



UNIVERSIDADE FEDERAL DO PIAUÍ
CENTRO DE CIÊNCIAS DA NATUREZA
CURSO DE GRADUAÇÃO EM ESTATÍSTICA

O uso da distribuição Poisson composta na teoria de valores extremos, para previsão de arrecadação de multas por excesso de velocidade.

Filipe Mateus de Sousa Costa

TERESINA - 2022

Filipe Mateus de Sousa Costa

Monografia:

O uso da distribuição Poisson composta na teoria de valores extremos, para previsão de arrecadação de multas por excesso de velocidade.

Monografia submetida à Coordenação do curso de graduação em Estatística, da Universidade Federal do Piauí, como requisito parcial para obtenção do grau de Bacharel em Estatística.

Orientador: Profa. Dra. Valmária Rocha da S. Ferraz

Co-orientador: Prof. Dr. Fernando F. do Nascimento

TERESINA - 2022

Agradecimentos

Primeiramente, eu preciso agradecer a Deus por ter permitido a possibilidade de realizar uma segunda graduação e ter me sustentado durante todo o curso. Foram vários os desafios, mas Ele permitiu que eu conseguisse superar tudo. Soli Deo Glória.

Preciso agradecer a minha esposa que ficou do meu lado durante este período, sacrificando-se para permitir que eu estudasse, motivando-me e superando o ciúme por ter sido trocada, muitas vezes, por um livro cheio de fórmulas esquisitas. Incluo meu filho Mateus, que chegou durante o curso para dar mais emoção à vida. Mesmo pequeno, me ensinou a ser mais produtivo. E claro, meu pai, minha mãe e minha irmã, que mesmo a distância, me apoiavam.

Agradeço a Professora Valmaria por ter aceitado o desafio de me orientar nesse TCC, juntamente com o professor Fernando que ofereceu o apoio teórico para a realização desse projeto. Incluo a STRANS, na pessoa do superintendente Claudio Pessoa, que gentilmente, disponibilizou os dados para este trabalho.

Finalizo os agradecimentos aos amigos que foram criados durante o curso. Edvaldo que, como eu, é chefe de família. Compartilhamos muito do desafio de trabalhar, cuidar do lar e estudar. Eva, cheia de energia, se esforçava para tirar as melhores notas nas disciplinas. Envelheceu alguns anos durante as madrugadas acordadas, mas ok.

Obrigado a todos, de coração.

É Cristo, e somente Cristo, que torna o encarar a morte suportável. Cristo infunde aos aspectos mais dolorosos e desconcertantes do fim desta vida esperança e paz. - Nancy Guthrie

Lista de Figuras

1	Distribuição de excesso na Alameda Parnaíba	21
2	Distribuição de excesso na Av. Raul Lopes	21
3	Distribuição de excesso na Av. Maranhão	22
4	Distribuição de excesso na Av. B. Castelo Branco	22
5	Boxplot comparativo das distribuições dos excessos	26
6	Parâmetros estimados Para Alameda Parnaíba	36
7	Av. Maranhão - Função fmgpd	37
8	Av. Raul Lopes - Função fmgpd	39
9	Av. Barão - Função fmgpd	40

Lista de Tabelas

1	Velocidade medida e Velocidade Considerada	15
2	Divisão do tipo de infração por excesso de velocidade a 60 km/h	15
3	Divisão do tipo de infração por excesso de velocidade a 40 km/h	16
4	Informações sobre os dados registrados por dia	18
5	Divisão por tipo de infração para os radares	18
6	Média e desvios-padrão diários por radar	19
7	Probabilidade de um dia ter 0 autos	20
8	Medidas descritivas das infrações de excesso por radar	20
9	Valor-p do Teste de Corridas	24
10	Teste de Dunn	25
11	Arrecadação Possível	26
12	Intervalos de credibilidade da média de autos diários	28
13	Comparação de DIC	34
14	Parâmetros para Alameda Parnaíba	35
15	Parâmetros para Av. Maranhão	37
16	Parâmetros para Av. Raul Lopes	38
17	Parâmetros para Av. Barão de C. Branco	39
18	Arrecadação diária	41
19	Arrecadação Mensal	42
20	Arrecadação Anual	42

Resumo

As infrações de excesso de velocidade no trânsito tem como finalidade criar uma medida punitiva e educacional aos condutores que excedem a velocidade máxima da via. Entretanto, existe o aspecto arrecadatário, influenciando sobre o orçamento do município. A Teoria dos Valores Extremos foi uma ferramenta importante para modelar a distribuição do excesso de velocidade determinando o valor possível de arrecadação. Além disso, utilizamos a distribuição de Poisson Composta para prever o número esperado de infrações diárias, possibilitando a previsão de arrecadação em nível diário, mensal e anual. Uma distribuição de mistura de Gamas com cauda GPD foi proposta para modelar a distribuição do valor do excesso de velocidade. O resultados indicaram valores bem significativos de arrecadação.

Palavras-Chave: Teoria de valores extremos; Infrações de Trânsito; Poisson Composta; Abordagem bayesiana.

Abstract

Excessive speeding violations in traffic are intended to create a punitive and educational action for drivers who exceed maximum speed allowed on the roads. However, there is the tax collection aspect, influencing the municipality budget. The Extreme Value Theory was an important tool used to model the distribution of over speeding by determining the possible amount of tax collection, in addition, we used the Compound Poisson Distribution to predict the expected number of daily infractions, enabling the tax collection forecast on a daily, monthly, and annual level. A distribution of mixtures of Gamma with GPD tail was proposed to model the distribution of the overspeed values. The results indicated very significant values on tax collection.

Key words: Extreme Value Theory; Traffic Violations; Compound Poisson Distribution; Bayesian approach.

Sumário

Lista de Figuras	6
Lista de Tabelas	7
1 Introdução	12
1.1 Teoria de Valores Extremos	13
1.2 Infrações de excesso de velocidade	14
1.3 Organização do trabalho	16
2 Detalhamento dos dados e análise descritiva	17
2.1 Análise descritiva do número de infrações diárias.	17
2.1.1 Análise descritiva do valor dos excessos.	20
2.2 Alguns testes	23
2.2.1 Teste de Corridas	23
2.2.2 Teste de Kruskal–Wallis	24
2.2.3 Pós-Teste	25
2.3 Arrecadação possível	26
3 Modelando o número diário de Infrações	27
3.1 Distribuição de Poison Composta	27
3.2 Inferência Bayesiana	28
3.3 Análise dos dados	28
4 Distribuição Pareto Generalizada (GPD) e Misturas de GPD	30
4.1 Domínio da atração	30
4.2 Determinação do Limiar	31
4.3 Estimação da GPD	32
4.4 Mistura de Gammas com GPD	33
4.5 Análise de Dados	34
4.5.1 Alameda Parnaíba	35
4.5.2 Av. Maranhão	36
4.5.3 Av. Raul Lopes - Shopping	38
4.5.4 Av. Barão de Castelo Branco	39
5 Estimação da arrecadação pela distribuição Poisson Composta	41

6	Conclusão	43
7	Referências Bibliográficas	44

1 Introdução

Infrações de trânsito são um elemento que visam limitar comportamentos que podem causar problemas para o funcionamento do fluxo público no trânsito. Como resultado de várias discussões visando uma melhor organização para as vias, surgiu o CTB - Código de Trânsito Brasileiro - que regulamenta todo um conjunto de regras que visam o bem público. As infrações de excesso de velocidade, por exemplo, visam limitar que os condutores circulem nas vias com velocidades muito altas, aumentando chances de acidentes.

Por outro lado, um dos resultados das infrações de trânsito, além da diminuição de situações de risco às pessoas, é o incremento de receita aos cofres públicos. Toda infração de trânsito, quando penalizada, resulta em uma multa de trânsito, que consiste num ônus ao cidadão autuado, um valor em dinheiro a ser recebido e utilizado pelas instituições públicas.

Este trabalho analisa os dados das autuações de 4 radares em Teresina através das medições que realizadas ao captar uma infração de excesso de velocidade. Dessa forma, buscaremos realizar uma previsão da arrecadação possível que estes radares podem oferecer ao município. Existem outros fatores que influenciam na arrecadação por autos de trânsito como o percentual de autos que são pagos ou a data de pagamento das penalidades em relação ao dia de vencimento da penalidade, mas esses outros fatores não pertencem ao objetivo de análise deste projeto.

As ferramentas estatísticas que utilizamos para a previsão orçamentária correspondem a, primeiramente, realizar uma análise descritiva dos dados, identificando possíveis problemas e falhas nos dados que possam prejudicar uma previsão eficiente. Em um segundo momento, iremos utilizar alguns modelos de distribuições probabilísticas.

Existe um conjunto de distribuições que trabalham com a ideia de ocorrência de um evento esperado, ou seja, calculam o tempo ou número de elementos necessários para que o objeto de análise ocorra. A distribuição de Poisson é usada para este tipo de análise e será a base de uma parte do estudo aplicado neste trabalho.

Eventos extremos são situações ou comportamentos que não ocorrem com tanta frequência. São aqueles que ficam próximos às caudas, distantes do pico onde estão

localizadas as medidas de tendência central. Entretanto, é possível que estes eventos ocorram. Na verdade, é esperado que eles ocorram. Diante dessa situação, é importante que possamos prever seus acontecimentos, pois quando ocorrem, as consequências podem ser trágicas.

Exemplos que demonstram com facilidade os efeitos de eventos extremos estão na área da climatologia. Para os portais de notícias, o aquecimento global tem trazido mais manchetes de catástrofes climáticas que causaram prejuízos substanciais e, infelizmente, mortes. Além da área citada, podemos encontrar estudos que envolvem valores extremos em áreas de pesquisa como engenharia de materiais, controle de tráfego e economia.

