

# Interpolação e Correlação Utilizando Dados Abertos: Chuvas, Dengue e Acidentes de Trânsito

Airton Martins Neris<sup>1</sup>, Carlos Ivan S.A<sup>1</sup>, José Carlos M. S. Junior<sup>1</sup>, Manoel Antônio de F. Neto<sup>1</sup>

<sup>1</sup>Departamento de Estatística e Informática – Universidade Federal Rural de Pernambuco (UFRPE)

R. Manuel de Medeiros, s/n – 52.171-900 – Recife – PE – Brazil

{airton.neris, carlosamaral.gps, josecmontejr, mbetmann}@gmail.com

**Abstract.** *This article describes a academic work that aims to collect, analyze and compare data: meteorological, health (dengue) and traffic accidents, observing their frequency to conclude if there are relation between they. The developer group used some correlation and interpolation techniques with the information that was extract to execute math calculations and measure the link between this three themes. The intention was discover if there is relation between this variables and if was possible predict theses facts.*

**Resumo.** *Este artigo descreve um trabalho acadêmico que tem o objetivo de coletar, analisar e comparar dados: meteorológicos, de saúde (dengue) e acidentes de trânsito, observando sua periodicidade para concluir se há relação entre estes. O grupo desenvolvedor do projeto utilizou algumas técnicas de correlação e interpolação com as informações extraídas para executar cálculos matemáticos e medir a ligação entre esses três temas. O intuito foi descobrir se há relação entre essas variáveis e se seria possível prever tais fatos.*

## 1. Introdução

A dengue, uma doença viral transmitida pela picada do mosquito *Aedes Aegypti*<sup>1</sup>, tem se tornado uma preocupação para o governo do Brasil todo ano, principalmente por causa da variante “hemorrágica”, mesmo com as medidas de prevenção e controle, a epidemia aparece, principalmente em áreas carentes de saneamento básico, no mínimo uma vez por ano. O modo como o vetor transmissor da patologia se reproduz é um fator importante no crescimento dessa estatística, visto que, o inseto coloca seus ovos em focos de água limpa, sendo esse o motivo para o aumento de incidência nas épocas de chuva.

As chuvas também são grandes responsáveis por grande parte dos acidentes de trânsito anualmente. Segundo o Observatório Nacional de Segurança Viária, aproximadamente trinta e sete mil pessoas morrem por ano e seiscentas mil ficam com

---

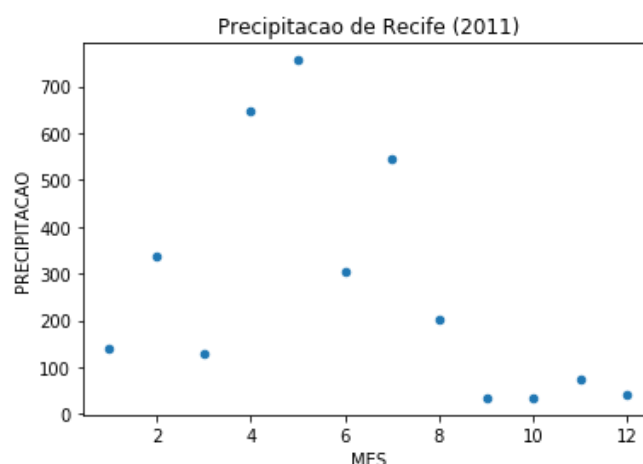
<sup>1</sup> "Período de chuvas traz risco da dengue | Notícias do ... - ECA/USP." <http://www2.eca.usp.br/njsaoremo/?p=2626>. Acessado: 21 Jan. 2019.

sequelas devido a acidentes de trânsito<sup>2</sup>. Fatores como dificuldade de visibilidade e diminuição de aderência dos pneus por causa das estradas molhadas são causas dos acidentes registrados em épocas de precipitação.

Levando em consideração essas duas problemáticas, a equipe resolveu fazer um estudo em cima destas e desenvolver uma solução que pudesse prever a precipitação de chuvas e conectá-la, estatisticamente, aos números já registrados de incidência de casos de dengue e acidentes de trânsito, podendo assim, entender melhor qual a relação entre esses três fatos. Para isso, foram executados cálculos matemáticos de correlação de Spearman (Rô de Spearman) e interpolação em cima de bases de dados abertos a respeito dos municípios do estado de Pernambuco - Brasil extraídas de fontes seguras do governo.

## 2. Metodologia

Para medir o nível de correlação entre as variáveis manipuladas no estudo, foi utilizado o coeficiente de correlação de Spearman (Rô de Spearman) através do Pandas, uma biblioteca da linguagem Python, que serve para medir a intensidade entre dois dados ordinais. Esse método foi selecionado entre outros disponíveis, por funcionar bem com dados assimétricos, ou seja, se as amostras colhidas resultariam em dados não lineares, a fórmula entregaria uma resposta mais perto possível da real e esperada. Para os dados referentes às precipitações, por exemplo, foram gerados gráficos de dispersão com o módulo da mesma linguagem chamado Matplotlib e observado que a distribuição é anormal não linear, isso explica a escolha do mecanismo.



