

**FUNDAÇÃO GETULIO VARGAS  
ESCOLA DE MATEMÁTICA APLICADA  
MESTRADO EM MODELAGEM MATEMÁTICA DA INFORMAÇÃO**

**MÉTODOS PARA DETECÇÃO DE *OUTLIERS* EM SÉRIES DE PREÇOS  
DO ÍNDICE DE PREÇOS AO CONSUMIDOR**

**TAÍSE FERRAZ LYRA**

**Rio de Janeiro  
2014**

**TAÍSE FERRAZ LYRA**

**MÉTODOS PARA DETECÇÃO DE *OUTLIERS* EM SÉRIES DE PREÇOS DO ÍNDICE  
DE PREÇOS AO CONSUMIDOR**

Dissertação apresentada na Escola de Matemática Aplicada da Fundação Getúlio Vargas como requisito parcial para obtenção do Grau de Mestre em Modelagem Matemática da Informação.

Orientador: Dr. Paulo Cezar Pinto Carvalho

Coorientador: Dr. Crysttian Arantes Paixão

**Rio de Janeiro  
2014**

Lyra, Taíse Ferraz

Métodos para detecção de outliers em séries de preços do Índice de Preços ao Consumidor / Taíse Ferraz Lyra. – 2014.  
209 f.

Dissertação (mestrado) – Fundação Getulio Vargas, Escola de Matemática Aplicada.

Orientador: Paulo Cezar Pinto Carvalho.

Coorientador: Crysttian Arantes Paixão.

Inclui bibliografia.

1. Valores estranhos (Estatística). 2. Índices de preços ao consumidor. 3. Índices de preços. I. Cesar, Paulo, 1952- . II. Paixão, Crysttian Arantes. III. Fundação Getulio Vargas. Escola de Matemática Aplicada. IV. Título.

CDD – 519.5

**TAÍSE FERRAZ LYRA**

**MÉTODOS PARA DETECÇÃO DE OUTLIERS EM SÉRIES DE PREÇOS DO ÍNDICE  
DE PREÇOS AO CONSUMIDOR.**

Dissertação apresentada ao Curso de Mestrado em Modelagem Matemática da Informação da Escola de Matemática Aplicada da Fundação Getúlio Vargas para obtenção do grau de Mestre em Modelagem Matemática da Informação.

Data da defesa: 24/02/2014.

**ASSINATURA DOS MEMBROS DA BANCA EXAMINADORA**



---

**Paulo Cezar Pinto Carvalho**  
Orientador (a)



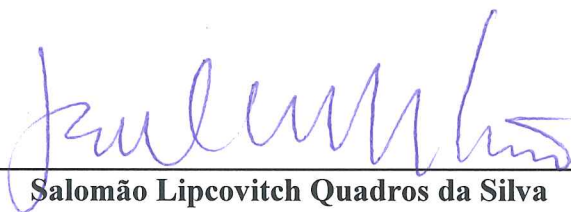
---

**Crystian Arantes Paixão**  
Co-Orientador (a)



---

**Moacyr Alvim Horta Barbosa da Silva**



---

**Salomão Lipcovitch Quadros da Silva**



---

**Sheila Zani**

*A Deus,  
à minha família,  
aos meus amigos,  
aos meus colegas de trabalho,  
e aos profissionais da área,  
ofereço.*

*Ao meu pai, Heribaldo; À minha mãe, Vitoria; Às minhas irmãs, Isabel e Talita; ao meu cunhado, Raphael; À minha madrinha, Eugenia; Ao meu Tio Jaime; ao meu amor, Rodrigo; às minhas queridas amigas Manuella, Perla e Selma e ao meu saudoso primo Rômulo,  
**dedico.***

## AGRADECIMENTOS

A Deus, por me proteger, amparar e guiar nos momentos mais difíceis.

Ao meu pai, Heribaldo do Nascimento Lyra, dono da maior torcida pelo meu sucesso profissional e maior incentivador dos meus estudos, pelo amor incondicional, por fazer dos meus sonhos os seus sonhos, por estar ao meu lado em todas as minhas escolhas, por viver preocupado comigo e por ser tão meu amigo. Se eu tivesse que dedicar todas as minhas vitórias profissionais a uma única pessoa, esta seria você.

À minha mãe, Vitoria Maria Ferraz Lyra, pelas inúmeras demonstrações de amor e carinho, pela cumplicidade e parceria, pelos mimos, pela proteção e confiança. Obrigada por ser uma mãe inigualável, por ser a melhor de todas as amigas e a conselheira mais sábia.

À minha irmã, Talita Ferraz Lyra, por sempre acreditar e confiar no meu potencial mais até do que eu mesma, e ao meu cunhado, Raphael Ramires Lucena, por ser sempre tão prestativo e bondoso.

À minha irmãzinha do coração, Isabel Cristina de Oliveira Ferraz, que há muito torce para que eu termine logo a dissertação e possa dar atenção a ela. O fato de hoje ter você por perto é uma das minhas maiores alegrias.

À minha madrinha, Eugenia Maria da Silva Ferraz Baptista, por ser a tradução do que é bondade e sinônimo de generosidade, e ao meu padrinho do coração, Jaime Loureiro Nobre Baptista, por toda a ajuda que sempre me deu junto com a minha tia. Vocês são exemplos de pessoas, tenho grande admiração pelos dois.

Ao meu amor, Rodrigo Haus da Silva Bacellar, por todo o apoio que sempre me deu, pela paciência para ouvir meus desabafos, por me compreender, por acreditar em mim e na minha capacidade, por me motivar a alcançar meus objetivos e por me dar forças para seguir em frente e nunca pensar em desistir. Obrigada por ser este companheiro maravilhoso e estar ao meu lado não só nos momentos de alegria, mas, principalmente, nos momentos difíceis da vida. Agradeço também a toda a sua família, que sempre torceu por mim e acreditou na minha vitória.

Às minhas queridas amigas Perla Freire da Silva, Manuella Pinheiro Lopes e Selma Mara Borges Coelho, pelas incontáveis madrugadas de estudo no IBRE, por todos os momentos de descontração e, principalmente, por me permitirem entender o quanto se aprende ao ensinar. Vocês foram fundamentais para que eu chegasse até aqui. Um obrigada especial à Manuella, por todas as explicações que me deu, contribuindo imensamente para a realização deste trabalho, e pela enorme paciência em tirar minhas intermináveis dúvidas.

À minha amiga Joyce de Figueiró Santos, pelos estudos virtuais que vararam madrugadas, por toda força e incentivo prestados durante o mestrado e pelas muitas barras de chocolate que dividimos em sala.

Ao meu colega de turma Diego Ferreira Heredia, por sempre se colocar não só à minha disposição como à de todos para ajudar no que fosse preciso, esquecendo-se, na maioria das vezes, de si mesmo. Muito obrigada pela força e pelo incentivo que você me deu.

À minha amiga Carolina Nascimento Nogueira Lima, pelos inúmeros almoços e por me acalmar nas horas de aflição com palavras de carinho, conforto e incentivo.

À minha amiga Livia dos Santos Perri, que, sem dúvidas, fez toda a diferença nestes últimos meses de puro estresse, tensão e nervosismo. Obrigada por me colocar pra cima, por me incentivar e motivar diariamente, por me mandar energias positivas quando eu não tinha mais forças, por me ouvir e me entender, por ter estado tão presente na minha vida. Já não me imagino mais sem a sua amizade.

À minha caloura preferida, Anna Carolina dos Santos Barros, por me divertir com suas histórias (e espirros), por estar sempre disposta a me ajudar e, principalmente, por se preocupar, se interessar e me perguntar todos os dias como estava o andamento da dissertação.

Aos meus amigos de longa data, Alessandra Mundy Torrero, Caroline Ponce de Moraes, Rafael Serra de Carvalho, Renata Carvalho Lopes de Lima e Roberta Guarany Oberlaender, por todo apoio e torcida demonstrados nesses muitos anos de amizade.

À minha inesquecível e tão querida professora da faculdade, Carla Silva Oliveira, pela enorme consideração em sempre perguntar como estava o mestrado, bem como a dissertação e por sempre manifestar sua torcida por mim.

Ao meu chefe, Salomão Lipcovitch Quadros da Silva, pela oportunidade de cursar o mestrado e por ter sido extremamente compreensivo não só nos momentos em que precisei me ausentar para assistir às aulas como também, e principalmente, nesta reta final, em que precisei me dedicar inteiramente à dissertação. Muito obrigada também por ter aceitado o convite para participar da banca, o que, para mim, é uma grande honra.

Ao meu companheiro de equipe André Furtado Braz, por todas as dicas, sugestões e ajuda durante este período.

Ao meu colega de trabalho Angelo Luiz Rocha Polydoro, por me apresentar o Mestrado em Modelagem Matemática da Informação da Escola de Matemática Aplicada da Fundação Getúlio Vargas.

Ao meu orientador, Paulo Cezar Pinto Carvalho, pela confiança depositada.