O surgimento da teoria de valores extremos surge da necessidade de pesquisas sobre eventos extremos de forma mais eficiente. As distribuições mais comuns trabalham melhor na análise dos eventos centrais, mais frequentes. Dessa forma, os acontecimentos mais raros ficam mais difíceis de prever. Em PINHEIROS (2014), podemos encontrar um resumo da teoria de valores extremos.

1.1 Teoria de Valores Extremos

O objetivo da teoria Valores Extremos é analisar valores extremos observados e prever possíveis valores ainda mais extremos, é fazer inferência para eventos cujas probabilidades são menores do que qualquer evento observado anteriormente. Fenômenos em que a probabilidade de um valor extremo é relativamente alta são caracterizados por distribuições com caudas pesadas.

A distribuição Pareto Generalizada (GPD) analisa a distribuição dos excessos de acordo com um limiar determinado. Esse formato de análise é mais eficaz pois evita a perda de informações que uma análise em períodos (ou blocos) pode gerar, afetando, principalmente, pesquisas com um grande volume de dados.

A GPD foi desenvolvida por Pickands [1975] é baseada no seguinte teorema:

Teorema 1: Se x for uma variável aleatória (v.a.) com função distribuição (f.d.) F_x , que pertence ao domínio de atração de uma distribuição GEV, então, quando $\mu \rightarrow \infty$, $F(x|u) = \Pr X > u + x | X > u$, possui distribuição GPD, com a seguinte função de distribuição:

$$P(x|\xi, \sigma, \mu) = \begin{cases} 1 - (1 + \xi \frac{(x-\mu)}{\sigma})^{-\frac{1}{\xi}}, & \text{se } \xi \neq 0 \\ 1 - \exp\left\{-\frac{(x-\mu)}{\sigma}\right\}, & \text{se } \xi = 0 \end{cases} \quad (1)$$

onde $\mu > 0, x - \mu \geq 0$, se $\xi > 0$, e $0 \leq x - \mu \leq -\frac{\sigma}{\xi}$, se $\xi < 0$. O caso $\xi = 0$ é interpretado como sendo o limite quando $\xi \rightarrow 0$, e tem como caso particular a distribuição exponencial de parâmetro $\frac{1}{\sigma}$. Os parametros são ξ , σ e μ representam, respectivamente, a forma, a escala e o limiar da distribuição.

Para este trabalho, as medições de excesso serão analisadas sobre a perspectiva de valores extremos, distiguindo de outras pesquisas de TVE mais tradicionais onde o limiar é estabelicido através de uma análise estatística. Entretanto, para esta pesquisa, consideramos todas a medições realizadas como extremos, baseado na norma de velocidade máxima determinada para a via.

1.2 Infrações de excesso de velocidade

Os dados trabalhados neste projeto têm como origem as medições de excesso de velocidade registradas por radares no município de Teresina. Baseado no Código de Trânsito Brasileiro (CTB), ruas, avenidas e estradas no Brasil possuem limites de velocidade especificados, no entanto, para a realização do registro de infração, é preciso que a via esteja sinalizada informando a velocidade máxima permitida.

Um ponto importante sobre as infrações de velocidade é a existência de uma margem de erro para medição da velocidade registrada excedente. Para fim de registro de infração, não é utilizado a velocidade obtida pelo instrumento de medição, mas subtrai-se um valor na velocidade medida (para medições até 100km/h, o valor reduzido é de 7 km/h). A velocidade final obtida após a subtração será a considerada para o registro ou não da infração.

Por exemplo, para vias com velocidade máxima de 60 km/h, as infrações serão registradas somente quando a velocidade calculada no ponto de medição for, pelo menos, de 68 km/h, registrando um excesso de 1km/h, mas para uma velocidade registrada de 67 km/h, o valor considerado é de 60km/h, não caracterizando uma infração. O resultado dessa distinção é o uso de dois termos: **Velocidade medida**, que consiste na velocidade Registrada e **Velocidade Considerada**, referente à velocidade medida menos o fator de correção. Na Tabela 1, temos a relação entre velocidade

medida e velocidade considerada de algumas velocidades, as mais comuns de vermos nas vias brasileiras. A infração é feita quando a velocidade considerada é maior que a velocidade permitida na via.

Tabela 1: Velocidade medida e Velocidade Considerada

Medida	Considerada
40	33
48	41
60	53
68	61
80	73
100	93
107	100

Em relação ao tipo de infração, o CTB normatiza três níveis baseado no percentual registrado na passagem do veículo. Para medições até 20% acima da permitida, a infração é média, entre 20% e 50% é considerada grave e, superior a 50%, é gravíssima. O aumento da gravidade resulta no aumento do valor pago da multa. Nas Tabela 1 e Tabela 2, temos um resumo da divisão de tipificação, o intervalo de excesso e o valor pago. A primeira refere-se para vias com velocidade máxima até 60 km/h e a segunda considera a máxima 40km/h.

- **Radares de 60 km/h**

Tabela 2: Divisão do tipo de infração por excesso de velocidade a 60 km/h

Tipo	Faixa percentual	Faixa de Excesso	Valor da Multa
Média	até 20%	$1 \leq v \leq 12$	R\$ 130,16
Grave	entre 20% e 50%	$13 \leq v \leq 30$	R\$ 195,23
Gravíssima	superior a 50%	≥ 31	R\$ 880,41

- **Radares de 40 km/h**

Tabela 3: Divisão do tipo de infração por excesso de velocidade a 40 km/h

Tipo	Faixa percentual	Faixa de Excesso	Valor da Multa
Média	até 20%	$1 \leq v \leq 8$	R\$ 130,16
Grave	entre 20% e 50%	$9 \leq v \leq 20$	R\$ 195,23
Gravíssima	superior a 50%	≥ 21	R\$ 880,41

1.3 Organização do trabalho

Este trabalho está dividido da seguinte maneira: no capítulo 2 iremos apresentar quais foram os pontos escolhidos para este trabalho. Realizaremos uma análise descritiva dos dados, apresentando a distribuição dos excessos e as medidas descritivas. Apresentaremos, também, medidas informações para seus comportamentos diários em relação a captação de excessos de velocidade.

No capítulo 3 faremos um resumo da distribuição de Poisson e seu funcionamento de forma conjunta. Apresentamos, também, o conceito de Inferência Bayesiana e realizaremos a estimação para o número de autos diários. No capítulo 4, apresentaremos a Teoria de Valores Extremos e uma noção da teoria de mistura de gammas para estimação de valores dos parâmetros da funções de distribuição envolvendo excessos.

Na parte final do trabalho, apresentamos os resultados obtidos e fazemos uma estimação dos valores diários, mensais e anuais que podem ser arrecadado em cada radar.

2 Detalhamento dos dados e análise descritiva

Os radares da cidade de Teresina possuem a função de identificar e registrar infrações, principalmente de transitar em vias destinadas ao transporte público e excesso de velocidade. O número exato de radares não é medida fácil de afirmar, pois existem fatores que influenciam na quantidade de radares em funcionamento, mas, em média, há 50 em funcionamento no município.

Para a realização deste trabalho escolhemos 4 endereços em pontos estratégicos da cidade. São eles:

- Alameda Parnaíba, próximo a Ponte Estaiada João Isidoro França – Zona Norte;
- Av. Raul Lopes, em frente ao Teresina Shopping – Zona Leste;
- Av. Maranhão, no trecho entre o centro Administrativo e ponte da Amizade – Zona Sul;
- Av. Barão de Castelo Branco, próximo a Igreja Católica do Cristo Rei. – Zona Sul.

Estes radares são bastante conhecidos no município de Teresina com velocidade máxima de 60 km/h, a exceção do radar da Av. Barão de Castelo Branco, onde o excesso é de 40 km/h. Seus bancos de dados são extensos e iremos apresentar algumas medidas descritivas para termos uma noção melhor da distribuição dos excessos.

É importante destacar que não é realizado uma distinção do sentido que o veículo está se locomovendo. Para análises futuras, pode-se aprofundar neste ponto. Destaco, também, que utilizaremos todos os registros realizados, independente se o auto não foi expedido por motivos técnicos ou administrativos.

2.1 Análise descritiva do número de infrações diárias.

Os dados utilizados apresentam algumas diferenças de período de registro por via, entretanto, todos os dados têm a data limite de 31/12/2021, sendo seus inícios variados, mas não anterior a 01/01/2016. Isto ocorre devido a fatores administrativos e técnicos. Como consequência, isso gerou um fator que não podemos mensurar facilmente que são os dias zerados para autuações, pois não podemos identificar se ocorreu algum problema técnico ou se realmente não houve nenhum registro naquele dia específico.

Na Tabela 4, apresentamos informações específicas de cada radar como a data

de início e fim dos registros, a quantidade de dias entre as datas e a quantidade de dias zerados.

Tabela 4: Informações sobre os dados registrados por dia

Endereço	Data de Início	Data Final	Dias	Dias Zerados
Alameda Parnaíba	05/05/2016	31/12/2021	2089	282
Av. Raul Lopes	12/05/2017	19/10/2020	1696	595
Av. Maranhão	05/05/2017	31/12/2021	1702	292
Av. B. C. Branco	26/09/2016	31/12/2020	1830	334

A Tabela 5 apresenta a quantidade de dias com registros de autuação, a quantidade de autos registrados em todo o período analisado e a divisão percentual por tipo de infração apresentada anteriormente, média, grave e gravíssima.

Tabela 5: Divisão por tipo de infração para os radares

Endereço	Dias	Nº de Autos	Percentual até 20%	Percentual entre 20% e 50%	Percentual acima 50%
Alameda Parnaíba	1807	48647	90.23	9.30	0.45
Av. Raul Lopes	1101	51605	92.75	7.03	0.22
Av. Maranhão	1410	27273	88.34	10.94	0.72
Av. B. de Castelo Branco	1496	30954	79.18	17.66	3.16

Todos os radares possuem mais de 1000 dias de registros, entretanto, os dias sem nenhuma autuação constituem uma fator presente em todos, destacando a Av. Raul Lopes, com 595 dias, mas é lá o ponto analisado com o maior número de autos registrados, 51.605 autos. Alameda Parnaíba possui um número próximo, com 48.647 autos em 1807 dias.