**Figura 1. Exemplo de gráfico de dispersão não linear**

O coeficiente de correlação de Spearman é uma medida de dependência estatística entre duas variáveis. Uma variável X (variável independente) e uma variável Y (variável dependente). O coeficiente tem como objetivo avaliar com que intensidade

<sup>2</sup> "Brasil reduz mortes no trânsito, mas está longe da meta para 2020 ...." 18 Sep. 2018, <http://agenciabrasil.ebc.com.br/geral/noticia/2018-09/brasil-reduz-mortes-no-transito-mas-esta-longe-da-meta-para-2020>. Acessado: 21 Jan. 2019.

uma relação entre duas variáveis pode ser descrita através do uso de uma função monótona.

A correlação de Spearman avalia relações monótonas lineares e não lineares, como já citado. Os valores do coeficiente obtidos variam entre -1 e +1. Quanto mais próximo do valor 0, temos o indício de que mais fraca será a relação, quanto mais distante do 0, mais forte. Quando o coeficiente é positivo, temos uma tendência crescente entre X e Y, ou seja, quando X cresce, Y tende a acompanhar o crescimento de X. Já um coeficiente negativo corresponde a uma tendência decrescente entre X e Y, quando X cresce, Y tende a diminuir.

$$\rho = 1 - \frac{6 \sum_{i=1}^n d^2}{n(n^2 - 1)}$$

**Figura 2. Fórmula do coeficiente de Spearman**

### **3. Dados Meteorológicos**

Os dados meteorológicos utilizados neste artigo são provenientes de 7 estações meteorológicas no estado de Pernambuco (Recife, Ouricuri, Cabrobó, Arcoverde, Triunfo, Surubim e Petrolina), disponibilizados pelo Instituto Nacional de Meteorologia - INMET. Como o estado de Pernambuco possui 185 cidades, tem-se uma motivação para o encontro de técnicas para preenchimento dos dados que não possuímos, melhorando a cobertura na intersecção entre os dados de precipitação e as outras bases de doenças e acidentes. Para isso foram efetuados experimentos de interpolação e validação dos dados de precipitações de chuva entre essas 7 cidades, utilizando validação cruzada para avaliação da técnica utilizada.

#### **3.1 Experimento de interpolação**

Para interpolação dos dados foi utilizado o algoritmo Krigagem Ordinária, técnica de interpolação geoespacial bastante utilizada no contexto de dados ambientais. Para montagem dos experimentos de interpolação foi utilizado como referência a metodologia de Clodoalves (2018), onde se faz experimentos comparativos entre técnicas para interpolação de dados de evapotranspiração, utilizando a validação cruzada e uma das suas métricas de avaliação é o erro quadrático médio. Para execução dos experimentos foram utilizados os dados de 2015 até 2017, essa base possui registros do acúmulo da precipitação diária. A validação cruzada é uma técnica que ajuda a avaliar a capacidade de generalização de uma determinada técnica, essa divide os dados em k fatias, dado que  $1 < k \leq N$ , sendo N quantidade de dados. Definindo o valor de k, que no nosso experimento  $k = 7$ , referente às 7 estações, a validação cruzada se encarrega de utilizar um subconjunto de fatias para treinar a técnica e outro para prever, fazendo uma combinação exclusiva para cada K fatias. Tendo os resultados de