Ao meu co-orientador, Crysttian Arantes Paixão, um dos maiores responsáveis, se não o maior, pela realização deste trabalho. Para você, palavras não são suficientes para expressar minha gratidão. Obrigada por toda a sabedoria transmitida, pelo exemplo de profissionalismo dado, pelo comprometimento com a minha dissertação mesmo estando em outro estado, por acreditar e confiar em mim e na minha capacidade, por me incentivar, pela enorme paciência para me acalmar nas minhas crises de insegurança, pelos inúmeros e-mails trocados e pelas intermináveis conversas no g-talk para esclarecer minhas dúvidas. Obrigada por ter se colocado à disposição, do início ao fim, para me ajudar em tudo o que fosse preciso. Eu não teria conseguido sem você.

Aos demais membros da banca examinadora, Moacyr Alvim Horta Barbosa da Silva e Sheila Zani, por aceitarem o convite, dispondo de seu tempo para analisar este trabalho.

À equipe de professores da EMap, por todo o conhecimento transmitido durante o curso.

E, por fim, a todos aqueles que, mesmo indiretamente, contribuíram não só para a realização deste trabalho como também para esta longa caminhada.



“Nunca deixe que lhe digam que não vale  
à pena acreditar no sonho que se tem, ou  
que seus planos nunca vão dar certo, ou  
que você nunca vai ser alguém...quem  
acredita, sempre alcança!”

Renato Russo - Mais Uma Vez

“Só quem toma um sonho como sua forma  
de viver, sabe desvendar o segredo de ser  
feliz...”

Roupa Nova - Nos Bailes da Vida

## RESUMO

*Outliers* são observações que parecem ser inconsistentes com as demais. Também chamadas de valores atípicos, extremos ou aberrantes, estas inconsistências podem ser causadas por mudanças de política ou crises econômicas, ondas inesperadas de frio ou calor, erros de medida ou digitação, entre outras. *Outliers* não são necessariamente valores incorretos, mas, quando provenientes de erros de medida ou digitação, podem distorcer os resultados de uma análise e levar o pesquisador à conclusões equivocadas. O objetivo deste trabalho é estudar e comparar diferentes métodos para detecção de anormalidades em séries de preços do Índice de Preços ao Consumidor (IPC), calculado pelo Instituto Brasileiro de Economia (IBRE) da Fundação Getúlio Vargas (FGV). O IPC mede a variação dos preços de um conjunto fixo de bens e serviços componentes de despesas habituais das famílias com nível de renda situado entre 1 e 33 salários mínimos mensais e é usado principalmente como um índice de referência para avaliação do poder de compra do consumidor. Além do método utilizado atualmente no IBRE pelos analistas de preços, os métodos considerados neste estudo são: variações do Método do IBRE, Método do Boxplot, Método do Boxplot SIQR, Método do Boxplot Ajustado, Método de Cercas Resistentes, Método do Quartil, do Quartil Modificado, Método do Desvio Mediano Absoluto e Algoritmo de Tukey. Tais métodos foram aplicados em dados pertencentes aos municípios Rio de Janeiro e São Paulo. Para que se possa analisar o desempenho de cada método, é necessário conhecer os verdadeiros valores extremos antecipadamente. Portanto, neste trabalho, tal análise foi feita assumindo que os preços descartados ou alterados pelos analistas no processo de crítica são os verdadeiros *outliers*. O Método do IBRE é bastante correlacionado com os preços alterados ou descartados pelos analistas. Sendo assim, a suposição de que os preços alterados ou descartados pelos analistas são os verdadeiros valores extremos pode influenciar os resultados, fazendo com que o mesmo seja favorecido em comparação com os demais métodos. No entanto, desta forma, é possível computar duas medidas através das quais os métodos são avaliados. A primeira é a porcentagem de acerto do método, que informa a proporção de verdadeiros *outliers* detectados. A segunda é o número de falsos positivos produzidos pelo método, que informa quantos valores precisaram ser sinalizados para um verdadeiro *outlier* ser detectado. Quanto maior for a proporção de acerto gerada pelo método e menor for a quantidade de falsos positivos produzidos pelo mesmo, melhor é o desempenho do método. Sendo assim, foi possível construir um ranking referente ao desempenho dos métodos, identificando o melhor dentre os analisados. Para o município do Rio de Janeiro, algumas das variações do Método do IBRE apresentaram desempenhos iguais ou superiores ao do método original. Já para o município de São Paulo, o Método do IBRE apresentou o melhor desempenho. Em trabalhos futuros, espera-se testar os métodos em dados obtidos por simulação ou que constituam bases largamente utilizadas na literatura, de forma que a suposição de que os preços descartados ou alterados pelos analistas no processo de crítica são os verdadeiros *outliers* não interfira nos resultados.

Palavras-chave: intervalo de tolerância, proporção de acerto, falsos positivos.

## ABSTRACT

Outliers are observations that appear to be inconsistent with the others. Also called atypical, extreme or aberrant values, these inconsistencies can be caused, for instance, by political changes or economic crises, unexpected cold or heat waves, and measurement or typing errors. Although outliers are not necessarily incorrect values, they can distort the results of an analysis and lead researchers to erroneous conclusions if they are related to measurement or typing errors. The objective of this research is to study and compare different methods for detecting abnormalities in the price series from the Consumer Price Index (Índice de Preços ao Consumidor - IPC), calculated by the Brazilian Institute of Economy (Instituto Brasileiro de Economia - IBRE) from Getulio Vargas Foundation (Fundação Getulio Vargas - FGV). The IPC measures the price variation of a fixed set of goods and services, which are part of customary expenses for families with income levels between 1 and 33 monthly minimum wages and is mainly used as an indice of reference to evaluate the purchasing power of consumer. In addition to the method currently used by price analysts in IBRE, the study also considered variations of the IBRE Method, the Boxplot Method, the SIQR Boxplot Method, the Adjusted Boxplot Method, the Resistant Fences Method, the Quartile Method, the Modified Quartile Method, the Median Absolute Deviation Method and the Tukey Algorithm. These methods were applied to data of the municipalities Rio de Janeiro and São Paulo. In order to analyze the performance of each method, it is necessary to know the real extreme values in advance. Therefore, in this study, it was assumed that prices which were discarded or changed by analysts in the critical process were the real outliers. The method from IBRE is correlated with altered or discarded prices by analysts. Thus, the assumption that the changed or discarded prices by the analysts are the real outliers can influence the results, causing the method from IBRE be favored compared to other methods. However, thus, it is possible to compute two measurements by which the methods are evaluated. The first is the method's accuracy score, which displays the proportion of detected real outliers. The second is the number of false-positive produced by the method, that tells how many values needed to be flagged to detect a real outlier. As higher the hit rate generated by the method and as the lower the amount of false positives produced therefrom, the better the performance of the method. Therefore, it was possible to construct a ranking relative to the performance of the methods, identifying the best among those analyzed. In the municipality of Rio de Janeiro, some of the variations of the method from IBRE showed equal or superior to the original method performances. As for the city of São Paulo, the method from IBRE showed the best performance. It is argued that a method correctly detects an outlier when it signals a real outlier as an extreme value. The method with the highest accuracy score and with smaller number of false-positive was from IBRE. For future investigations, we hope to test the methods in data obtained from simulation and from widely used data bases, so that the assumption related to the discarded or changed prices, during the critical process, does not alter the results.

Keywords: tolerance interval, hit rate, false positives.

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>18</b>
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA . . . . .</b>	<b>20</b>
<b>3</b>	<b>DESCRIÇÃO DOS DADOS . . . . .</b>	<b>24</b>
<b>3.1</b>	<b>Índices de Preços . . . . .</b>	<b>24</b>
<b>3.1.1</b>	<b>Índice de Preços ao Produtor Amplo (IPA) . . . . .</b>	<b>26</b>
<b>3.1.2</b>	<b>Índice Nacional de Custo da Construção (INCC) . . . . .</b>	<b>27</b>
<b>3.1.3</b>	<b>Índice de Preços ao consumidor (IPC) . . . . .</b>	<b>27</b>
<b>3.1.3.1</b>	<b>Estrutura da Amostra de Bens e Serviços . . . . .</b>	<b>28</b>
<b>3.1.3.2</b>	<b>Sistema de Pesos . . . . .</b>	<b>29</b>
<b>3.1.3.3</b>	<b>Sistema de Preços . . . . .</b>	<b>30</b>
<b>3.1.3.4</b>	<b>Sistema de Cálculo . . . . .</b>	<b>30</b>
<b>3.2</b>	<b>Base de Dados . . . . .</b>	<b>33</b>
<b>4</b>	<b>METODOLOGIA . . . . .</b>	<b>39</b>
<b>4.1</b>	<b>Método do IBRE . . . . .</b>	<b>44</b>
<b>4.2</b>	<b>Variações do Método do IBRE . . . . .</b>	<b>45</b>
<b>4.3</b>	<b>Método do Boxplot . . . . .</b>	<b>45</b>
<b>4.4</b>	<b>Método do Boxplot SIQR . . . . .</b>	<b>46</b>
<b>4.5</b>	<b>Método do Boxplot Ajustado . . . . .</b>	<b>46</b>
<b>4.6</b>	<b>Método de Cercas Resistentes . . . . .</b>	<b>48</b>
<b>4.7</b>	<b>Método do Quartil . . . . .</b>	<b>48</b>
<b>4.8</b>	<b>Método do Quartil Modificado . . . . .</b>	<b>48</b>
<b>4.9</b>	<b>Método do Desvio Mediano Absoluto . . . . .</b>	<b>49</b>
<b>4.10</b>	<b>Algoritmo de Tukey . . . . .</b>	<b>49</b>
<b>5</b>	<b>RESULTADOS . . . . .</b>	<b>50</b>
<b>5.1</b>	<b>Método do IBRE e Variações . . . . .</b>	<b>52</b>