Evidencia-se que a primeira faixa de autos é dominantes nos registros de infrações com o maior percentual de de registros com um mínimo de 79,18% na Av. Barão de

C. Branco. Pelo fato de possuir uma velocidade máxima de 40 km/h, ocorre um fator de confundimento aos condutores causando uma proporção de excessos mais altos. Os outros pontos possuem proporções próximas, sendo a infração do tipo gravíssima menor que 1% em todos.

Uma informação complementar a Tabela 5 está na Tabela 6 e consiste na média e no desvio-padrão diária de autos por radar.

Tabela 6: Média e desvios-padrão diários por radar

Endereços	Média diária	Desvio padrão
Alameda Parnaíba	23.29	21.93
Av. Raul Lopes	30.43	25.91
Av. Maranhão	16.02	13.77
Av. B. de Castelo Branco	16.91	14.98

Como consequência do menor número de dias, mas com mais resgistros, Av. Raul Lopes apresenta a maior média diária de autos com o valor de 30,43 autos/dia, Av. Maranhão e Av. Barão de C. Branco obtiveram valores próximos na faixa de 16 autos/dia. Todos os radares obtiveram desvio-padrões muitos altos comparados as médias obtidas, indicando alta variabilidade.

Considerando a presença de dias zerados, é preciso realizar uma avaliação para determinar se estes dias devem ser considerados na análise. Com o valor da média, utilizamos a função exponencial para indicar a probabilidade de um dia conter 0 autuações. Determinando-se que a probabilidade é muito baixa, a decisão tomada foi pela exclusão destes dias na análise.

A função exponencial se estrutura como $f(x|\lambda) = \lambda e^{-\lambda x}, x \geq 0$ onde o parâmetro λ é a média da distribuição. No software R, utilizamos a função $pexp(x, \lambda)$ onde x é o valor que buscamos definir sua probabilidade de ocorrência e λ é a média. No caso desta pesquisa, a média diária de autos de infração de cada radar.

Tabela 7: Probabilidade de um dia ter 0 autos

Endereço	$p(0;\lambda)$
Alameda Parnaíba	1/12987052869
Av. Raul Lopes	1/1.638646e+13
Av. Maranhão	1/9102770
Av. Barão C. Branco	1/22181166

De acordo com a tabela 7, a função exponencial utilizada indicou que a probabilidade de um dia conter 0 autos é mínima, praticamente 0, dessa forma, podemos utilizar o vetor diário de autuações retirando os dias zerados, pois o alto número de zeros diários na amostra original foi causado por um não funcionamento do sistema nesses dias.

2.1.1 Análise descritiva do valor dos excessos.

O objetivo deste tópico é apresentar algumas medidas mais importantes dos radares que estamos analisando. Primeiramente, na Tabela 8 com as informações descritivas dos radares. Num segundo momento, construímos algumas imagens com as distribuições dos excessos dos radares.

Nos gráficos, as linhas pontilhadas referem-se ao ponto onde ocorre a mudança de tipificação das infrações de média para grave e de grave para gravíssima.

Tabela 8: Medidas descritivas das infrações de excesso por radar

Endereço	Média	Desvio Padrão	Mediana	Mínimo	Máximo
Alameda Parnaíba	5.61	5.39	4	1	67
Av. Raul Lopes	5.02	4.70	4	1	92
Av. Maranhão	6.05	5.94	4	1	70
Av. B. de Castelo Branco	5.74	5.89	4	1	109

