

Relatório 3 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 Dispersão, que foi obtido com a realização dos cálculos de correlação e 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. Correlação entre apoio a Editoração e apoio a Eventos

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

- O gráfico ilustra a correlação dos valores em reais que foram destinados para pagamentos de auxílios nos programas de editoração e eventos, englobando todos os estados do Brasil entre 2006 e 2015.
- O estado de São Paulo no gráfico representado pelo ponto amarelo apresentou um *Outlier* relevante, o que indica que o apoio a este estado foi bem mais acentuado que em outros estados do País.

- Como o cálculo de correlação desta medida se aproximou de 1, significa que os dados dos totais aplicados a editoração e eventos tem uma relação forte, ou seja, quanto maior o investimento em editoração, maior será o investimento em eventos, essa forte correlação pode ser observado com pontos quase lineares no gráfico de dispersão.

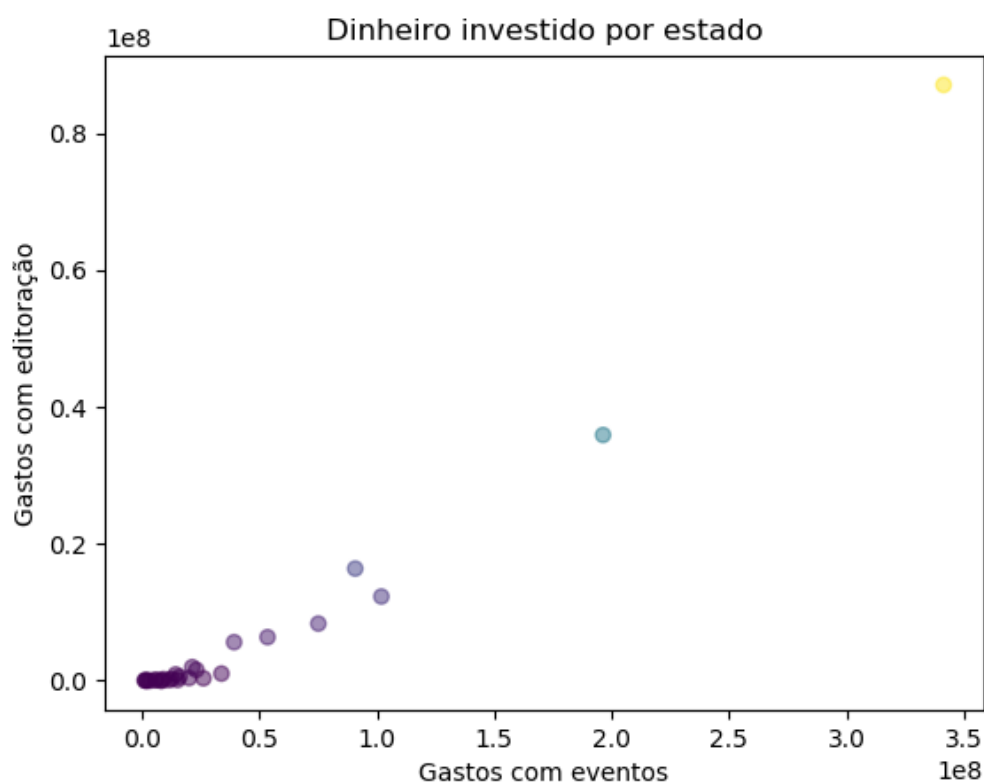


Gráfico 1. Correlação entre os totais gastos com apoios a editoração e eventos por Unidade Federativa (UF), a nível nacional, entre 2006 e 2015

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

- A correlação também é forte quando comparada com a correlação por estados, porém pode-se observar a presença de um Outlier no ano de 2011;
- A correlação do ano 2016 é 0 pois não existem dados sobre esse ano no DataFrame de Editoração
- A correlação do ano 2015 é 0 pois nenhuma das unidades federativas foi informada nesse ano no DataFrame de Editoração