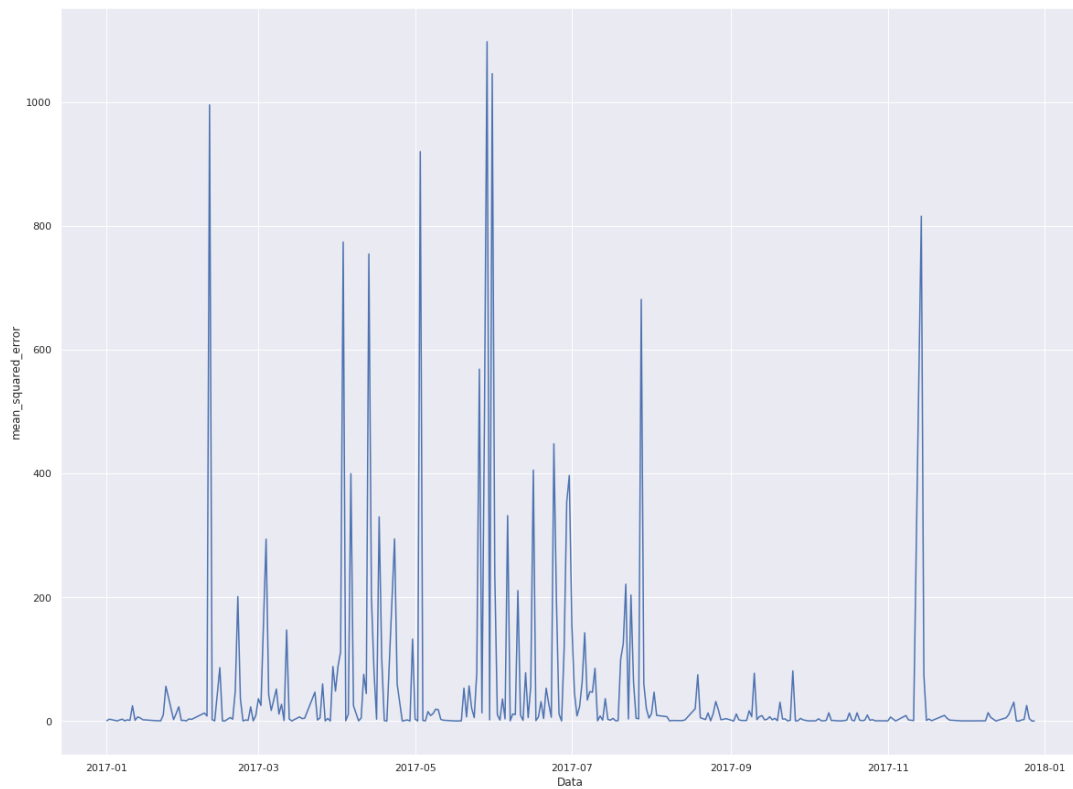
cada predição, utiliza-se uma métrica de avaliação para problemas de regressão, no nosso caso foi utilizado o erro quadrático médio. Como mostra a Figura 3, o erro quadrático médio tem o N como número de pontos interpolados,  $f_i$  o valor interpolado e  $y_i$  o valor real.

$$MSE = \frac{1}{N} \sum_{i=1}^N (f_i - y_i)^2$$

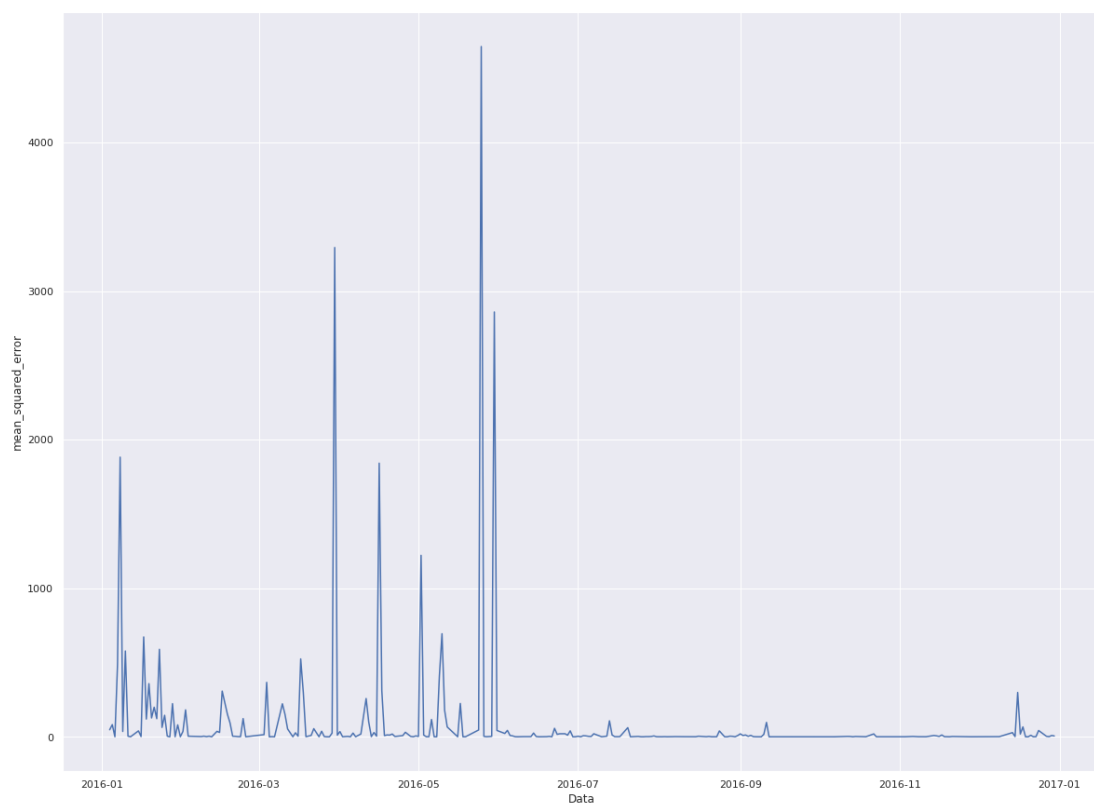
**Figura 3. Fórmula do erro quadrático médio**