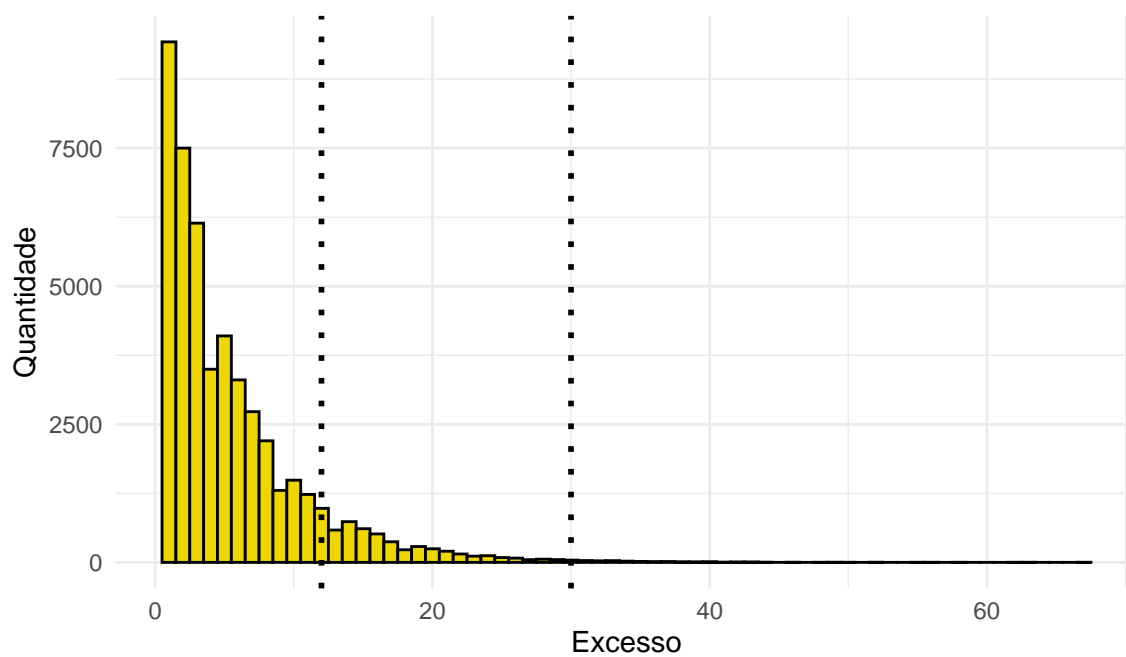


Figura 1: Distribuição de excesso na Alameda Parnaíba

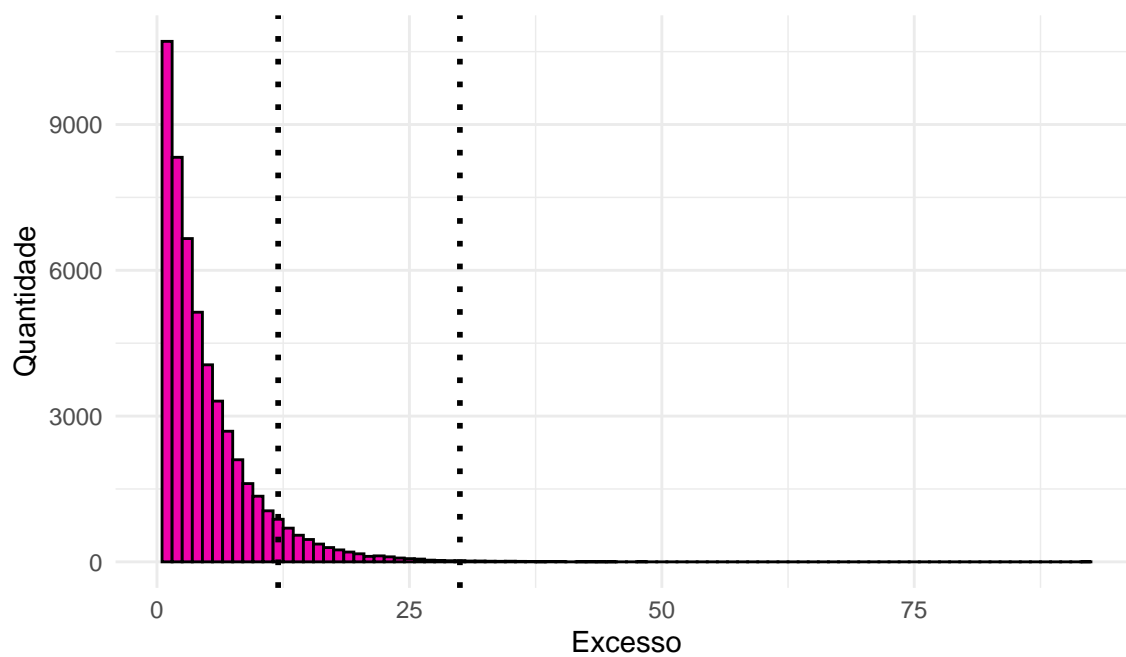


Figura 2: Distribuição de excesso na Av. Raul Lopes

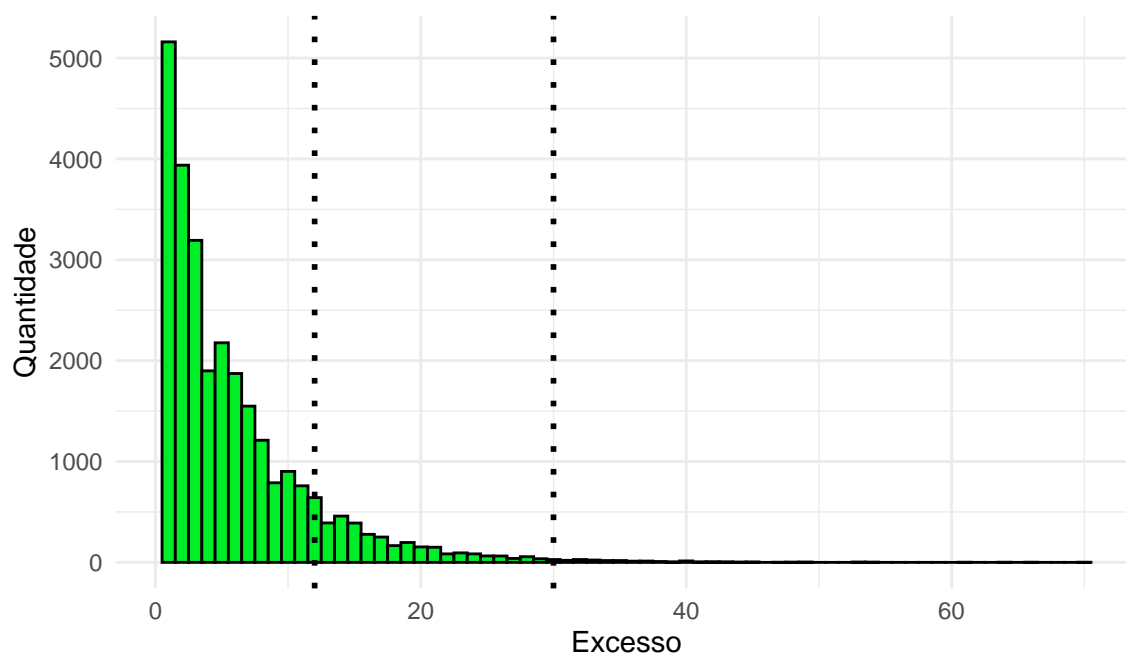


Figura 3: Distribuição de excesso na Av. Maranhão

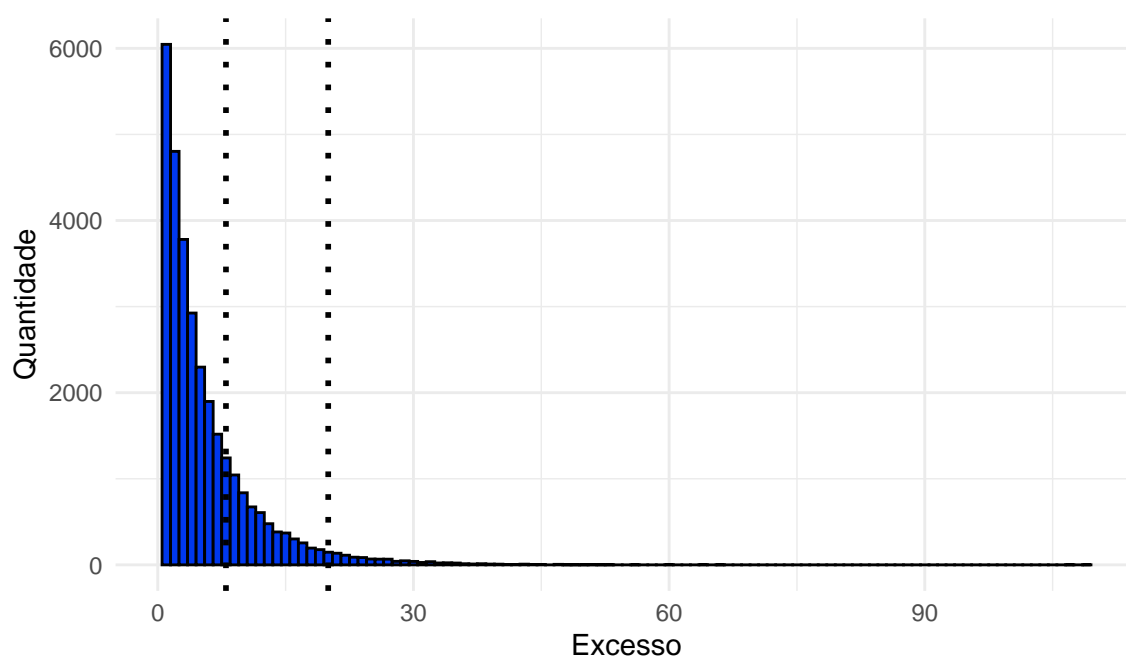


Figura 4: Distribuição de excesso na Av. B. Castelo Branco

Da tabela 8 pode-se identificar medidas semelhantes para todos os radares, como o mesmo mínimo e mesma mediana, com excesso de 4 km/h, indicando comportamentos semelhantes em relação a, pelo menos, 50% da distribuição dos radares. As médias são muito próximas, entre 5 e 6 km/h de excesso e o desvio padrão próximos da faixa de 5 km/h, o que, comparado a média são valores muito altos.

A medida que se diferencia entre os radares é o máximo excesso registrado com uma diferença entre o menor, na Alameda Parnaíba de 67 km/h, e o maior, na Av. Barão de B. Branco de 109 de 42 km/h. Analisando com base na velocidade considerada, esse excesso máximo é referente a uma velocidade de 149 km/h. Para os radares de 60 km/h, o maior excesso é na Av. Raul Lopes, com 92 km/h de excesso, resultado de uma velocidade considerada de 152 km/h.

Os graficos indicam comportamentos bem semelhantes na distribuição dos excessos. Utilizando as linhas pontilhadas que indicam a mudança de faixa, vemos a maioria das medidas estão na primeira faixa, de multas de nível médio. O único gráfico onde temos multas do tipo gravíssimo com um pouco de exposição é no radar do da Av. Barão de Castelo Branco, no outros radares, a perctagem é menor que 1%.

2.2 Alguns testes

Baseado em GIBBNS(2010), iremos realizar alguns testes, o primeiro teste buscar identificar se a ocorrência de multas dos excessos ocorre de uma forma aleatória ou não. Para isso, utilizaremos o teste de corridas, que se baseia no aumento ou diminuição de uma medição comparada com a medição anterior.

O segundo teste realizado será o de Kruskal-Wallis, com o objetivo de comparar a distribuição dos excessos dos radares e identificar se são semelhantes.

2.2.1 Teste de Corridas

O primeiro teste que iremos realizar tem como objetivo analisar se a distribuição em análise possui um comportamento aleatório ou não. De acordo com a resposta obtida por radar, podemos ter uma noção do funcionamento dos radares no registro do excessos.

O teste de corridas faz uma comparação de todos os valores medidos em posição ordenada e compara com a medição anterior analisando se o valor aumentou ou diminuiu e, a partir da sequência de aumentos ou diminuições, será obtido o número de corridas, sendo este o valor utilizado para determinar a aleatoriedade ou não do processo.

As respostas possíveis do teste aplicado são,

$H_0 = A \text{ sequência de dados é aleatória}$

$H_1 = A \text{ sequência de dados não é aleatória}$

Para a realização do teste, utilizando a função `DescTools::RunSTest()` no R e os resultados são apresentados a seguir:

Tabela 9: Valor-p do Teste de Corridas

Endereço	p-valor
Alameda Parnaíba	0.004
Av. Raul Lopes	0.009
Av. Maranhão	0.002
Av. Barão de C. Branco	0.028

Com os testes realizados, podemos afirmar que os registro dos radares não são aleatórios para todos os radares. Dessa forma, podemos construir uma análise que possa descrever o funcionamento dos radares.

2.2.2 Teste de Kruskal–Wallis

Outro teste que podemos realizar é o de Kruskal–Wallis, com ele iremos analisar a distribuição dos quatro radares buscando definir se o comportamento é semelhante ou não. Dessa forma, avalia-se de H_0 que determina o comportamento semelhante para todos as distribuições em análise ou H_1 , que determina que ao menos dois objetos de análise possui diferença. Este teste utiliza a posição dos medidas das distribuição conjuntas com a esperada.

Se o resultado indicar que existe diferença, utilizaremos um pós-teste (Teste de Dunn) para identificar entre quais ocorrem a diferença. Para realização do teste, utilizamos a função `Kruskal.test` do pacote `stats` do R

O resultado do teste indicou um p-valor muito baixo, 0, indicando que existe diferença entre as distribuições entre os radares. Assim, o próximo o passo é realizar o pós teste para identificar onde estão as diferenças.

2.2.3 Pós-Teste

Utilizamos os teste de Dunn, neste procedimento buscaremos avaliar em quais radares estão as diferenças significativas. Utilizamos a função `dunn_test` do pacote `rstatix`.

Tabela 10: Teste de Dunn

group1	group2	n1	n2	statistic	p
Alameda Parnaíba	Av. Barão C. Branco	48647	30954	-0.867789	0.3855098
Alameda Parnaíba	Av. Maranhão	48647	27273	7.843274	0.0000000
Alameda Parnaíba	Av. Raul Lopes	48647	51605	-13.916882	0.0000000
Av. Barão C. Branco	Av. Maranhão	30954	27273	7.903749	0.0000000
Av. Barão C. Branco	Av. Raul Lopes	30954	51605	-11.355477	0.0000000
Av. Maranhão	Av. Raul Lopes	27273	51605	-19.672889	0.0000000

O pós teste indicou que somente Alameda Parnaíba e Av. Maranhão possuem um comportamento de distribuição parecidos com um p-valor de 0,385, a comparação entre os outros mostrou diferenças significantes.

Completando esta análise comparativa, construímos um gráfico Boxplot comparado as distribuições dos excessos. Na Figura 5, podemos identificar que existe um comportamento semelhante entre as medições do radares, entretanto, podemos enxergar diferenças nos valores outliers, que certamente influenciaram no cálculos dos testes utilizados.