5.1.1	Método do IBRE e Variações para o Rio de Janeiro . . . . .	52
5.1.2	Método do IBRE e Variações para São Paulo . . . . .	54
5.2	Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado . . . . .	56
5.2.1	Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado para o Rio de Janeiro . . . . .	57
5.2.2	Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado para São Paulo .	60
5.3	Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e do Algoritmo de Tukey . . . . .	64
5.3.1	Métodos do Quartil, do Quartil Modificado, de Cercas Resistentes, do Desvio Mediano Absoluto e do Algoritmo de Tukey para o Rio de Janeiro . . . . .	65
5.3.2	Métodos do Quartil, do Quartil Modificado, de Cercas Resistentes, do Desvio Mediano Absoluto e do Algoritmo de Tukey para São Paulo . . . . .	70
5.4	Resultado Geral . . . . .	75
5.4.1	Resultado Geral para o Rio de Janeiro . . . . .	75
5.4.2	Resultado Geral para São Paulo . . . . .	76
6	CONCLUSÕES . . . . .	77
	REFERÊNCIAS . . . . .	80
	ANEXO . . . . .	83
	APÊNDICE A . . . . .	95
	APÊNDICE B . . . . .	96
	APÊNDICE C . . . . .	160

## Lista de Figuras

1	Períodos de Coleta dos Preços para o Cálculo dos IGPs . . . . .	25
2	Períodos de Coleta dos Preços para o Cálculo das Versões do IPC-S . . . . .	28
3	Recorte da Estrutura do Índice de Preços do Consumidor (IPC) . . . . .	36
4	Tipos de Críticas Realizadas no IPC do IBRE . . . . .	45
5	Proporção de Acerto $\times$ Número de Falsos Positivos para o Método do IBRE e suas Variações no Município do Rio de Janeiro . . . . .	54
6	Proporção de Acerto $\times$ Número de Falsos Positivos para o Método do IBRE e suas Variações no Município de São Paulo . . . . .	56
7	Proporção de Acerto $\times$ Número de Falsos Positivos para os Métodos do Boxplot, Boxplot SIQR e Boxplot Ajustado para o Município do Rio de Janeiro . . . . .	60
8	Proporção de Acerto $\times$ Número de Falsos Positivos para os Métodos do Boxplot, Boxplot SIQR e Boxplot Ajustado para o Município de São Paulo . . . . .	63
9	Desempenho do Método de Cercas Resistentes no Município do Rio de Janeiro. . .	66
10	Desempenho do Método do Quartil no Município do Rio de Janeiro. . . . .	66
11	Desempenho do Método do Quartil Modificado no Município do Rio de Janeiro. .	67
12	Desempenho do Método do Desvio Mediano Absoluto no Município do Rio de Janeiro. . . . .	67
13	Desempenho do Algoritmo de Tukey no Município do Rio de Janeiro. . . . .	68
14	Desempenho dos Métodos de Cercas Resistentes, do Quartil, do Quartil Modifi- cado, do Desvio Mediano Absoluto e do Algoritmo de Tukey no município do Rio de Janeiro quando aplicados ao conjunto de dados TA. . . . .	69
15	Desempenho do Método de Cercas Resistentes no Município de São Paulo. . . .	71
16	Desempenho do Método do Quartil no Município de São Paulo. . . . .	71
17	Desempenho do Método do Quartil Modificado no Município de São Paulo. . . .	72
18	Desempenho do Método do Desvio Mediano Absoluto no Município de São Paulo.	72
19	Desempenho do Algoritmo de Tukey no Município de São Paulo. . . . .	73

20	Desempenho dos Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Aboluto e do Algoritmo de Tukey no município de São Paulo quando aplicados ao conjunto de dados TA. . . . .	74
----	--	----

## Lista de Tabelas

1	Grupos de Despesa da Estrutura de Bens e Serviços do IPC . . . . .	29
2	Ponderações Regionais para o Cálculo do IPC em Nível Nacional . . . . .	32
3	Amostra da Base de Dados - Parte 1 . . . . .	37
4	Amostra da Base de Dados - Parte 2 . . . . .	38
5	Relativos Gerados pela Equação 3 - Parte 1 . . . . .	42
6	Relativos Gerados pela Equação 3 - Parte 2 . . . . .	43
7	Quantidade Total de <i>Outliers</i> Identificados ( <i>Outliers</i> Id.), Número de <i>Outliers</i> Identificados Corretamente ( <i>Outliers</i> Id. Corretamente), Proporção de Acerto (Prop. de Acerto) e Número de Falsos Positivos Gerados pelo Método do IBRE ( $p = 50\%$ ) e suas Variações ( $p = 30\%, 35\%, 40\%, 45\%, 46\%, 47\%, 48\%, 49\%, 49,25\%, 49,5\%$ e $49,75\%$ ) no Município do Rio de Janeiro . . . . .	53
8	Quantidade Total de <i>Outliers</i> Identificados ( <i>Outliers</i> Id.), Número de <i>Outliers</i> Identificados Corretamente ( <i>Outliers</i> Id. Corretamente), Proporção de Acerto (Prop. de Acerto) e Número de Falsos Positivos Gerados pelo Método do IBRE ( $p = 50\%$ ) e suas Variações ( $p = 30\%, 35\%, 40\%, 45\%, 46\%, 47\%, 48\%, 49\%, 49,25\%, 49,5\%$ e $49,75\%$ ) no Município de São Paulo . . . . .	55
9	Quantidade Total de <i>Outliers</i> Identificados ( <i>Outliers</i> Id.), Número de <i>Outliers</i> Identificados Corretamente ( <i>Outliers</i> Id. Corretamente), Proporção de Acerto (Prop. de Acerto) e Número de Falsos Positivos Gerados pelos Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado no Município do Rio de Janeiro . . . . .	57
10	Quantidade Total de <i>Outliers</i> Identificados ( <i>Outliers</i> Id.), Número de <i>Outliers</i> Identificados Corretamente ( <i>Outliers</i> Id. Corretamente), Proporção de Acerto (Prop. de Acerto) e Número de Falsos Positivos Gerados pelo Método do Boxplot quando Aplicado aos Conjuntos de Dados MR, MA, 2MA, TA e MMAA no Município do Rio de Janeiro . . . . .	59

11	Quantidade Total de <i>Outliers</i> Identificados ( <i>Outliers</i> Id.), Número de <i>Outliers</i> Identificados Corretamente ( <i>Outliers</i> Id. Corretamente), Proporção de Acerto (Prop. de Acerto) e Número de Falsos Positivos Gerados pelos Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado no Município de São Paulo . . . . .	61
12	Quantidade Total de <i>Outliers</i> Identificados ( <i>Outliers</i> Id.), Número de <i>Outliers</i> Identificados Corretamente ( <i>Outliers</i> Id. Corretamente), Proporção de Acerto (Prop. de Acerto) e Número de Falsos Positivos Gerados pelo Método do Boxplot quando Aplicado aos Conjuntos de Dados MR, MA, 2MA, TA e MMAA no Município de São Paulo . . . . .	63
13	Resultados da Análise do Desempenho dos Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Aboluto e do Algoritmo de Tukey no Município do Rio de Janeiro quando Aplicados ao Conjunto de Dados TA	70
14	Resultados da Análise do Desempenho dos Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Aboluto e do Algoritmo de Tukey no Município de São Paulo quando Aplicados ao Conjunto de Dados TA . .	75
15	Desempenho do Método de Cercas Resistentes (MR) no Município do Rio de Janeiro	160
16	Desempenho do Método do Quartil (MR) no Município do Rio de Janeiro . . . . .	161
17	Desempenho do Método do Quartil Modificado (MR) no Município do Rio de Janeiro	162
18	Desempenho do Método do Desvio Mediano Absoluto (MR) no Município do Rio de Janeiro . . . . .	163
19	Desempenho do Algoritmo de Tukey (MR) no Município do Rio de Janeiro . . . .	164
20	Desempenho do Método de Cercas Resistentes (MA) no Município do Rio de Janeiro	165
21	Desempenho do Método do Quartil (MA) no Município do Rio de Janeiro . . . . .	166
22	Desempenho do Método do Quartil Modificado (MA) no Município do Rio de Janeiro	167
23	Desempenho do Método do Desvio Mediano Absoluto (MA) no Município do Rio de Janeiro . . . . .	168
24	Desempenho do Algoritmo de Tukey (MA) no Município do Rio de Janeiro . . . .	169
25	Desempenho do Método de Cercas Resistentes (2MA) no Município do Rio de Janeiro . . . . .	170
26	Desempenho do Método do Quartil (2MA) no Município do Rio de Janeiro . . . .	171