### **3.3 Resultado da interpolação**

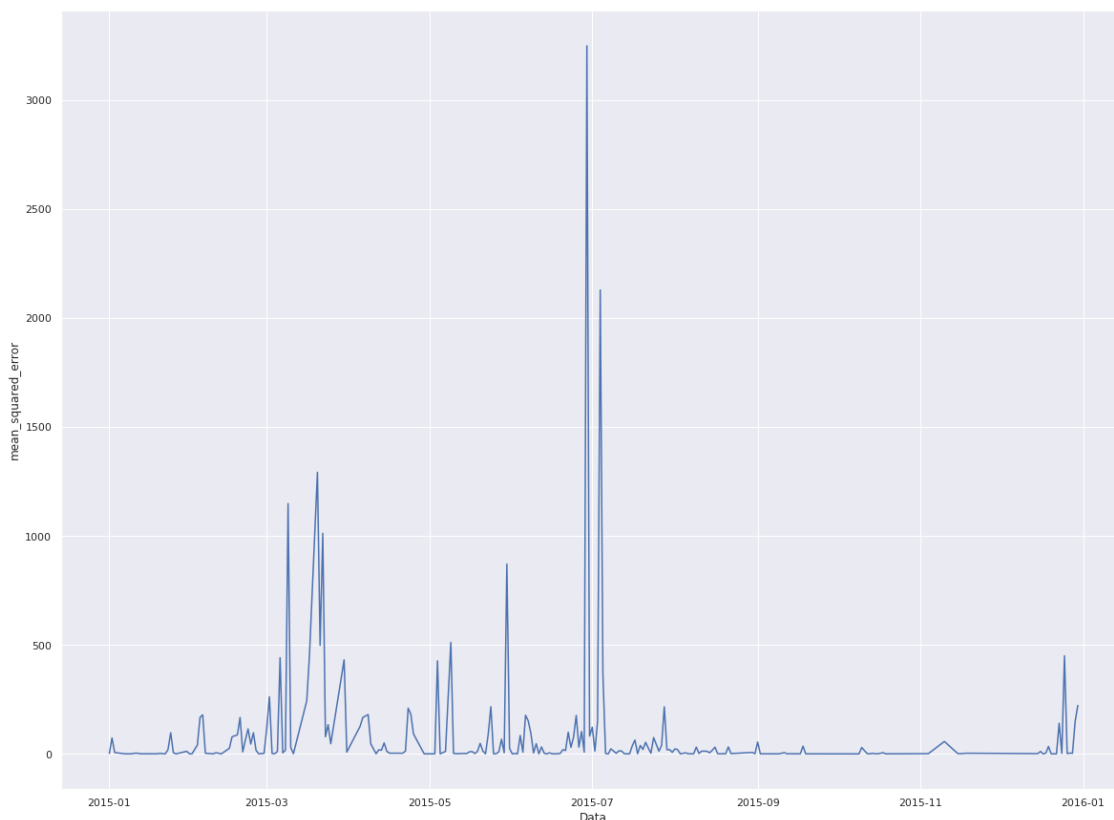
Como podemos observar nas figuras 4, 5 e 6, o erro quadrático médio foi bastante elevado em determinados pontos nos experimentos dos 3 anos. No experimento do ano de 2017, houve uma grande variância do erro para boa parte dos dias do ano interpolados. Nos experimentos do ano de 2016 e 2017 ocorreu uma melhoria nos resultados a partir do mês de julho. Os experimentos obtiveram uma taxa muito variada de erros pelo fato de que não existe uma regularidade na esfera de dados de precipitação, outro motivo seria o fato de que, das 7 cidades, Recife possui uma quantidade de precipitação mais elevada do que das outras cidades, em muitos casos em determinado dia, ou até mês, ocorreu precipitações apenas em Recife. Outro fator para elevação dos erros foram a quantidade de estações, apenas 7 estações que estão dispostas em pontos bastante distintos, nos experimentos de Clodoalves (2018), foram utilizados dados de todas as estações meteorológicas do nordeste, e pode-se observar que os resultados foram mais satisfatórios. Com esses resultados, conclui-se que nesse contexto, a generalização do modelo não foi satisfatória, assim não utilizando os dados de interpolação para os experimentos de correlação com as outras bases dispostas a análise neste artigo.



**Figura 4. Erro quadrático médio diário - 2017**



**Figura 5. Erro quadrático médio diário - 2016**



**Figura 6. Erro quadrático médio diário - 2015**

#### **4. Dados de Saúde (Dengue)**

Para construir essa parte do estudo, foi utilizada a base de dados sobre casos de dengue registrados pelo Departamento de Informática do SUS - DATASUS, restringido apenas para os municípios do estado de Pernambuco referentes aos anos de 2010, 2011 e 2012. Esses dados foram comparados com as informações extraídas da base do INMET, citado no tópico acima e colhido os resultados.