- Gráfico comparativo

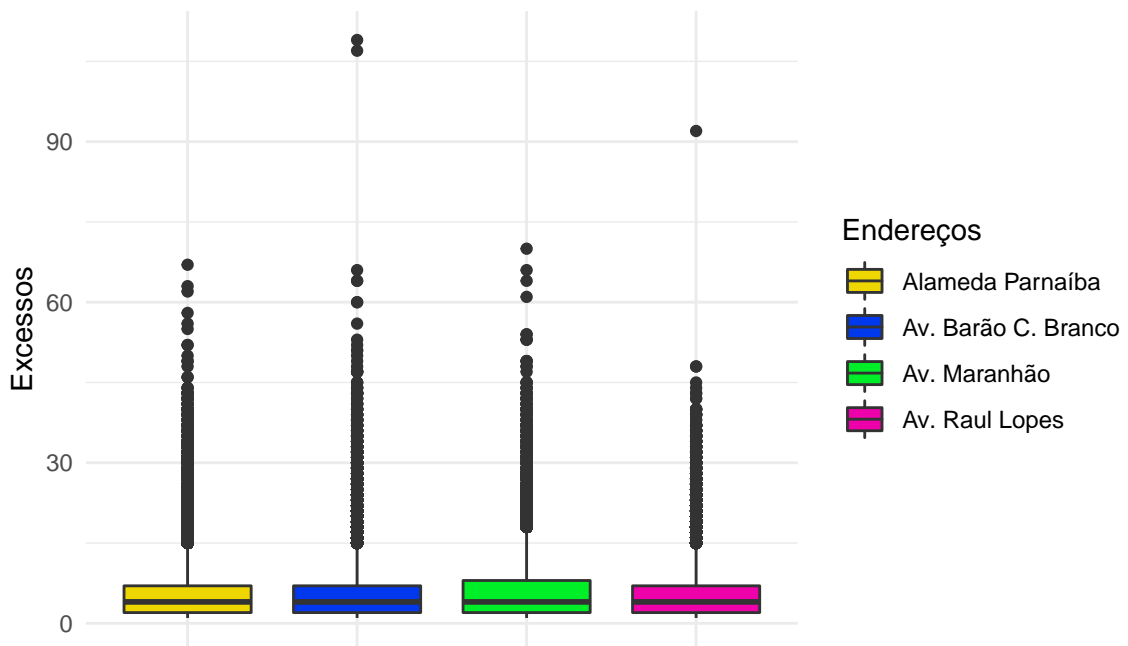


Figura 5: Boxplot comparativo das distribuições dos excessos

2.3 Arrecadação possível

O último ponto da análise descritiva e que consiste o objetivo final do trabalho, refere-se a arrecadação possível. A idéia do trabalho é utilizar as informações detalhadas acima para que seja possível realizar uma previsão de arrecadação e, desta forma, aperfeiçoar a forma de administração das receitas e despesas públicas. Abaixo, temos a arrecadação possível se todas as multas fosses pagas com o valor integral.

Tabela 11: Arrecadação Possível

Local	Valores
Alameda Parnaíba	R\$ 6.794.927
Av. Raul Lopes	R\$ 7.036.259
Av. Maranhão	R\$ 3.891.757
Av. B. de Castelo Branco	R\$ 5.119.140

Os valores são bem significativos, sendo o de maior poder arrecadatório a Av. Raul Lopes, com R\$ 7.036.259. O menor valor está na Av. Maranhão com R\$ 3.891.757. A soma total de arrecadação é de R\$ 22.842.083.

3 Modelando o número diário de Infrações

Inicialmente, iremos modelar a variável N , referente ao número diário de infrações. Na próxima seção, serão modelados os valores de excesso X_i utilizando TVE.

A distribuição Poisson pertence a família das distribuições quantitativas discretas. Seus resultados indicam a probabilidade de um evento ocorrer após um período de tempo, segundo CASELLA (2022), "uma das suposições básicas sobre a qual este modelo é desenvolvida é a de que, para pequenos intervalos de tempo, a probabilidade de uma chegada é proporcional ao tempo de espera.", em outras palavras, o tempo que se aguardar para um evento influencia o resultado.

A distribuição Poisson possui um único parâmetro, λ , que corresponde a um parametro de intensidade. Sua função de distribuição é:

$$P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \dots \quad (2)$$

Para esta distribuição, temos que $E(X) = \lambda$ e $VAR(X) = \lambda$

3.1 Distribuição de Poison Composta

Supondo que $N \sim Poisson(\lambda)$ e N seja um conjunto de variáveis baseada em *Poisson* com valor esperado de λ e X_1, X_2, X_3, \dots sejam identicamente distribuídas e independentes, então a distribuição de probabilidade da soma de N

$$Y = \sum_{n=1}^N x_n \quad (3)$$

é uma distribuição de Poisson composta.

A esperança dessa distribuição é dada por:

$$E(Y) = E(N)E(X) \quad (4)$$

e a variância por:

$$VAR(Y) = E(N)Var(X) + (E(X))^2Var(N) \quad (5)$$

3.2 Inferência Bayesiana

Uma diferente perspectiva de análise estatística foi iniciada com a *Inferência Bayesiana*, cujo os valores iniciais são originados por distribuições incertas e descritas por uma função de densidade em relação aos parâmetros da função utilizada.

$$\pi(\theta|x) = \frac{f(x|\theta)p(\theta)}{\int f(x|\theta)p(\theta)d\theta} \quad (6)$$

Preferencialmente, com algum conhecimento do parâmetro, o pesquisador insere a informação na pesquisa através de uma densidade $p(\theta)$, denominada de distribuição a priori.

As distribuições sob as regras da inferência Bayesiana contém dois elementos na sua composição: uma distribuição observacional $f(x|\theta)$ e a distribuição $p(\theta)$, sendo esta especificada com a ajuda de parâmetros da distribuição de parâmetros, ou hiperparâmetros.

3.3 Análise dos dados

Utilizando software R na versão 4.2.1, consideramos uma posterior e calcularmos intervalos de credibilidade para média diária de autos, dessa forma, temos que $\lambda|x \sim G(a_1, b_1)$ onde $a_1 = \sum x_i + a$ e $b_1 = n + b$. Assim $\hat{\lambda} = \frac{a_1}{b_1} = \frac{\sum x_i + a}{n + b}$, se a e b são pequenos ($Ex.0.0001$), resultando em $\hat{\lambda} \cong \bar{X}$

Tabela 12: Intervalos de credibilidade da média de autos diários

Endereços	A1	B1	$\hat{\lambda}$	L. I.	L. S.
Alameda Parnaíba	48647	1807.001	26.921	26.683	27.161
Av. Raul Lopes	51605	1101.001	46.871	46.467	47.276
Av. Maranhão	27273	1410.001	19.343	19.114	19.573
Av. Barão	30954	1496.001	20.691	20.461	20.922
c. Branco					

Os valores limites obtidos na Tabela 12 são baseados em um intervalo de confiança de 95%. Comparando os valores, obtemos o maior λ para Av. Raul Lopes

com 46.87 autos dia e o menor para a Av. Maranhão 19.46 autos dia, entretanto, muito próximo a Av. Barão de C. Branco com 20.69 autos dia. Alameda Parnaíba, com 26,92, evidencia como a Av. Raul Lopes é um ponto com grande número de autos, com uma média quase 2 vezes maior que os outros.

Neste trabalho, iremos considerar a distribuição de Poisson composta como em (3), onde a variável X_n possui uma distribuição de valores extremos que será apresentada na seção seguinte.

4 Distribuição Pareto Generalizada (GPD) e Misturas de GPD

Nesta seção iremos modelar os valores dos excesso de velocidade das infrações, para isso utilizamos a distribuição da GPD

A função de densidade da distribuição GPD é dada por:

$$p(x|\xi, \sigma, \mu) = \begin{cases} \frac{1}{\sigma}(1 + \xi \frac{x-\mu}{\sigma})^{-\frac{1}{\xi}}, & \text{se } \xi \neq 0 \\ \frac{1}{\sigma} \exp\left\{-\frac{(x-\mu)}{\sigma}\right\}, & \text{se } \xi = 0 \end{cases} \quad (7)$$

onde $x - \mu > 0$ para $\xi \geq 0$ e $0 \leq x - \mu < -\frac{\mu}{\xi}$ para $\xi < 0$.

A distribuição GPD possui as seguintes características em relação aos parâmetros:

$$E(X) = \frac{\sigma}{1-\xi}; \xi < 1; Md(x) = \frac{\sigma(2\xi - 1)}{\xi}; V(X) = \frac{\sigma^2}{(1-\xi)^2(1-2\xi)} \quad (8)$$

Justificando o uso da GPD Pickands (1975) e Davidson e Smith (1990) demonstram as propriedades e provam que GPD é única que satisfaz estas propriedades. “Por exemplo, estabilidade do limiar, ou seja, se Y possui distribuição GPD, e se $\mu > 0$, então a distribuição de $P(Y - \mu | Y > \mu)$ também possui distribuição GPD” Nascimento (2012).

4.1 Domínio da atração

As distribuições de valores extremos são obtidas como distribuição limite ($n \rightarrow \infty$) do máximo de um conjunto de variáveis aleatórias (v.a.s) independente e identicamente distribuídas (i.i.d) e são unicamente determinadas, a menos de transformações afins. O teorema de Fisher-Tippet implica que se $F_x^n(C_n x + d_n)$ não degenera quando ($n \rightarrow \infty$), para certas constantes $C_n > 0$, $d_n \in \mathbb{R}$, então

$$|F_x^n(x) - H(\frac{x - d_n}{c_n})| \rightarrow 0, n \rightarrow \infty \quad (9)$$

para alguma distribuição H . A coleção das distribuições F_x tais que os respectivos máximos possuem a mesma distribuição limite é chamada de domínio de atração.

Definição 2.1.2 Se (função) se verifica dizemos que F_x pertence ao domínio de atração do

máximo da distribuição de valores extremos H. Notação: $F_x \in MDA(H)$

Existem 3 casos possíveis para as distribuições limites das excedências de um limiar. Para domínio do tipo I ($\gamma = 0$), a distribuição se torna

$$H(y) = 1 - e^{-\frac{y}{\sigma}}, y > 0 \quad (10)$$

sendo assim, o domínio, uma distribuição Exponencial com parâmetro $\frac{1}{\sigma}$. Para o domínio tipo II ($\gamma > 0$), a distribuição limite será a distribuição de Pareto. Já para o domínio tipo III ($\gamma < 0$), quando $\sigma = -\frac{1}{\gamma}$, a distribuição limite será uma Beta e quando $\sigma \neq -\frac{1}{\gamma}$, a distribuição limite será uma Beta reescalada com suporte em $(0, \frac{\sigma}{\gamma})$.