27	Desempenho do Método do Quartil Modificado (2MA) no Município do Rio de Janeiro . . . . .	172
28	Desempenho do Método do Desvio Mediano Absoluto (2MA) no Município do Rio de Janeiro . . . . .	173
29	Desempenho do Algoritmo de Tukey (2MA) no Município do Rio de Janeiro . . .	174
30	Desempenho do Método de Cercas Resistentes (MMAA) no Município do Rio de Janeiro . . . . .	175
31	Desempenho do Método do Quartil (MMAA) no Município do Rio de Janeiro . . .	176
32	Desempenho do Método do Quartil Modificado (MMAA) no Município do Rio de Janeiro . . . . .	177
33	Desempenho do Método do Desvio Mediano Absoluto (MMAA) no Município do Rio de Janeiro . . . . .	178
34	Desempenho do Algoritmo de Tukey (MMAA) no Município do Rio de Janeiro . .	179
35	Desempenho do Método de Cercas Resistentes (TA) no Município do Rio de Janeiro	180
36	Desempenho do Método do Quartil (TA) no Município do Rio de Janeiro . . . . .	181
37	Desempenho do Método do Quartil Modificado (TA) no Município do Rio de Janeiro	182
38	Desempenho do Método do Desvio Mediano Absoluto (TA) no Município do Rio de Janeiro . . . . .	183
39	Desempenho do Algoritmo de Tukey (TA) no Município do Rio de Janeiro . . . . .	184
40	Desempenho do Método de Cercas Resistentes (MR) no Município do São Paulo .	185
41	Desempenho do Método do Quartil (MR) no Município do São Paulo . . . . .	186
42	Desempenho do Método do Quartil Modificado (MR) no Município do São Paulo .	187
43	Desempenho do Método do Desvio Mediano Absoluto (MR) no Município do São Paulo . . . . .	188
44	Desempenho do Algoritmo de Tukey (MR) no Município do São Paulo . . . . .	189
45	Desempenho do Método de Cercas Resistentes (MA) no Município do São Paulo .	190
46	Desempenho do Método do Quartil (MA) no Município do São Paulo . . . . .	191
47	Desempenho do Método do Quartil Modificado (MA) no Município do São Paulo .	192

48	Desempenho do Método do Desvio Mediano Absoluto (MA) no Município do São Paulo . . . . .	193
49	Desempenho do Algoritmo de Tukey (MA) no Município do São Paulo . . . . .	194
50	Desempenho do Método de Cercas Resistentes (2MA) no Município do São Paulo	195
51	Desempenho do Método do Quartil (2MA) no Município do São Paulo . . . . .	196
52	Desempenho do Método do Quartil Modificado (2MA) no Município do São Paulo	197
53	Desempenho do Método do Desvio Mediano Absoluto (2MA) no Município do São Paulo . . . . .	198
54	Desempenho do Algoritmo de Tukey (2MA) no Município do São Paulo . . . . .	199
55	Desempenho do Método de Cercas Resistentes (MMAA) no Município do São Paulo	200
56	Desempenho do Método do Quartil (MMAA) no Município do São Paulo . . . . .	201
57	Desempenho do Método do Quartil Modificado (MMAA) no Município do São Paulo	202
58	Desempenho do Método do Desvio Mediano Absoltuo (MMAA) no Município do São Paulo . . . . .	203
59	Desempenho do Algoritmo de Tukey (MMAA) no Município do São Paulo . . . . .	204
60	Desempenho do Método de Cercas Resistentes (TA) no Município do São Paulo . .	205
61	Desempenho do Método do Quartil (TA) no Município do São Paulo . . . . .	206
62	Desempenho do Método do Quartil Modificado (TA) no Município do São Paulo .	207
63	Desempenho do Método do Desvio Mediano Absoltuo (TA) no Município do São Paulo . . . . .	208
64	Desempenho do Algoritmo de Tukey (TA) no Município do São Paulo . . . . .	209

## 1 INTRODUÇÃO

As séries históricas de preços são influenciadas por diversos fatores, como, por exemplo, os *outliers*. *Outliers* são observações que parecem ser inconsistentes com as demais. Também chamadas de valores atípicos, extremos ou aberrantes, estas inconsistências podem ser causadas por mudanças de política ou crises econômicas, ondas inesperadas de frio ou calor, erros de medida ou digitação, entre outras. *Outliers* não são necessariamente valores incorretos, mas, quando provenientes de erros de medida ou digitação, podem distorcer os resultados de uma análise e levar o pesquisador à conclusões equivocadas.

Portanto, uma das etapas mais importantes de uma pesquisa é a crítica dos dados. Nela, todas as informações são analisadas a fim de encontrar possíveis inconsistências. Uma vez identificadas, torna-se factível eliminar erros capazes de gerar resultados que não condizem com a realidade do fenômeno pesquisado. Fundamental para garantir a qualidade dos dados, o processo de crítica é considerado essencial para órgãos produtores de estatísticas, como o Instituto Brasileiro de Economia (IBRE<sup>1</sup>), alcançarem resultados confiáveis.

O IBRE foi criado em 1951 e é a unidade da Fundação Getulio Vargas (FGV<sup>2</sup>) que tem por missão pesquisar, analisar, produzir e disseminar estatísticas macroeconômicas e pesquisas econômicas aplicadas, de alta qualidade, que sejam relevantes para o aperfeiçoamento das políticas públicas ou da ação privada na economia brasileira, estimulando o desenvolvimento econômico e o bem-estar social do país.

Um dos destaques das estatísticas produzidas pelo IBRE é o Índice Geral de Preços (IGP). O IGP é formado por três índices: o Índice de Preços ao Produtor Amplo (IPA), o Índice de Preços ao Consumidor (IPC) e o Índice Nacional de Custos da Construção (INCC).

A base de dados utilizada neste trabalho é um recorte da estrutura de bens e serviços que integram o IPC. Este índice mede a variação dos preços de um conjunto fixo de bens e serviços componentes de despesas habituais das famílias com nível de renda situado entre 1 e 33 salários mínimos mensais. O IPC é usado principalmente como um índice de referência para avaliação do poder de compra do consumidor.

Atualmente, no IBRE, a crítica dos preços coletados é feita com base em parâmetros fixos desenvolvidos pelos analistas de preços. No entanto, esta etapa é a mais demorada do cálculo do índice. Isto porque uma grande quantidade de valores críticos é sinalizada e os analistas gastam muito tempo examinando todas as informações detectadas, de forma a verificar se as mesmas são realmente *outliers*.

Existem diferentes técnicas para a detecção de *outliers*. Neste estudo, além do método utili-

---

<sup>1</sup><http://portalibre.fgv.br/>

<sup>2</sup><http://portal.fgv.br/>

zado atualmente no IBRE, são considerados os seguintes métodos: variações do Método do IBRE, o Método do Boxplot, o Método do Boxplot SIQR, o Método do Boxplot Ajustado, o Método de Cercas Resistentes, o Método do Quartil, o Método do Quartil Modificado, o Método do Desvio Mediano Absoluto e o Algoritmo de Tukey. Todos eles envolvem a construção de um intervalo de tolerância. Valores pertencentes a este intervalo são considerados aceitáveis, enquanto valores menores que o limite inferior ou maiores que o limite superior deste intervalo são considerados *outliers*.

As variações do Método do IBRE consistem em utilizar, também, parâmetros fixos, porém diferentes dos considerados atualmente. Os intervalos de tolerância gerados pelo Método do Boxplot e pelo Método de Cercas Resistentes são baseados no intervalo interquartil, enquanto os gerados pelo Método do Boxplot SIQR, pelo Método do Quartil e pelo Método do Quartil Modificado são baseados nos intervalos semi-interquartílicos. O Método do Boxplot Ajustado leva em consideração uma medida de assimetria denominada *medcouple* (*MC*). O Método do Desvio Mediano Absoluto é fundamentado na mediana calculada a partir dos valores absolutos das diferenças entre a quantidade observada e a mediana das observações. Por fim, o Algoritmo de Tukey é um método de detecção de *outliers* em dados racionais, mais especificamente em dados de preços relativos.

O objetivo deste trabalho é estudar e comparar os métodos citados com a finalidade construir um ranking referente ao desempenho dos métodos, permitindo uma busca pela redução do tempo gasto na identificação de *outliers* por meio de um processo mais eficiente, aumentando, assim, a precisão do IPC. Para compará-los, é preciso conhecer os verdadeiros valores extremos antecipadamente. Por isso, neste trabalho, tal comparação é feita assumindo que os preços descartados ou alterados pelos analistas no processo de crítica são os verdadeiros *outliers*. Desta forma, é possível computar duas medidas através das quais os métodos são avaliados. A primeira é a porcentagem de acerto do método, que informa a proporção de verdadeiros *outliers* detectados. A segunda é o número de falsos positivos produzidos pelo método, que informa quantos valores precisaram ser sinalizados para um verdadeiro *outlier* ser detectado.

Esta dissertação está dividida em seis Seções, contando com a Introdução. Na Seção 2, é feita a revisão bibliográfica do tema abordado neste trabalho. Na Seção 3, é definido o conceito de índice de preços e o IGP da FGV é apresentado, bem como cada um de seus índices componentes. Ainda nesta Seção, é definida a base de dados utilizada neste estudo e a descrição de cada uma das variáveis pertencentes a ela. Na Seção 4, são apresentados os métodos para detecção de *outliers* considerados neste trabalho. Na Seção 5, são discutidos os resultados encontrados e, por fim, na Seção 6, são exibidas as conclusões do estudo desenvolvido e são dadas sugestões para trabalhos futuros.