Com a coleta e comparação desses dados, o objetivo foi conseguir entender se há relação entre a precipitação mensal e os registros, também mensais, de dengue nas cidades de Recife, Ouricuri, Cabrobó, Arcoverde, Triunfo, Surubim e Petrolina (localizadas no estado de Pernambuco). Para isso, foi aplicada a fórmula de correlação de Spearman entre essas duas tabelas de dados, selecionando as variáveis necessárias e executando os cálculos em cima delas.

##### **4.1 Correlação entre Clima e Dengue**

As duas bases passaram por um processo de tratamento de dados para que chegassem a um formato desejável para que o experimento pudesse ser realizado.

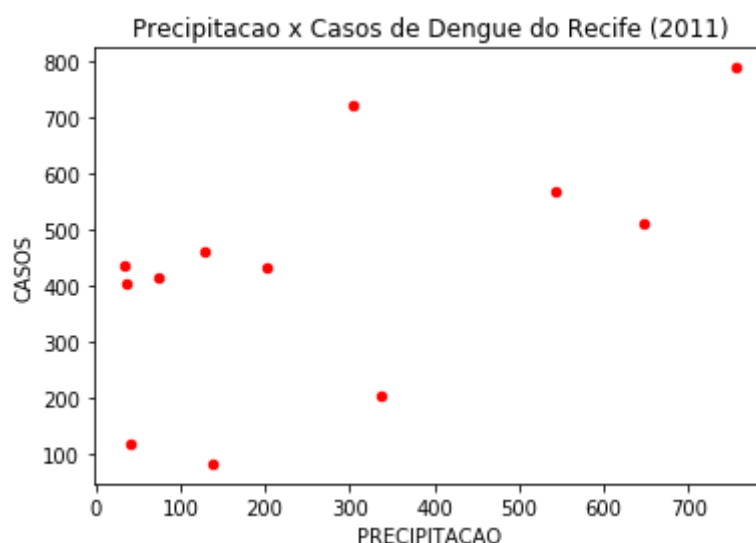
Como descrito no tópico acima, foi utilizada a série histórica de três anos, que abrange os anos de 2010 à 2012 para ambas as bases.

O método utilizado para o experimento foi o Coeficiente de Correlação de Spearman. Os valores de precipitação (mm) foram escolhidos como a variável X (variável independente) e os números de casos de dengue foram escolhidos como variável Y (variável dependente). A intenção é verificar se existe uma relação de dependência dos casos de dengue com o nível de precipitação de chuvas nas sete cidades de Pernambuco utilizadas no experimento.

Para cada cidade, exceto a cidade de Arcoverde, foi escolhido como amostra o ano que teve o maior nível de precipitação de chuvas (mm) dentre os anos da série histórica. No caso desta, por questões de insuficiência de dados no ano de maior quantidade de chuvas, foi escolhido então o ano com a menor precipitação. Após escolhido o ano da base de dados de precipitação a ser utilizado para cada cidade, foi selecionado o ano correspondente na base de dados de casos de dengue do mesmo município, para que os dados equivalentes pudessem ser relacionados e então foi aplicado o método do coeficiente de correlação de Spearman. Após a aplicação do método, os seguintes resultados foram obtidos:

**Tabela 1. Coeficiente de correlação de Spearman Precipitação (mm) x Casos de Dengue**

| <b>Cidade</b> | <b>Ano</b> | <b>Coeficiente (<math>\rho</math>)</b> | <b>Interpretação</b> |
|---------------|------------|--|----------------------|
| Arcoverde     | 2012       | 0,10639                                | Relação fraca        |
| Cabrobó       | 2011       | 0,13101                                | Relação fraca        |
| Ouricuri      | 2011       | -0,26276                               | Relação fraca        |
| Petrolina     | 2010       | 0,20979                                | Relação fraca        |
| Recife        | 2011       | 0,55944                                | Relação moderada     |
| Surubim       | 2011       | 0,55931                                | Relação moderada     |
| Triunfo       | 2011       | 0,40387                                | Relação moderada     |



**Figura 7. Gráfico de dispersão Precipitação x Casos de Dengue do Recife 2011**

Para a cidade do Recife, cidade que apresentou o maior coeficiente entre as amostras selecionadas, o ano com o maior nível de precipitação de chuvas (mm) presente na série histórica, foi o ano de 2011. Relacionando os dados de precipitação com os dados de casos de dengue do mesmo ano, conseguimos visualizar através do gráfico de dispersão, uma distribuição anormal dos dados. Aplicando o método do coeficiente de correlação de Spearman, chegamos a um coeficiente de 0,55944.... Este valor por si só, nos apresentou uma relação moderada entre a precipitação com os casos de dengue.

## **5. Acidentes de Trânsito**

Os dados referentes a acidentes de trânsito são disponibilizados pelo portal da Polícia Rodoviária Federal - PRF, que disponibiliza registros anuais de acidentes ocorridos em rodovias do Brasil. Para esse estudo, os dados analisados foram restritos ao município do Recife - PE nos anos de 2015 à 2017. Os registros de causas de acidentes foram comparados com os dados disponibilizados pelo INMET.