4.2 Determinação do Limiar

A análise via GPD exige um cuidado inicial pois é preciso determinar um limiar para os dados. O valor escolhido pode ser determinado pelo pesquisador, entretanto, correm-se riscos que podem influenciar os cálculos, resultando em análises incorretas.

A escolha de um limiar " μ " muito alto implica em um número muito pequeno de observações resultando em estimadores com grande variabilidade. Um limiar muito pequeno resulta na violação do Teorema de Pickands (1945), modelando de forma errada os valores com limiar baixo, dessa forma, não se garante a convergência dos excessos Y para a família da GPD, levando a um vício alto.

Métodos mais convencionais de determinação do limiar utilizam-se de análises gráficas da linearidade de N_u . Um método muito utilizado é o gráfico de médias de excessos (MRL < *Mena Residual Life Plot*), baseado na espera da GPD, NASCIMENTO (2012). Sua construção segue o seguinte formato:

$$\left\{ \left(\mu, \frac{1}{n_u} \sum_{i=1}^{n_u} \right) : \mu < x_{max} \right\} \quad (11)$$

onde $x_1 \leq x_2 \leq \dots x_n$ consistem nas N_u observações que excedem μ , e x_{max} é o valor mais elevado das observações.

Considerando a distribuição GPD válida para os excessos, esta também é válida para os excessos acima de todos os limiares $\mu > \mu_0$, sujeito a mudanças no parâmetro de escala $\sigma_\mu = \sigma_{\mu_0} + \xi_\mu$. Então, para $\mu > \mu_0$

$$E(X - \mu | X > \mu) = \frac{\sigma_\mu}{1 - \xi} = \frac{\sigma_{\mu_0} + \xi_\mu}{1 - \xi} \quad (12)$$

Se o modelo é adequado a partir de μ_0 o gráfico apresentará um comportamento linear a partir de u . Um problema recorrente com a utilização desse gráfico é que o limiar pode limitar o número de excessos devido a escolha de limiar muito alto.

Outra técnica gráfica utilizada é Dispersion Index Plot (DIP), baseado em CUNNANE (1979) (Citado por NASCIMENTO (2012)), que diz que, o número de excessos sobre um limiar alto em um determinado período (geralmente meses ou anos), pode ser distribuído através de um processo de Poisson. Assim, a razão entre a variância e a média é igual a 1. Assim, pode-se fazer um gráfico

$$\left\{ \left(\mu, \frac{Var(Y)}{E(Y)} \right) : \mu < x_{max} \right\} \quad (13)$$

4.3 Estimação da GPD

Após determinar o limiar, a estimação dos parâmetros da distribuição GPD podem ser estimados por vários métodos, entre eles, tem-se o da máxima verossimilhança. Existem outros métodos como de momentos proposto por SMITH(1987) (citado por MENDES (2004)) e o métodos dos momentos ponderados SINGH e GUO (1995) (citado por MENDES (2004)), em que a eficiência de cada método depende da situação estudada.

Os estimadores de máxima verossimilhança (EMV) que maximizam a função de log-verossimilhança, quando $\xi \neq 0$, é dado por:

$$l(\mu, \xi) = -n_u \log(\sigma) - \left(1 + \frac{1}{\xi} \sum_{i=1}^{n_u} \log\left(1 + \xi \frac{y_i}{\sigma}\right)\right) \quad (14)$$

definida em $(1 + \xi \frac{y_i}{\sigma}) > 0$, para todo $i = 1, 2, \dots, N_u$. No caso particular onde $\xi = 0$, a a log-verossimilhança é dada por

$$l(\sigma) = -n_u \log(\sigma) - \sum_{i=1}^{n_u} \left(\frac{y_i}{\sigma}\right) \quad (15)$$

No caso de $\xi = 0$, a maximização dos parâmetros não pode ser obtida analiticamente, sendo necessárias técnicas numéricas de maximização.

4.4 Mistura de Gammas com GPD

Para valores maiores que um determinado limiar, os dados são estimados por GPD. Para distribuição abaixo do limiar, podemos propor diversas abordagens, sendo uma delas a mistura finita de distribuição.

NASCIMENTO (2012) propôs um modelo para dados extremos que utiliza aproximação não-paramétrica baseado em mistura de distribuição Gama para valores extremos maiores que um limiar e GPD para cauda dos dados, a função densidade é definida como:

$$f(x|\theta, p, \Psi) = \begin{cases} h(x|\mu, \eta, p), & \text{se } x \leq \mu \\ (1 - H(x|\mu, \eta, p))p(x|\Psi), & \text{se } x > q\mu \end{cases} \quad (16)$$

onde H é a função de distribuição acumulada da mistura de Gammas, $\Psi = (\xi, \sigma, \mu)$, $\sigma > 0$, $(x - \mu) \leq -\sigma/\xi$ e $x > \mu$.

O métodos de Nascimento (2012) realiza a estimação dos parâmetros utilizando MCMC -Markov chain Monte Carlo - e em @Nascimento 2012, encontramos o procedimento de estimação.

Em valores extremos, além de encontrar a estimativa dos parâmetros do modelo, também é importante encontrar uma forma para determinar os quantis altos, acima do limiar, de tal forma que se X possui distribuição GPD, é importante saber com qual probabilidade ocorre um evento maior ou igual a q , ou seja, $P(X > q) = 1 - q$.

Com os cálculos destes quantis, podemos realizar previsões com os dados de autos de excesso velocidade de trânsito de Teresina nos próximos anos, considerando uma manutenção da estrutura dos radares e, incluindo novos endereços de medição, uma previsão de quantos autos poderão ser registrados. Outra variável que pode ser analisada consiste no excessos de velocidade, determinando possíveis valores máximos e a previsão da quantidade de infrações.

Na distribuição GPD, pode-se encontrar um quantil q com probabilidade $P(X < q)$ em função dos parâmetros. Invertendo a função acumulada, obtém-se a seguinte função dos quantis da cauda:

$$q_x p = \frac{((1 - p^*)^{-\xi} - 1)}{\xi} \quad (17)$$

onde $p^* = 1 - (1 - p)N/N_u$.

4.5 Análise de Dados

Para realização deste trabalho realizamos testes para determinar qual seria o número de Gammas que ofereceria o modelo eficiente. Utilizamos DIC (*Deviance Information Criterion*) para apontar quantas componentes na mistura de Gamma seriam utilizados.

A Tabela 13 tem como resultado que para todos os modelos, a utilização de 2 Gammas é a mais eficiente para distribuições de excessos analisadas.

Tabela 13: Comparação de DIC

Endereço	K = 1	K = 2
Alameda Parnaíba	247517.2	218385.2
Av. Raul Lopes	253704.1	210970.1
Av. Maranhão	145195.4	122570.4
Av. Barão C. Branco	160787.6	157461.5

Nos próximos tópicos apresentamos os parâmetros estimados para cada radar utilizando 2 Gammas.

Para estimação dos parâmetros dos modelos, foi utilizado a função `fmgpd` do pacote `extrememix`, criado por Manuele Leonelli, este pacote utiliza a estimação bay-siana para modelos de valores extremos utilizando algoritmos de MCMC - Simulação Monte Carlo via cadeias de Markov. Com este pacote também podemos estimar o limiar que definirá o início da utilização da função GPD, outras medidas importantes como quantis altos também podem ser conhecidas.

Uma segunda fase é utilizar funções do pacote `POT` criada por Mathieu Ribatet, é usada no software R e fornece algumas funções utilizadas em técnicas de análise univariada e bivariada de valores extremos, em especial a distribuição GPD, além de alguns recursos gráficos para determinação do limiar. Neste trabalho, não iremos

determinar o limiar pois o modelo proposto de estimação Bayesiana de NASCIMENTO (2012) estima o limiar.

4.5.1 Alameda Parnaíba

Para Alameda Parnaíba, o valor de ξ é estabelecido em -0,06, σ em 6,33, o limiar em 3,01. Através desses valores estabelecidos podemos calcular que o limite máximo de velocidade é de 55,33 km/h. Como o limite máximo registrado é 67 km/h, pode-se dizer que o condutor desse veículo conseguiu superar o limite do modelo criado neste trabalho.

Tabela 14: Parâmetros para Alameda Parnaíba

	Estimado	L.I.	L.S.
ξ	-0.06	-0.07	-0.05
σ	6.33	6.18	6.45
μ	3.01	3.01	3.01
μ_1	1.00	1.00	1.00
μ_2	3.55	3.52	3.57
η_1	219.16	219.05	219.18
η_2	10.07	9.85	10.40
W_1	0.19	0.19	0.19
W_2	0.81	0.81	0.81

Através dos gráficos de estimação de valores dos parâmetros, visualizamos que ξ possui um comportamento semelhante a uma escada, tendo a média estabelecida em 0,061, para σ , o comportamento é mais variado e os valores estimados são próximos dos empíricos.

Para os valores acima do μ (3,01) estabelecido, podemos utilizar a função `fitgpd` para analisar a divisão da distribuição e medidas importantes para o modelos. Para a utilização, `udsmod` o vetor dos excessos e o limiar determinado pela função `fmgpd` do pacote `extrememix`.

Verificamos que os valores obtidos para os estimadores ξ e σ foram, respectivamente, -0,056 e 6,311, valores válidos para o intervalo de confiança da função `fmgpd`.