## 2 REVISÃO BIBLIOGRÁFICA

Na literatura, são encontradas diversas definições para *outliers*, além de diferentes técnicas para detectá-los. Nesta Seção, são apresentadas algumas destas definições bem como alguns textos já publicados que falam sobre os métodos considerados neste estudo e a forma como os mesmos foram abordados e utilizados em trabalhos anteriores.

Segundo Gumbel (1960), *outliers* são valores que parecem ser ou muito grandes ou muito pequenos em comparação com o restante das observações. Para Grubbs (1969), uma observação aberrante, ou *outlier*, é aquela que parece desviar significativamente dos outros membros da amostra na qual ocorre. De acordo com Hawkins (1980), *outlier* é uma observação que desvia muito das outras observações despertando suspeitas de que são geradas por um mecanismo diferente. Conforme Barnett e Lewis (1994), *outlier* é uma observação (ou um subconjunto de observações) que parece ser inconsistente quando comparado ao restante do conjunto de dados. Consoante Tan, Steinbach e Kumar (2006), *outliers* são objetos de dados que tem características diferentes da maioria dos outros objetos num conjunto de dados.

Neste estudo, os Métodos do Boxplot, do Boxplot SIQR, do Boxplot Ajustado, de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e o Algoritmo de Tukey são aplicados além do método utilizado atualmente no IBRE e de suas variações.

O boxplot é uma ferramenta de análise gráfica através da qual são extraídas informações sobre a distribuição dos dados, tais como: posição, dispersão, assimetria, caudas e dados discrepantes. Para construí-lo, é necessário calcular cinco medidas: limite da haste inferior, primeiro quartil, mediana, terceiro quartil e limite da haste superior. A diferença entre os três boxplots considerados neste estudo está na maneira através da qual o limite das hastes inferior e superior é calculado.

A versão mais utilizada do boxplot é a sugerida por Tukey (1977). O intervalo de tolerância deste método é construído por meio do primeiro e do terceiro quartil, de uma constante  $c$  determinada e considerada neste estudo como sendo igual a 1.5 e do intervalo interquartílico. Quando a distribuição dos dados é assimétrica, o Método do Boxplot detecta um grande número de falsos positivos. Portanto, de forma a melhorar a distinção entre observações regulares e aberrantes, algumas modificações foram propostas para este método na literatura.

Kimber (1990) e Aucremanne et al. (2004), por exemplo, propuseram o uso dos intervalos semi-interquartílicos inferior e superior ao invés do intervalo interquartílico, dando origem ao Método do Boxplot SIQR. Este boxplot captura de forma suave a assimetria dos dados, mas a mesma ainda não é ajustada de forma suficiente.

Hubert e Vandervieren (2008) propõem um boxplot ajustado, que possui limites inferior e superior mais robustos e que aderem a todas as distribuições. Para a construção do intervalo de tolerância do Método do Boxplot Ajustado, uma medida de assimetria é levada em consideração,

bem como o primeiro e o terceiro quartil e o intervalo interquartílico.

O Método de Cercas Resistentes, utilizado por Thompson e Sigman (1999) do *United States Census Bureau (USBC*<sup>3</sup>), utiliza o primeiro e o terceiro quartil, o intervalo interquartílico e constantes predeterminadas  $c_L$  e  $c_U$  no cálculo dos limites inferior e superior de seu intervalo de tolerância, respectivamente.

Já o cálculo dos limites inferior e superior do intervalo de tolerância do Método do Quartil, que, segundo Saïdi e Rubin-Bleuer (2005) do *Statistics Canada*<sup>4</sup>, tem um desempenho tão bom quanto o Método de Cercas Resistentes e o Algoritmo de Tukey e é indicado por Rais (2007) como o melhor, é baseado na mediana, nos intervalos semi-interquartílicos inferior e superior e nas constantes predeterminadas  $c_L$ ,  $c_U$  e  $a$ . O Método do Quartil Modificado também utiliza a mediana, os intervalos semi-interquartílicos inferior e superior e as constantes predeterminadas  $c_L$ ,  $c_U$  e  $a$  na construção de seu intervalo de tolerância, porém, de uma maneira ligeiramente diferente.

O Algoritmo de Tukey é o método para detecção de *outliers* utilizado pelo *Office for National Statistics (ONS*<sup>5</sup>). Para o cálculo dos limites inferior e superior do intervalo de tolerância deste método, não só a estabilidade dos dados é removida como também parte das observações mais extremas é excluída. Além disso, são utilizadas constantes predeterminadas  $c_L$  e  $c_U$ .

O intervalo de tolerância do Método do Desvio Mediano Absoluto é calculado por meio da mediana, de constantes predeterminadas  $c_L$  e  $c_U$  e do desvio mediano absoluto, redescoberto e popularizado por Hampel (1974).

Em Hubert e Vandervieren (2008), os Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado são aplicados a diferentes conjuntos de dados: intervalos, medidos em dias, entre explosões ocorridas em minas de carvão; valor do pH e o teor de cálcio em amostras de solo coletadas em diferentes comunidades da região de Condroz, na Bélgica; a velocidade do vento, medida em milhas por hora, por 111 dias consecutivos; tempo de permanência de 201 pacientes que estiveram no Hospital Universitário de Lausanne (*University Hospital of Lausanne*), em 2000; e dados da pesquisa de orçamentos familiares - *Consumer Expenditure Survey (CES)*, realizada em 1995, pelo *Bureau of Labor Statistics, U.S. Department of Labor*. Foi feita também uma simulação, na qual os métodos foram aplicados em 100 amostras de tamanho 1.000 das distribuições *Qui-Quadrado* ( $\chi^2$ ), *Gamma* ( $\Gamma$ ), *Pareto*, *F de Snedecor* ( $F$ ) e distribuição  $G_g$  (HOAGLIN; MOSTELLER; TUKEY, 1995). Os resultados da aplicação dos métodos em dados reais e simulados indicam que o uso do Método do Boxplot Ajustado para distribuições assimétricas permite uma melhor distinção entre observações regulares e *outliers*, o que faz do boxplot ajustado uma ferramenta rápida e automática para a detecção de *outliers* sem que qualquer suposição sobre a distribuição dos dados seja feita.

<sup>3</sup><http://www.census.gov/> em 23 de janeiro de 2014

<sup>4</sup><http://www.statcan.gc.ca/> em 23 de janeiro de 2014

<sup>5</sup><http://www.ons.gov.uk> em 23 de janeiro de 2014

Na aplicação dos métodos do Quartil, do Quartil Modificado, de Cercas Resistentes e do Desvio Mediano Absoluto, para considerar a constante  $c_L = c_U$ , é necessário que a distribuição dos dados seja simétrica. Já o Algoritmo de Tukey não requer que os dados sejam simétricos.

De forma a obter uma distribuição simétrica ou, no mínimo, menos enviesada, os dados são transformados por meio de uma função apropriada. Segundo Hutton (2008), a função logaritmo natural e as funções de raiz quadrada são amplamente utilizadas por terem propriedades desejáveis com relação à variação. Ainda segundo Hutton (2008), o Método de BOX-COX (BOX; COX, 1964) está disponível nos casos de grandes coeficientes de assimetria. Outra transformação utilizada é a de Hidiroglou-Berthelot (*HB*), introduzida por Hidiroglou e Berthelot (1986) e específica para razões.

Em Rais (2008), o Método de Cercas Resistentes, o Método do Quartil e o Algoritmo de Tukey são aplicados a microdados mensais do *Consumer Price Index (CPI)* do *Statistics Canada*. O estudo foi limitado para preços referentes a oito itens específicos, representativos em termos de peso na cesta do *CPI* e em termos de diferentes graus de variabilidade. As informações pertencem ao período compreendido entre julho de 2007 e janeiro de 2008.

Ainda de acordo com Rais (2008), o método para detecção de *outliers* utilizado atualmente pelo *Statistics Canada* envolve o uso de intervalos de tolerância cujos limites inferiores e superiores são definidos por especialistas no assunto, que usam suas experiências e seus conhecimentos da indústria para tal. Os limites destes intervalos são específicos para cada item ou conjunto de itens relacionados e, normalmente, permanecem estáticos durante longos períodos de tempo. Este método está sendo analisado para determinar se alterações podem ser feitas para melhorar a eficiência do sistema e a qualidade das estimativas.

Portanto, segundo Rais (2008), o Método de Cercas Resistentes, o Método do Quartil e o Algoritmo de Tukey são testados com diferentes valores para a constante  $c$ . Desta maneira, o estudo serviu não só para ajudar a selecionar o método mais apropriado para os dados do *CPI*, mas também ajudou a identificar os valores de  $c$  que permitiram que cada método apresentasse um bom desempenho. O estudo foi feito assumindo que os *outliers* detectados pelo método atual do *Statistics Canada* são os verdadeiros *outliers*. Os resultados, analisados por meio da proporção de acerto e da quantidade de falsos positivos geradas, mostraram que o Método do Quartil aplicados a dados transformados produziu a melhor proporção de acerto. Rais (2008) recomenda o uso da transformação *HB*, uma vez que a transformação logarítmica é mais robusta para distribuições fortemente viesadas, embora a performance de ambas tenham sido bem parecidas. Além disso, os resultados mostraram que os valores de  $c$  devem ser proporcionais à variabilidade dos dados e tais valores podem ser atualizados de dois em dois anos usando dados recentes.