O objetivo da análise comparativa entre essas bases é compreender se há alguma correlação entre as precipitações mensais que ocorrem no município (variável independente X) e alguns fatores principais que provocam os acidentes na cidade (variável dependente Y), também no período mensal. Para essa análise estatística foi utilizado o coeficiente de correlação de Spearman, que mede a intensidade da relação entre variáveis aleatórias.

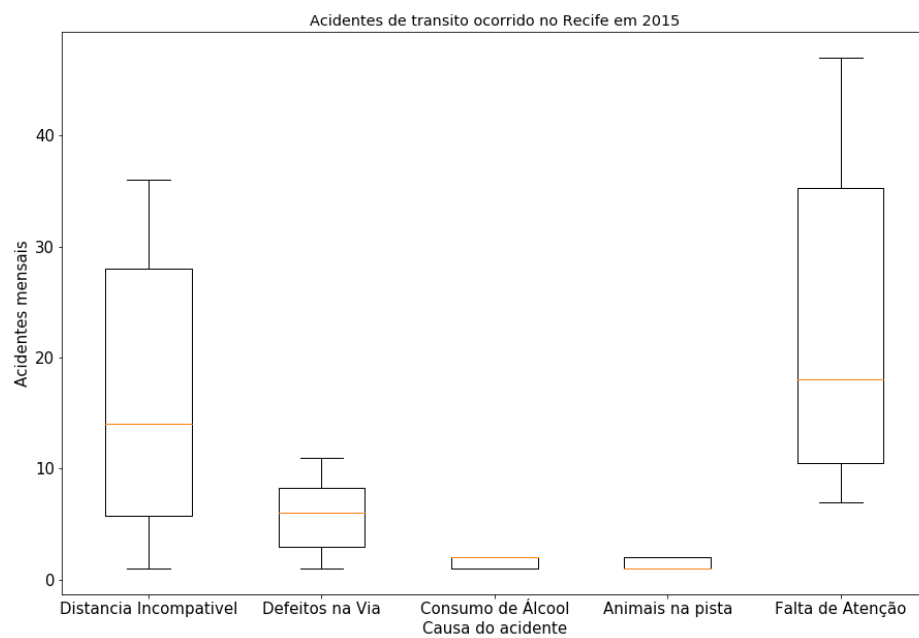
### **5.1. Correlação entre Clima e Acidentes de Trânsito**

Para montagem da análise de correlação foi utilizado como referência a metodologia de [Alves e Raia Jr 2012], onde se faz experimentos comparativos entre acidentes de trânsito, uso e ocupação do solo, polos geradores de viagens e população de Uberlândia - MG, contudo o método de correlação utilizado pelos autores não se aplica às variáveis utilizada nesse estudo, sendo substituído pelo coeficiente de spearman.

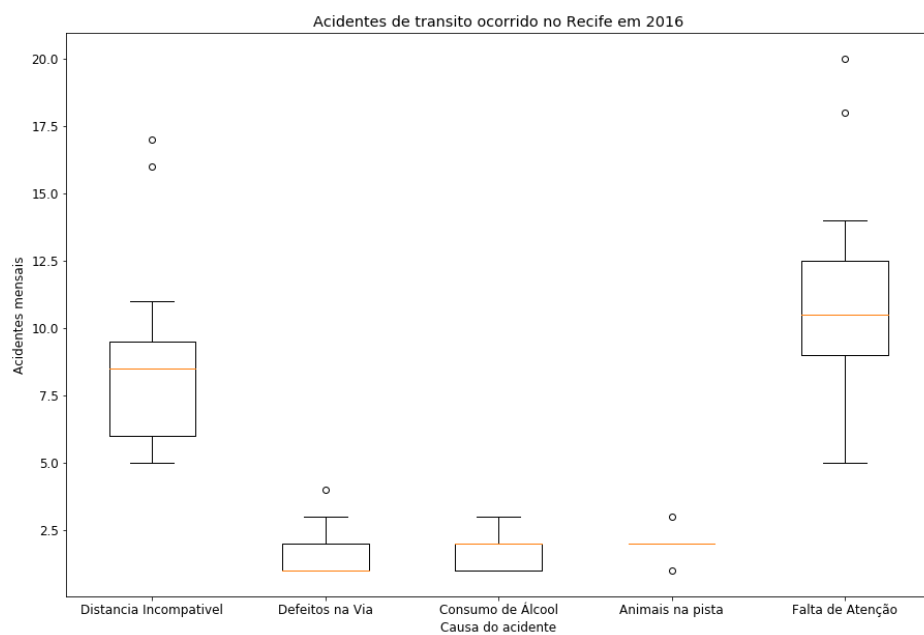
Para que a análise fosse realizada, foi necessário aplicar um tratamento nas bases devido ao padrão de organização de alguns registros sofrerem modificações durante os anos escolhidos para o estudo. O objetivo do tratamento era padronizar a série histórica dos dados para que esses estejam em um formato desejável para aplicar o método de correlação escolhido.