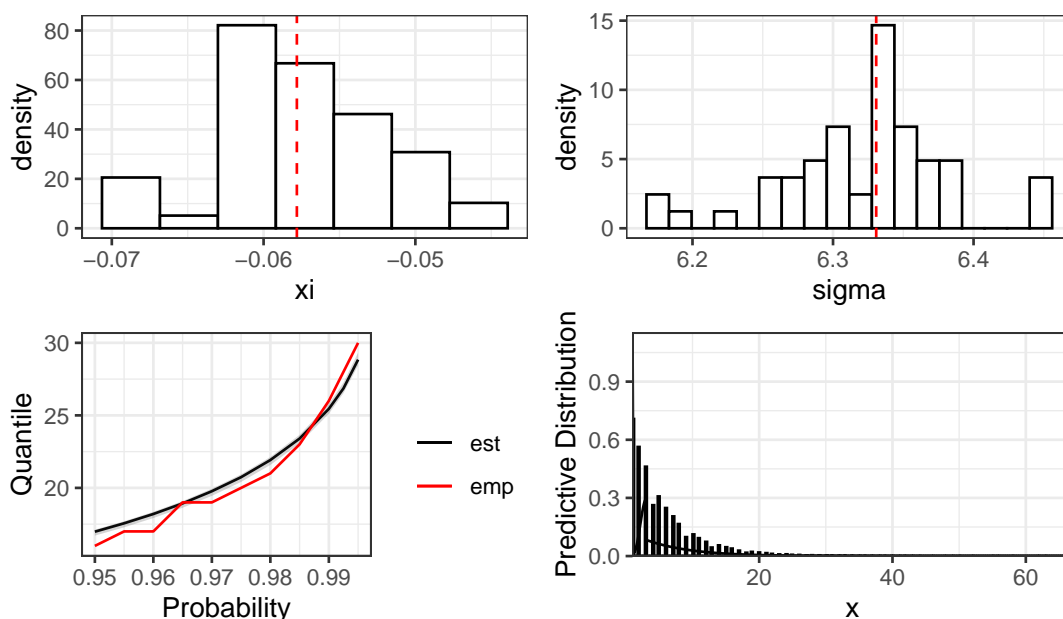


Figura 6: Parâmetros estimados Para Alameda Parnaíba

Temos uma proporção de 52,28% acima do limiar, valor bem distinto daqueles que se costuma usar como valor extremo, que normalmente supera o quantil de 90%.

Com a distribuição dos excessos, outro ponto fundamental é a divisão baseada nas faixas de excesso das atuações de trânsito. Como na Tabela 3 vimos, a divisão dos excessos pode ser dividida em três níveis e influência no resultado da arrecadação.

Para a Alameda Parnaíba, 87,94% das atuações se encontram no primeiro nível de atuação, 11,67% consiste em multa do tipo grave e o último nível, gravíssima, corresponde a 0,39%.

4.5.2 Av. Maranhão

Para a Av. Maranhão, temos o limiar mais baixo dos radares que estamos analisando, de 2, o que corresponde a um total de 66,64% de todos elementos acima do limiar, ou seja, trabalhadas na função GPD, somente os excessos de 1 km/h e 2km/h teriam distribuição não-paramétrica com mistura de gammas.

O maior excesso registrado empiricamente na Av. Maranhão foi de 70 km/h, no entanto, temos um ξ positivo, o que significa que não há um limite definido para o excesso neste ponto.

Tabela 15: Parâmetros para Av. Maranhão

	Estimado	L.I.	L.S.
ξ	0.19	0.16	0.21
σ	4.25	4.17	4.35
μ	2.00	1.99	2.00
μ_1	1.00	1.00	1.00
μ_2	10.36	8.30	11.93
η_1	353.15	353.04	353.17
η_2	10.40	8.54	11.94
W_1	0.19	0.19	0.19
W_2	0.81	0.81	0.81

No gráfico de avaliação dos parâmetros, observamos comportamentos que não seguem um formato de normalidade para ξ e σ e a curva de estimação não acompanha a curva de empírica, estando fora do intervalo de confiança acima do quantil de 98%.

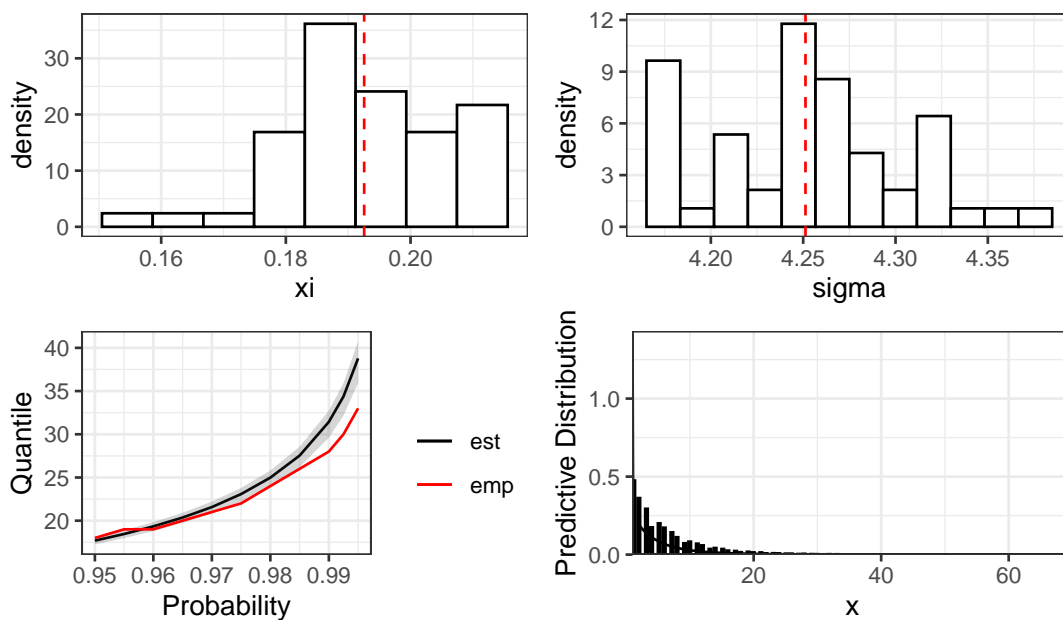


Figura 7: Av. Maranhão - Função fmgpd

Baseado na divisão dos excessos dos autos, os resultados obtidos indicam 88,29% para o nível médio, 10,51 para nível grave e 1,16% para nível gravíssima.

4.5.3 Av. Raul Lopes - Shopping

Para a Av. Raul Lopes, temos o μ em 3, o que resulta numa divisão de 50% dos dados desse radar serão trabalhados com mistura de Gammas e a outra metade, superior ao limiar determinado será estimado com gpd. A estimação de ξ é -0,04 e σ é 5,40, resultando num limite de velocidade máxima de excesso de 60 km/h

Tabela 16: Parâmetros para Av. Raul Lopes

	Estimado	L.I.	L.S.
ξ	-0.04	-0.05	-0.04
σ	5.40	5.32	5.46
μ	3.00	3.00	3.00
μ_1	1.00	1.00	1.00
μ_2	3.47	3.46	3.49
η_1	392.15	392.07	392.24
η_2	10.35	10.12	10.47
W_1	0.21	0.20	0.21
W_2	0.79	0.79	0.80

No gráfico Av. Raul Lopes - Função fmgpd vizualiamos que o estimador de ξ possui um comportamento mais centralizado no valor estimado de 0,04, para σ , temos um comportamento com uma variação mais alta.

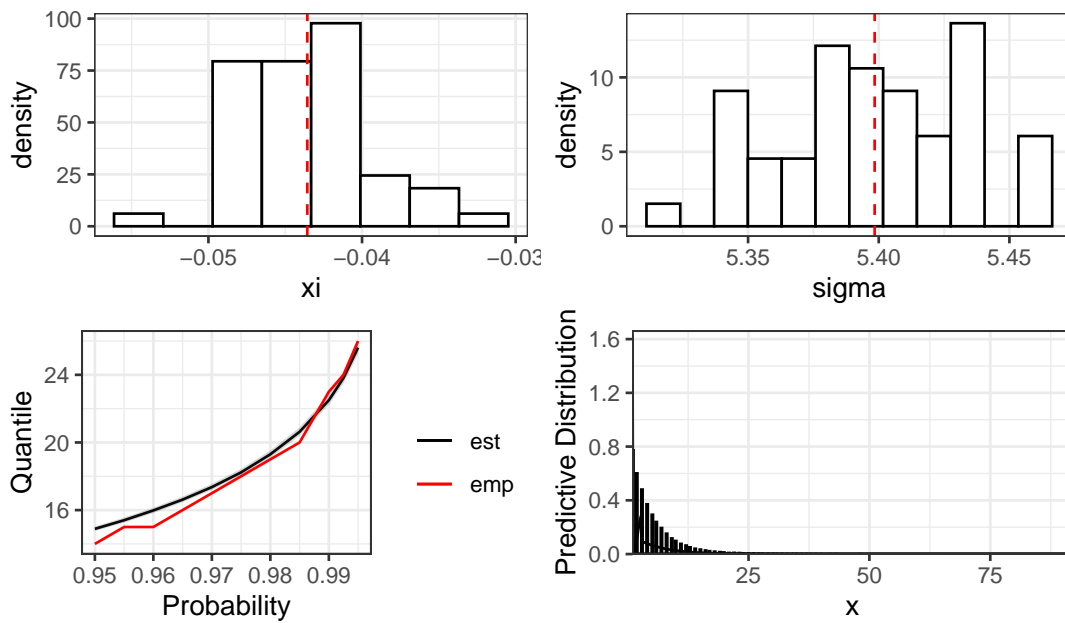


Figura 8: Av. Raul Lopes - Função fmgpd

Considerando a divisão dos níveis das multas, temos que 91,01 % dos autos estão no nível médio, 8,81% estão no nível grave e apenas 0,18% é do tipo gravissima. Neste radar, temos a maior proporção de radares para nível médio e a menor para nível gravissima.