Em Hutton (2008), o foco é a construção de um sistema específico para pesquisas de índices de preços. Tal sistema é composto por dois ambientes: um de teste e outro de produção. No ambiente de teste, a função através da qual os dados são transformados é selecionada dentre a de

logaritmo natural, as de raiz quadrada e a transformação *HB*. Em seguida, o melhor método para detecção de *outliers* é eleito dentre o Método do Quartil, o Algoritmo de Tukey, o Método de Cercas Resistentes e o Método do Desvio Mediano Absoluto. Os métodos são aplicados utilizando valores *default* para  $c_L$ ,  $c_U$  e  $a$ , mas o usuário tem a opção alterá-los para outros de sua preferência. Uma vez detectados, o usuário tem a possibilidade corrigir os valores atípicos identificados por meio dos métodos de tratamento mencionados no trabalho em questão. Por fim, os resultados são exibidos em duas tabelas: na primeira, estão os dados que foram considerados como *outliers* pelos quatro métodos citados e, na segunda, estão os resultados da pesquisa antes e depois do tratamento das informações. Já no ambiente de produção, espera-se que o usuário já saiba qual a transformação e qual o método para detecção de *outliers* mais apropriados para seus dados. Neste ambiente, é preciso que o usuário analise todas as informações consideradas extremas pelo método escolhido, de forma a determinar se aqueles valores são realmente aberrantes e devem ser tratados. O tratamento dos *outliers* é aplicado apenas nas observações indicadas pelo usuário, que podem ser alteradas manualmente pelo mesmo. Depois, os resultados da pesquisa são exibidos antes e após o tratamento das informações.

As fórmulas para o cálculo dos limites inferior e superior do intervalo de tolerância gerado por cada um dos métodos são apresentadas na Seção 4.



### 3 DESCRIÇÃO DOS DADOS

Esta Seção está dividida em duas partes: Índices de Preços e Base de Dados. Na primeira parte, é dada a definição de índices de preços e é apresentado o Índice Geral de Preços (IGP) da Fundação Getúlio Vargas (FGV), bem como cada um de seus índices componentes. A apresentação é feita mediante descrição de características tais como: principais usos, abrangência geográfica e setorial, período de coleta dos preços, periodicidade de cálculo e divulgação dos resultados. A base de dados utilizada neste estudo é referente ao Índice de Preços ao Consumidor (IPC) e, portanto, sobre ele, são dados maiores detalhes tais como: a origem da estrutura da amostra de bens e serviços e o funcionamento do sistema de pesos, do sistema de preços e do sistema de cálculo (em nível municipal e nacional). Na segunda parte, é definido o recorte feito na estrutura do IPC e é dada a descrição de cada uma das variáveis pertencentes ao arquivo de dados.

#### 3.1 Índices de Preços

De acordo com Diewert (2001), o número de bens fisicamente distintos e de tipos de serviços que os consumidores podem comprar está na casa dos milhões. Para diversos fins, é necessário resumir esta vasta quantidade de informações sobre preço e quantidade em um conjunto de números bem menor. Isto pode ser feito por meio da construção de um índice de preços. Segundo Diewert (2001), um índice de preços é uma medida ou função que sumariza a mudança nos preços de diversas *commodities* de uma situação 0 (período de tempo ou local) para outra situação 1. Mais especificamente, na prática, um índice de preços pode ser considerado como uma média ponderada da variação nos preços relativos das *commodities* em consideração nas duas situações. Para definir um índice de preços, é necessário conhecer:

- Quais *commodities* ou itens a serem incluídos no índice;
- Como determinar os preços dos itens;
- Quais transações que envolvem estes itens a serem consideradas no índice;
- Como determinar os pesos e de qual pesquisa estes pesos devem ser retirados;
- Qual fórmula ou qual tipo de média deve ser utilizado.

Os índices de preços mais simples são indicadores que visam medir a evolução do nível de preços, mensurando a variação média dos preços de uma cesta fixa de produtos, em um determinado período de tempo.

O Índice Geral de Preços da Fundação Getúlio Vargas é calculado pelo Instituto Brasileiro de Economia (IBRE) e registra as variações de preços de matérias-primas agropecuárias e industriais, de produtos intermediários e de bens e serviços finais, cobrindo, assim, grande parte do processo produtivo. É usado principalmente como indicador macroeconômico, deflator de valores monetários e indexador de contratos. Possui abrangência geográfica nacional e engloba setores da indústria, construção civil, agricultura, comércio varejista e serviços prestados às famílias. Tem periodicidade mensal e está disponível em três versões: Índice Geral de Preços - 10 (IGP-10), Índice Geral de Preços de Mercado (IGP-M) e Índice Geral de Preços - Disponibilidade Interna (IGP-DI). Por contar com três versões, o IGP é, muitas vezes, tratado no plural (IGPs). Estas três versões diferem entre si apenas no período de coleta dos preços. Para o IGP-10, os preços são coletados no período compreendido entre os dias 11 do mês anterior e 10 do mês de referência, para o IGP-M, no período compreendido entre os dias 21 do mês anterior e 20 do mês de referência e, para o IGP-DI, no período compreendido entre o primeiro e o último dia útil do mês de referência. Os períodos de coleta dos preços de cada uma das versões citadas estão representados na Figura 1.

Períodos de Coleta de Preços																							
Mês Anterior												Mês de Referência											
11										21		01								10			
IGP - 10																							
												IGP - M											
												IGP - DI											

Figura 1 Períodos de Coleta dos Preços para o Cálculo dos IGPs

Os IGPs permitem um acompanhamento decendial da inflação. O IGP-M é a única versão que possui duas prévias divulgadas antes do fechamento mensal e que apresentam resultados parciais do índice com base na coleta realizada em períodos decendiais. A primeira prévia (IGP-M1) calcula as variações obtidas por meio da comparação dos preços coletados no período de 21 a 31 do mês anterior ao de referência com os coletados ao longo dos trinta dias anteriores. A segunda prévia (IGP-M2) expande o período de coleta dos preços para 21 do mês anterior a 10 do mês de referência. A próxima apuração é o próprio IGP-M.

O IGP é formado por outros três índices: o Índice de Preços ao Produtor Amplo (IPA), o Índice de Preços ao Consumidor (IPC) e o Índice Nacional de Custos da Construção (INCC). O IPA representa 60% do IGP, o IPC representa 30% e os 10% complementares são representados pelo INCC.

Em termos algébricos, o IGP, para um mês  $t$  qualquer, resulta da Equação 1, que é a mesma para todas as versões do índice, considerando, é claro, as versões do IPA, IPC e INCC correspondentes:

$$I_t = 0.6X_t + 0.3Y_t + 0.1Z_t \quad (1)$$

onde  $I_t$ ,  $X_t$ ,  $Y_t$  e  $Z_t$  são, respectivamente, os números índices do IGP, IPA, IPC e INCC, no mês de referência  $t$ .

Tradicionalmente, estes pesos correspondem a parcelas da despesa interna bruta, calculadas com base nas Contas Nacionais, do Instituto Brasileiro de Geografia e Estatística (IBGE)<sup>6</sup>. As Contas Nacionais apresentam informações sobre a geração, distribuição e uso da renda no país, além de dados sobre a acumulação de ativos não financeiros e sobre as relações entre a economia nacional e o resto do mundo. Os 60% representados pelo IPA são referentes ao valor adicionado pela produção de bens agropecuários e industriais nas transações comerciais em nível de produtor. Os 30% de participação do IPC correspondem ao valor adicionado pelo setor varejista e pelos serviços de consumo. Finalmente, os 10% representados pelo INCC equivalem ao valor adicionado pela indústria da construção civil.

Os resultados mensais do IGP são publicados às 8 horas dos dias estabelecidos no calendário de disseminação de índices da FGV, no Portal do IBRE, por meio de um comunicado divulgado. Os resultados gerais podem ser obtidos gratuitamente. Já a consulta detalhada à maior parte das séries históricas é feita mediante contrato, estando restrita aos assinantes do IBRE/FGVDados, que é o serviço de banco de dados da produção estatística do IBRE e também de um conjunto selecionado de indicadores de outras entidades, como índices de preços, cotações de moedas, taxas de juros e índices de ações.

A seguir, são descritas as características de cada um dos índices componentes do IGP. Para maiores detalhes, ver Instituto Brasileiro de Economia (2013).

### 3.1.1 Índice de Preços ao Produtor Amplo (IPA)

O IPA tem por objetivo medir as variações médias de preços recebidos pelos produtores domésticos na primeira venda de seus produtos. É um indicador econômico de abrangência nacional, de periodicidade mensal e que engloba os setores agropecuário e industrial. O IPA é apurado com base em pesquisa sistemática de preços nas principais regiões de produção do país e está disponível nas mesmas versões do IGP: IPA-10, IPA-M e IPA-DI. O IPA é o principal componente do IGP, representando 60% do mesmo, e é apresentado em duas diferentes estruturas de classificação de seus itens componentes: Origem e Estágios de Processamento. A primeira contém os produtos agropecuários e industriais e a segunda os bens finais, bens intermediários e matérias primas brutas. Os resultados mensais do IPA são divulgados da mesma forma que os do IGP.