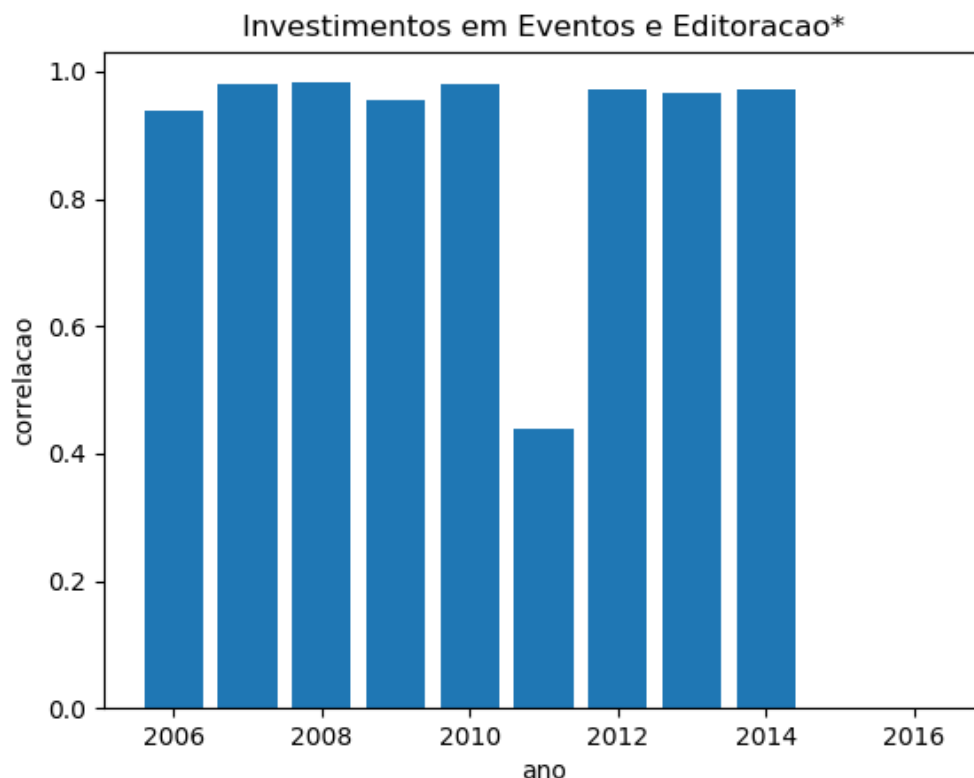


Gráfico 2. Propostas de bolsas por Instituições na PB em 2015

Obs: A correlação dos anos 2006 e 2011 não são confiáveis pois há um grande número de unidades federativas não informadas nesses anos no DataFrame de Editoração

3. Cálculos de correlações por Unidades Federativas (UF)

Abaixo segue imagem (Imagem 1), com os cálculos obtidos da correlação entre os valores investidos por estados aplicados ao apoio a editoração e apoio a pesquisa, o cálculo da correlação abaixo da imagem se aproxima de 1, logo essas medidas tem uma forte correlação.

```

rama$ python3 Correlacao_Editoracao_Eventos.py
Editoracao (R$)  Eventos (R$)  UF
0      0.0      1704418.0  AC
1      0.0      8469664.0  AL
2     39500.0     1176081.0  AP
3    909000.0     14375640.0  AM
4   5614831.0    39243016.0  BA
5     92362.0     15177501.0  CE
6   8333064.0     75102896.0  DF
7    258852.0     12797431.0  ES
8    562845.0     15995835.0  GO
9      0.0      8066512.0  MA
10     75774.0     11512506.0  MT
11    119610.0      5843895.0  MS
12   16377530.0    90810534.0  MG
13    377000.0     20061088.0  PA
14   1971632.0     21422807.0  PB
15   1570631.0     23185398.0  PR
16   1021739.0     33874862.0  PE
17     57042.0      5890086.0  PI
18    338384.0     26179660.0  RN
19   12292327.0    101949493.0  RS
20   35907999.0    196336099.0  RJ
21      0.0      2410758.0  RO
22    20500.0     1582210.0  RR
23   6352846.0     53491162.0  SC
24   87090370.0    341350157.0  SP
25    203600.0      9222772.0  SE
26      0.0      3388780.0  TO

Editoracao (R$)  Eventos (R$)
Editoracao (R$)      1.000000      0.978899
Eventos (R$)        0.978899      1.000000

```

Imagem 1. Totais gerais em reais de cada estado nacional, seguido do cálculo de suas correlações

4. Conclusão

Medidas de correlação abordam a relação entre duas variáveis enquanto os gráficos de dispersão podem representar de maneira satisfatória essa medida obtida do cálculo da correlação, ambos permitem uma análise de dados por uma perspectiva que aborda informações muito importantes. 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.