4.5.4 Av. Barão de Castelo Branco

Finalizando a estimação dos parâmetros, temos para o radar da Av. Barão de Castelo Branco, um único de 40 km/h. Para os excessos desse radar, temos que ξ é igual 0, o que indica a utilização de uma função diferente dos modelos anteriores quando ξ foi maior ou menor que 0. Para σ temos o valor de 6,20 e o limiar está localizado na medida 3,01, onde 52,74 % estão acima no limiar.

Tabela 17: Parâmetros para Av. Barão de C. Branco

	Estimado	L.I.	L.S.
ξ	0.00	-0.01	0.01
σ	6.20	6.09	6.31
μ	3.01	3.01	3.02
μ_1	1.02	1.00	1.09
μ_2	3.62	3.60	3.67

	Estimado	L.I.	L.S.
η_1	13.00	9.98	14.20
η_2	8.85	8.58	9.14
W_1	0.20	0.20	0.20
W_2	0.80	0.80	0.80

Graficamente, visualizamos que o estimador ξ não é 0, mas um valor um pouco maior que 0. Utilizando a função `fitgpd` do pacote POT, conseguimos um valor semelhante a 0,006325. Com essa medida, temos que não existe um excesso máximo para a Av. Barão C. Branco. Para a distribuição de σ , observamos uma variação que quase centraliza na média. Para o quantil de superior a 95%, a linha dos estimados foi a que mais se aproximou da linha empírica.

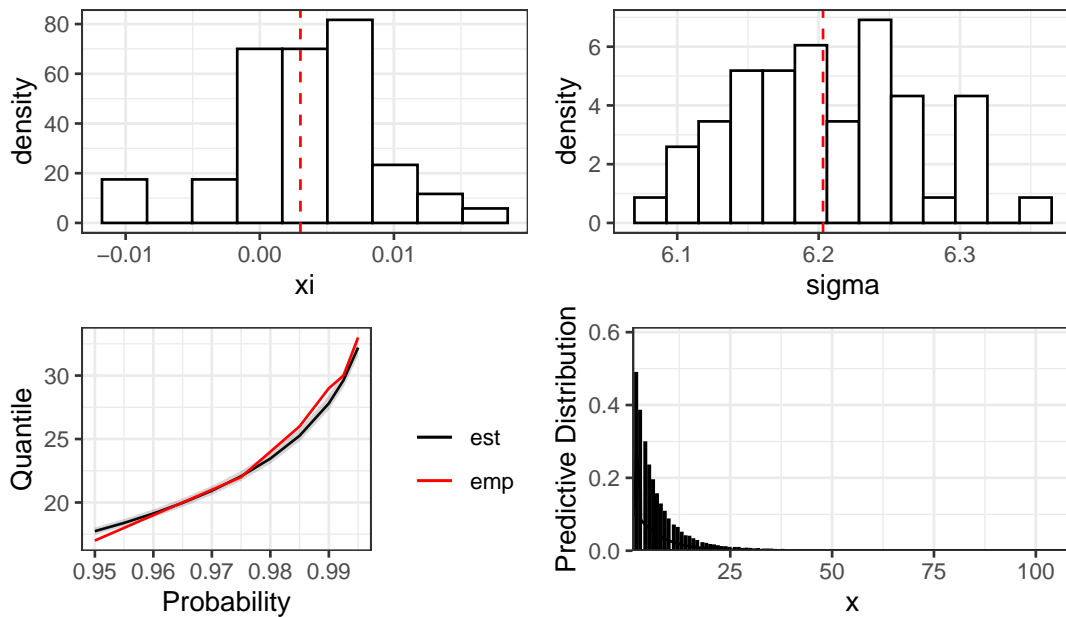


Figura 9: Av. Barão - Função `fmgpd`

Para os intervalos das infrações, temos 87,46% no nível gravíssima, 11,83% para nível médio e 0,70% para o nível gravíssima.

5 Estimação da arrecadação pela distribuição Poisson Composta

Considerando a estimação do número de autuções diárias pela distribuição de Poisson (Equação 5) e a estimação do valor do excesso pela distribuição de mistura por GPD (Equação 16), apresentamos a estimação do valor esperado de arrecadação diária, mensal e anual, considerando os valores de cada atuação e de cada faixa de multa. Será utilizado o valor esperado da distribuição poisson composta.

Conclui-se que o radar localizado na Av. Raul Lopes possui o maior potencial de arrecadação diário, com um valor possível de R\$ 6379,78. Em segundo lugar, temos Alameda Parnaíba com R\$ 3727,62. Av. Maranhão e Av. Barão de C. Branco possuem valores que ficam mais próximos da faixa de R\$ 2000, com R\$ 2114,53 e R\$ 211,53, respectivamente.

Na tabela Tabela 19 vemos que a Av. Raul Lopes se destaca com os valores mais altos e a única superando os R\$ 6000, Alameda Parnaíba passa do R\$ 3000, numa posição intermediária entre os extremos. Os dados apresentados são apresentados considerando um intervalo de confiança de 95% e está construída a partir da função Gamma utilizada no cálculo do parâmetro das médias diárias e através da função `fmgpd` na estimação dos excessos.

Tabela 18: Arrecadação diária

Endereço	IC -	IC +	Mediana
Alameda Parnaíba	3704.352	3764.832	3727.48
Av. Raul Lopes	6336.130	6427.172	6383.88
Av. Maranhão	2657.423	2728.497	2693.00
Av. Barão C. Branco	2841.207	2900.338	2875.46

Considerando os resultados obtidos para a arrecadação diária, as medições para as possíveis arrecadações a nível mensal e anual são resultantes da multiplicação por 30 e 365, respectivamente. Para Av. Raul Lopes, temos potenciais de R\$ 191.393,40. Para a Alameda Parnaíba, teremos R\$ 111.828,60, para Av. Maranhão, R\$ 86.253,90. Para o único radar com velocidade máxima de 40km/h, os valores são R\$ 63.425,90.

Tabela 19: Arrecadação Mensal

Endereço	IC -	IC +	Mediana
Alameda Parnaíba	111130.57	112944.98	111824.4
Av. Raul Lopes	190083.90	192815.17	191516.4
Av. Maranhão	79722.68	81854.93	80790.0
Av. Barão C. Branco	85236.23	87010.12	86263.8

Para as os valores em níveis anuais, Av. Raul Lopes é a única que passa da barreiras dos 2 milhões, com o potencial de R\$ 2.328.319,70, Alameda Parnaíba e Av. Barão de C. Branco ultrapassam a barreira do 1 milhão com R\$ 1.360.581,30 e R\$ 1.049.422,40. Av. Maranhão fica restrita a R\$ 771.803,40.

Tabela 20: Arrecadação Anual

Endereço	IC -	IC +	Mediana
Alameda Parnaíba	1352088.7	1374163.9	1360530
Av. Raul Lopes	2312687.5	2345918.0	2330116
Av. Maranhão	969959.2	995901.6	982945
Av. Barão C. Branco	1037040.7	1058623.2	1049543

6 Conclusão

Este trabalho teve como motivação realizar previsões de arrecadação que alguns radares podem disponibilizar ao município de Teresina. Certamente, existem fatores que iram impossibilitar a concretização desses valores devido à falta de pagamentos dos autos em geral, por exemplo.

Entretanto com análises semelhantes a esta e a inclusão de estudos de outras variáveis que influenciam nos pagamentos nas penalidades de trânsito, o poder público pode realizar previsões mais eficientes e organizar as despesas públicas baseados em critérios que buscam a eficiência da arrecadação e dos gastos.

Este trabalho se dividiu em duas partes. Num primeiro momento, realizamos uma análise através da Distribuição Poisson analisando a frequência diária das infrações de trânsito. A segunda parte consistiu em parametrizar a distribuição dos excessos através da teoria dos valores extremos utilizando a função GPD.

Acreditamos que os resultados obtidos são satisfatórios e que podem auxiliar o poder público na sua organização. Ressaltamos, também, que a análise feita pode ser replicada em outras pesquisas, independente da área, ou seja, não fica restrita a análise de radares, mas, qualquer área que possa ser analisada através da Teoria de Valores Extremos relacionando a com a Poisson.

7 Referências Bibliográficas

- [1] Nascimento FF, Gamerman D, Lopes HF (2011) **Regression models for exceedance data via the full likelihood**. Environ Ecol Stat 18:495-512.
- [2] CTB - **Código de Trânsito Brasileiro**. Disponível em *http* :
//www.planalto.gov.br/ccivil_03/leis/l9503compilado.htm
- [3] Beijo, Luiz Alberto e Avelar, Fabricio Goecking (2011). **Distribuição generalizada de valores extremos no estudo de dados climáticos : Uma breve revisão e aplicação.**, Revista da Estatística da UFOP, Minas Gerais, v. 1, p.10-15, jan. 2011.
- [4] Bussab, W. O. e Morettin, P. A. (2010). **Estatística Básica**, Editora Saraiva, 6a. Edição.
- [5] Nascimento, F. F. (2012). **Modelos Probabilísticos Para Dados Extremos: Teoria e Aplicações**.In: II COLOQUIO DE MATEMÁTICA DA REGIÃO NORDESTE, 2012, Teresina, Piauí. Universidade Federal do Piauí, Edufpi.
- [6] Gibbons, JD e Chakraborti. **Nonparametric Statiscal Inferece**. Fourth Edition, Revised and Expanded. New York, 2010.
- [7] CASELLA, G. e BERGER, R.L,. **Inferência Estatística**. Tradução: Solange Aparecida Visconte. Cengage Learning. São Paulo, 2020.
- [8] PINHEIRO, E.C. **Contribuições em inferência e modelagem de valores extremos**. São Paulo, 2014.
- [9] Página de informações do pacote 'extrememix'. *https* :
//github.com/manueleleonelli/extrememix