<sup>6</sup><http://www.ibge.gov.br/> em 16 de dezembro de 2013

### **3.1.2 Índice Nacional de Custo da Construção (INCC)**

O INCC mede a evolução dos custos de construções habitacionais. É muito usado na correção de contratos de financiamentos habitacionais. Abrange geograficamente as seguintes capitais: Belo Horizonte, Brasília, Porto Alegre, Recife, Salvador, Rio de Janeiro e São Paulo. Calcula-se um índice para cada uma delas e o índice nacional resulta da média aritmética ponderada dos sete índices municipais. O INCC tem periodicidade mensal e engloba os setores de materiais e equipamentos, serviços e mão-de-obra. Está disponível nas mesmas versões que o IGP e o IPA: INCC-10, INCC-M e INCC-DI. Os resultados mensais do INCC são divulgados da mesma forma que os do IGP e os do IPA.

### **3.1.3 Índice de Preços ao consumidor (IPC)**

O IPC mede a variação dos preços de um conjunto fixo de bens e serviços componentes de despesas de consumo habituais das famílias com nível de renda situado entre 1 e 33 salários mínimos mensais. É usado principalmente como um índice de referência para avaliação do poder de compra do consumidor.

Além das versões IPC-10, IPC-M e IPC-DI, o IPC também se encontram disponíveis o Monitor da Inflação, o Índice de Preços ao Consumidor Semanal (IPC-S), o Índice de Preços ao Consumidor da Terceira Idade (IPC-3i) e o Índice de Preços ao Consumidor - Baixa Renda (IPC-C1).

O Monitor da Inflação é direcionado principalmente a analistas de mercado e apresenta as variações de preços ao consumidor de uma cesta abrangente de produtos e serviços, tornando-se um importante aliado no acompanhamento, na análise e na previsão dos índices oficiais de inflação do país. O indicador apresenta as variações diárias dos preços dos itens pesquisados nas sete capitais cobertas pelo IPC já mencionadas anteriormente, em duas bases de comparação: últimos trinta dias em relação a igual período anterior, e últimos sete dias em relação a mesmo período do mês anterior.

O IPC-S é um índice quadrissemanal composto pelas seguintes versões: IPC-S1 (preços coletados entre os dias 8 do mês anterior e 7 do mês de referência), IPC-S2 (preços coletados entre os dias 16 do mês anterior e 15 do mês de referência), IPC-S3 (preços coletados entre os dias 23 do mês anterior e 22 do mês de referência) e IPC-S4 (preços coletados entre o primeiro e o último dia útil do mês de referência), sendo esta última equivalente ao IPC-DI. Os períodos de coleta dos preços de cada uma das versões do IPC-S citadas estão representados na Figura 2.

O IPC-S permite detectar com agilidade as mudanças de curso na trajetória dos preços, pois o intervalo entre o fim de sua coleta e divulgação é de um dia, sendo um dos mais curtos, inclusive

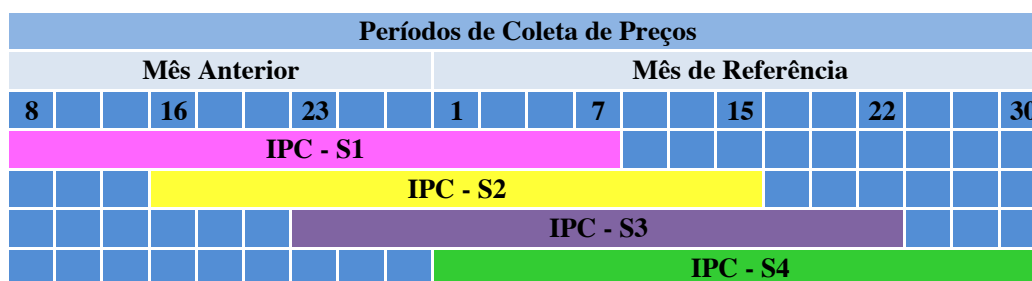


Figura 2 Períodos de Coleta dos Preços para o Cálculo das Versões do IPC-S

para padrões internacionais.

O IPC-3i mede a evolução do custo de vida para famílias compostas, majoritariamente, por indivíduos com mais de 60 anos de idade e serve como referência para a execução de políticas públicas nas áreas de saúde e previdência.

O IPC-C1 mede a variação dos preços de uma cesta de produtos e serviços para as famílias com renda entre 1 e 2,5 salários mínimos mensais e detecta o afastamento entre a taxa de inflação medida para o conjunto da população e a referente às famílias com menor poder aquisitivo, dado que a inflação não é a mesma para todas elas, pois cada uma tem seu próprio padrão de consumo.

A coleta dos preços para o IPC-3i e para o IPC-C1 é semelhante à do IPC-DI: ocorre do primeiro ao último dia útil do mês de referência. Todas as versões do IPC abrangem geograficamente as capitais Belo Horizonte, Brasília, Porto Alegre, Recife, Salvador, Rio de Janeiro e São Paulo e tem periodicidade mensal, com exceção do IPC-S, que tem periodicidade quadrissemanal.

O índice nacional de cada uma das versões resulta da média aritmética ponderada dos índices municipais. Os resultados mensais de todas as versões do IPC são divulgados da mesma forma que os do IGP, IPA e INCC, sendo que a divulgação do IPC-3i é trimestral.

Para a obtenção dos resultados do IPC, combina-se: uma amostra de bens e serviços (de onde foram selecionados e como estão distribuídos os bens e serviços que fazem parte do IPC), um sistema de pesos (como são obtidos os pesos de cada um dos bens e serviços integrantes do IPC), um sistema de preços (como é feita a coleta de preços dos bens e serviços que compõem o IPC) e um sistema de cálculo (como é feito o cálculo do IPC).

### 3.1.3.1 Estrutura da Amostra de Bens e Serviços

A cesta de bens e serviços integrantes do IPC foi selecionada da Pesquisa de Orçamentos Familiares (POF), realizada pelo IBGE nos anos de 2008 e 2009. A POF visa mensurar as estruturas de consumo, dos gastos e dos rendimentos das famílias, além de permitir traçar um perfil das condições de vida da população brasileira a partir da análise de seus orçamentos domésticos.

Para as versões IPC-S, IPC-10, IPC-M e IPC-DI, é considerada a estrutura média das despesas de consumo das famílias residentes nos sete municípios pesquisados já mencionados anteriormente, com renda mensal entre 1 e 33 salários mínimos.

Para o IPC-3i, é considerada a estrutura média das despesas de consumo das famílias compostas, majoritariamente, por indivíduos com mais de 60 anos de idade, residentes nos sete municípios pesquisados, com renda mensal entre 1 e 33 salários mínimos.

E, para o IPC-C1, é considerada a estrutura média das despesas de consumo das famílias residentes nos sete municípios pesquisados, com renda mensal entre 1 e 2,5 salários mínimos.

Os bens e serviços que integram a amostra do IPC estão distribuídos em oito grupos ou classes de despesa, exibidos na Tabela 1.

Tabela 1 Grupos de Despesa da Estrutura de Bens e Serviços do IPC

Grupos
Alimentação
Habitação
Vestuário
Saúde e Cuidados Pessoais
Educação, Leitura e Recreação
Transportes
Despesas Diversas
Comunicação

### 3.1.3.2 Sistema de Pesos

A importância monetária, em termos percentuais, dos produtos e serviços que compõem a cesta do IPC é expressa por um conjunto de valores, denominados pesos, obtidos a partir da POF. O peso de um produto ou serviço  $i$  no município  $k$  é dado pela divisão da soma das despesas de todos os domicílios  $d$  com o produto ou serviço  $i$  pela despesa total de todos os domicílios  $D$  com todos os produtos e serviços do município  $k$ :

$$W_o^{i,k} = \frac{\sum_{j=1}^n d^{i,j,k}}{\sum_{j=1}^n D^{j,k}} \quad (2)$$

onde:

- $W_o^{i,k}$  = peso do produto ou serviço  $i$ , no município  $k$ , no período base  $o$ ;
- $d^{i,j,k}$  = despesa com o produto ou serviço  $i$ , no domicílio  $j$ , no município  $k$ ;

- $D^{j,k}$  = despesa total do domicílio  $j$ , no município  $k$  e
- $\sum_i W_o^{i,k} = 1$ .

Os gastos estão estruturados de acordo com a natureza das despesas da seguinte maneira: um índice geral, 8 grupos, 25 subgrupos, 85 itens e 338 subitens. O peso do subitem  $s$  no município  $k$ ,  $W_o^{s,k}$ , é obtido através dos pesos de produtos ou serviços  $i$  no município  $k$ ,  $W_o^{i,k}$ , sendo  $\sum_s W_o^{s,k} = 1$ . Da mesma forma, do somatório dos pesos dos subitens, derivam os pesos dos itens, subgrupos, grupos e geral. Esses pesos são atualizados mensalmente.

### 3.1.3.3 Sistema de Preços

A coleta de preços dos bens e serviços identificados no varejo na forma como são comercializados, denominados genericamente de insumos, é realizada em duas partes. Primeiro, donas de casa especialmente treinadas e na condição de prestadoras autônomas de serviço para a FGV coletam, a cada dez dias, conforme calendário prévio e nos mesmos estabelecimentos, preços de uma mesma lista de produtos do gênero alimentício, materiais de limpeza e artigos de higiene, cuidados e serviços pessoais. Depois, funcionários do IBRE coletam os preços dos insumos que não foram pesquisados pelas donas de casa, diretamente nos estabelecimentos informantes, uma vez por mês.

