outliers, o que pode ser bastante importante, por exemplo, em análises para tomadas de decisão. A mesclagem desse tipo de gráfico com outros modelos formam um importante conjunto de ferramentas para uma análise abrangente e minuciosa dos mais diversos tipos de dados.