Através da análise estatística (correlação de Spearman) da precipitação com acidentes de trânsito causados por ingestão de álcool, falta de atenção do condutor, não manter a distância segura entre os veículos, defeito na via e animais na pista, verificou-se correlações fracas, moderadas e forte. A correlação forte ocorre na relação de acidentes causados por defeitos na via com as precipitações no município no ano de 2015. Também foi encontrado correlação moderada, do tipo negativa, para acidentes que tinham como causa animais na pista para o ano de 2016 como mostram as tabelas abaixo.

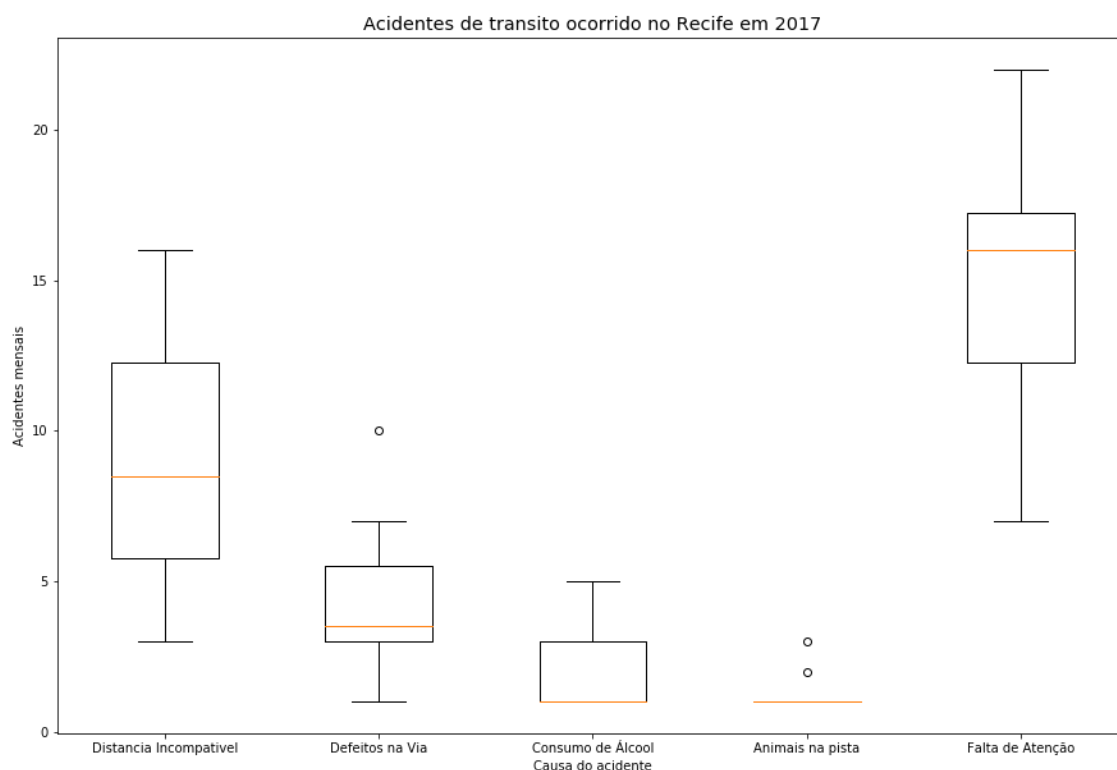




**Figura 8. Gráfico de caixa com causas de acidentes mensais analisadas em 2015**



**Figura 9. Gráfico de caixa com causas de acidentes mensais analisadas em 2016**



**Figura 10. Gráfico de caixa com causas de acidentes mensais analisadas em 2017**

**Tabela 1**

Calculo de correlação de Spearman para acidentes causados por ingestão de álcool x precipitação (Recife -PE)

| Ano  | Coeficiente ( $\rho$ ) | Interpretação    |
|------|------------------------|------------------|
| 2015 | -0,09                  | Correlação Fraca |
| 2016 | -0,08                  | Correlação Fraca |
| 2017 | 0,07                   | Correlação Fraca |

Nessa primeira etapa o objetivo era analisar se existia alguma relação dos acidentes causados pela ingestão de álcool com os períodos chuvosos do município. Assim poder investigar se há um aumento ou diminuição do consumo de álcool quando a precipitação está elevada. No entanto, como mostra a tabela 1, para a série temporal analisada as correlações tiveram um resultado fraco, quase sem relação entre os dados.

Tabela 2

Calculo de correlação de Spearman para acidentes causados pela falta de atenção do condutor x precipitação (Recife -PE)

| Ano  | Coeficiente ( $\rho$ ) | Interpretação    |
|------|------------------------|------------------|
| 2015 | 0,36                   | Correlação Fraca |
| 2016 | -0,04                  | Correlação Fraca |
| 2017 | 0,25                   | Correlação Fraca |

Tabela 3

Calculo de correlação de Spearman para acidentes causados pela aproximação insegura dos veículos x precipitação (Recife -PE)

| Ano  | Coeficiente ( $\rho$ ) | Interpretação    |
|------|------------------------|------------------|
| 2015 | 0,12                   | Correlação Fraca |
| 2016 | -0,26                  | Correlação Fraca |
| 2017 | 0,25                   | Correlação Fraca |

Nesta etapa foi investigado a atenção do condutor do veículo nos períodos de chuva do município e a aproximação insegura dos veículos nesse período. O objetivo era entender se existe relação dos acidentes causados pela distração do condutor ou aproximação incompatível dos veículos com as chuvas. Porém, as interpretações dos resultados obtidos para essa análise apontou uma correlação fraca para os anos analisados como mostrado nas tabelas 2 e 3.

Tabela 4

Calculo de correlação de Spearman para acidentes causados por defeito na via x precipitação (Recife -PE)

| Ano  | Coeficiente ( $\rho$ ) | Interpretação       |
|------|------------------------|---------------------|
| 2015 | 0,75                   | Correlação Forte    |
| 2016 | 0,14                   | Correlação Fraca    |
| 2017 | 0,56                   | Correlação Moderada |

A tabela 4 apresenta forte correlação no ano de 2015 para os acidentes que ocorreram na cidade do Recife causados por problemas nas vias de circulação de veículos e as chuvas que ocorreram neste ano no município. Uma observação importante sobre essa análise é a oscilação entre os valores das relações comparadas a cada ano, como mostrado acima em 2016 obtivemos uma correlação fraca com coeficiente muito inferior ao do ano de 2015, porém o coeficiente nos indica uma correlação moderada no ano de 2017. É importante destacar que periodicamente ocorrem operações de recapeamento e requalificação das vias no município, o que pode influenciar nos resultados dos dados, bem como a irregularidade das chuvas nessa região.

Tabela 5

Calculo de correlação de Spearman para acidentes causados por animais na pista x precipitação (Recife -PE)

| Ano  | Coefficiente ( $\rho$ ) | Interpretação       |
|------|-------------------------|---------------------|
| 2015 | 0,43                    | Correlação Moderada |
| 2016 | -0,5                    | Correlação Moderada |
| 2017 | 0,31                    | Correlação Fraca    |

Na tabela 5 são apresentados os resultados da relação dos acidentes provocados por animais nas vias de circulação de veículos nos períodos chuvosos. As correlações vão de moderada a fraca no período analisado, com um destaque para o ano de 2016, que apresentou uma correlação moderada negativa entre as variáveis analisadas, completamente oposto ao resultado apresentado no ano anterior.

## 6. Considerações Finais

Sobre a base de precipitação, concluímos que para obter um resultado mais satisfatório nas interpolações seria necessário dados de mais estações próximas aos pontos a serem interpolados.

Em relação ao método de Correlação de Spearman aplicado nas bases do INMET contendo os dados de precipitação das sete estações de Pernambuco com os dados de registros de casos de dengue destas mesmas cidades fornecidos pelo DATASUS, podemos verificar que para as amostras escolhidas, tivemos correlações fracas e moderadas. Nenhuma amostra obteve um coeficiente de correlação que pudesse ser interpretado como uma relação forte, o que nos leva a concluir que é improvável que haja uma correlação entre a precipitação da chuva e os casos de dengue nestas cidades.

Para a relação entre os registros de precipitação ocorridos no município do Recife com alguns fatores principais registrados como causa de acidentes foram identificadas valores de coeficientes de spearman baixos na maior parte da análise. No entanto alguns coeficientes apontaram altos valores para o coeficiente de spearman indicando forte relação entre as variáveis, contudo não se pode afirmar uma relação de causa e efeito entre os fatores analisados.

A construção e implementação dos experimentos deste artigo fizeram com que fosse possível obter conhecimentos sobre as bases utilizadas e sobre as técnicas e metodologias científicas aplicadas.

## 7. Referências

Borges de Paula, M. E., Duarte, Augusta. (1996) “Influência da chuva na ocorrência dos acidentes de trânsito”. Acesso em: 22 Jan. 2018

Clodoalves, J., (2018) **Avaliação de técnicas para interpolação espacial de dados de evapotranspiração**, Trabalho de Conclusão de Curso - Universidade Federal Rural de Pernambuco, Departamento de Estatística e informática.

Alves, Priscilla & Raia Jr, Archimedes. (2012). Análise de correlação entre acidentes de trânsito, uso e ocupação do solo, polos geradores de viagens e população em Uberlândia-MG. Revista dos Transportes Públicos. 34. 55-70.