### 3.1.3.4 Sistema de Cálculo

O cálculo do índice é realizado através de uma sequência de etapas. Primeiro, calcula-se o índice a nível municipal e, depois, a nível nacional. Para o cálculo em nível de município, o primeiro passo é calcular os relativos dos insumos, dados por:

$$R_t^{j,s,l,k} = \frac{P_t^{j,s,l,k}}{P_{t-1}^{j,s,l,k}} \quad (3)$$

onde:

- $R_t^{j,s,l,k}$  = relativo do insumo  $j$ , do subitem  $s$ , no local de compras  $l$ , no município  $k$ , no período de referência  $t$ ;
- $P_t^{j,s,l,k}$  = média dos preços coletados para o insumo  $j$ , do subitem  $s$ , no local de compras  $l$ , no município  $k$ , no período de referência  $t$  e
- $P_{t-1}^{j,s,l,k}$  = média dos preços coletados para o insumo  $j$ , do subitem  $s$ , no local de compras  $l$ , no município  $k$ , no período de referência  $t - 1$ , após qualquer descarte ou alteração feita pelos analistas de preços no momento da crítica dos mesmos.

Em seguida, são calculados os relativos em nível de subitens. Os subitens são formados por insumos afins, que cobrem a diversidade de marcas e formas de comercialização. Seus relativos são obtidos por meio da média geométrica dos relativos médios dos insumos afins integrantes de cada subitem:

$$\overline{R}_t^{s,k} = \left[ \prod_{j=1}^n \overline{R}_t^{j,s,k} \right]^{\frac{1}{n}} \quad (4)$$

onde:

- $\overline{R}_t^{s,k}$  = relativo do subitem  $s$  no município  $k$ ;
- $\overline{R}_t^{j,s,k} = \left[ \prod_{l=1}^p R_t^{j,s,l,k} \right]^{\frac{1}{p}}$  = relativo médio do insumo  $j$ , onde  $p$  é o número de cotações de preços do insumo  $j$ ;
- $R_t^{j,s,l,k}$  = relativo do insumo  $j$ , do subitem  $s$ , no local de compras  $l$ , no município  $k$ , no período de referência  $t$  e
- $n$  = número de insumos do subitem  $s$  no município  $k$ .

Para os subitens do grupo “Vestuário” e subitens cujos insumos são de alta tecnologia com alterações de modelos e características muito frequentes, os relativos são calculados da seguinte maneira:

$$\overline{R}_t^{s,k} = \left[ \prod_{j=1}^n \prod_{l=1}^{m_j} R_t^{j,s,l,k} \right]^{\frac{1}{\sum_{j=1}^n m_j}} \quad (5)$$

onde:

- $\overline{R}_t^{s,k}$  = relativo do subitem  $s$  no município  $k$ ;
- $R_t^{j,s,l,k}$  = relativo do insumo  $j$ , do subitem  $s$ , no local de compras  $l$ , no município  $k$ , no período de referência  $t$ ;
- $m_j$  = número de locais de compra do insumo  $j$  e
- $n$  = número de insumos do subitem  $s$  no município  $k$ .

Por último, são calculados os relativos em nível de itens, subgrupos, grupos e geral através da média aritmética ponderada dos relativos em nível de subitens:



Tabela 2 Ponderações Regionais para o Cálculo do IPC em Nível Nacional

Municípios	Participação (%)
Belo Horizonte	8,47
Brasília	8,21
Porto Alegre	17,61
Recife	7,24
Rio de Janeiro	19,90
Salvador	10,68
São Paulo	27,89
Total	100,00

$$\bar{R}_t^{x,k} = \frac{\sum_{s=1}^{\alpha x} \bar{R}_t^{s,k} W_t^{s,k}}{\sum_{s=1}^{\alpha x} W_t^{s,k}} \quad (6)$$

onde:

- $x$  = grupamento ao nível de item, subgrupo, grupo e geral;
- $\bar{R}_t^{x,k}$  = relativo do grupamento  $x$ , do município  $k$ , no período de referência  $t$ ;
- $\bar{R}_t^{s,k}$  = relativo do subitem  $s$ , do município  $k$ , no período de referência  $t$ ;
- $W_t^{s,k} = \frac{W_0^{s,k} \prod_{i=1}^{t-1} \bar{R}_i^{s,k}}{\sum_{s=1}^{\alpha} W_0^{s,k} \prod_{i=1}^{t-1} \bar{R}_i^{s,k}}$  = a ponderação do subitem  $s$ , do município  $k$ , no período de referência  $t$ ;
- $\alpha$  = número de subitens pertencente ao grupamento “geral” e
- $\alpha x$  = número de subitens pertencente ao grupamento  $x$ .

Terminado o cálculo do índice em nível municipal, o mesmo é calculado em nível nacional por meio de pesos fixos atribuídos a cada um dos sete municípios, equivalentes à renda da população residente em cada um deles com relação ao rendimento total do conjunto de municípios. Tais pesos são exibidos na Tabela 2.

Em nível nacional, as ponderações em todos os níveis resultam da média aritmética ponderada entre os pesos regionais do subitem, tendo como ponderador a participação relativa a cada

município (Tabela 2), e os relativos, também em todos os níveis, derivam da média aritmética ponderada entre relativos regionais do subitem, tendo como ponderadores o peso regional do subitem e a percentagem (Tabela 2) referente a cada município:

$$\overline{W}_t^{x,BR} = \frac{\sum_{l=1}^7 \sum_{s=1}^{\alpha x} W_t^{s,l} \cdot Q^l}{\sum_{l=1}^7 Q^l} \quad (7)$$

$$\overline{R}_t^{x,BR} = \frac{\sum_{l=1}^7 \sum_{s=1}^{\alpha x} \overline{R}_t^{s,l} \cdot W_t^{s,l} \cdot Q^l}{\sum_{l=1}^7 \sum_{s=1}^{\alpha x} W_t^{s,l} \cdot Q^l} \quad (8)$$

onde:

- $x$  = grupamento ao nível de item, subgrupo, grupo e geral;
- $\overline{W}_t^{x,BR}$  = ponderação nacional agregada no período de referência  $t$ ;
- $W_t^{s,l}$  = ponderação do subitem  $s$ , do município  $l$ , no período de referência  $t$ ;
- $Q^l$  = participação relativa do município  $l$ ;
- $\alpha x$  = número de subitens pertencente ao grupamento  $x$  do agregado nacional;
- $\overline{R}_t^{x,BR}$  = relativo nacional agregado no período de referência  $t$  e
- $R_t^{s,l}$  = relativo do subitem  $s$ , do município  $l$ , no período de referência  $t$ .

O Índice IPC-BR referente ao mês  $t$ ,  $I_t^{BR}$ , é correspondente ao  $\overline{R}_t^{x,BR}$  onde  $x$  é igual ao grupamento “geral”.

### 3.2 Base de Dados

As informações utilizadas neste trabalho provêm da versão IPC-DI, cuja coleta de preços é realizada no período compreendido entre o primeiro e o último dia útil do mês de referência. Não há um motivo especial para a escolha desta versão, apenas levou-se em conta o fato de sua coleta ocorrer no que se chama de mês calendário, ao contrário de todas as outras, que tem início no mês anterior ao de referência.

Conforme mencionado, os insumos são bens e serviços pertencentes à cesta do IPC identificados no varejo na forma como são comercializados e os subitens são formados por insumos afins que cobrem a diversidade de marcas e formas de comercialização.

Feito um recorte na estrutura da amostra de bens e serviços do IPC, a base de dados utilizada neste estudo contém informações referentes aos insumos afins dos subitens pertencentes aos itens “Hortaliças e Legumes” e “Frutas”, que fazem parte do subgrupo “Gêneros Alimentícios”, que, por sua vez, está dentro do grupo “Alimentação”. Na Figura 3, é exibido o trecho da estrutura do IPC considerado neste trabalho. A estrutura completa é exibida no Anexo.

Na estrutura do IPC, cada um dos grupos, subgrupos, itens e subitens possuem um código, como pode ser observado na primeira coluna da Figura 3. Os códigos de um dígito são referentes aos grupos, os de dois dígitos aos subgrupos, os de quatro dígitos aos itens e os de seis dígitos aos subitens.

Escolheu-se trabalhar com os itens “Hortaliças e Legumes” e “Frutas” por ambos serem formados por subitens homogêneos, isto é, subitens que possuem apenas um insumo afim, como, por exemplo, o “Tomate”, que é formado apenas pelo insumo “Tomate Fresco ou Refrigerado (Kg)”. Já outros itens da estrutura do IPC possuem subitens heterogêneos, ou seja, subitens que possuem diversos insumos afins, como é o caso de “Móveis para Residência”, pertencente ao item “Móveis”, que faz parte do subgrupo “Mobiliário”, que, por sua vez, está dentro do grupo “Habitação”. “Armário de Quarto Duplex”, “Sofá Estofado de Tecido”, “Mesa de Madeira de Jantar” e “Armário de Cozinha” são alguns dos insumos que formam este subitem.

O arquivo de dados contém as seguintes colunas: Código do Item Elementar, Descrição do Item, Código do Insumo, Descrição do Insumo, Código do Informante, Código do Insumo Informado, Data de Referência do Serviço, Município, Data de Coleta do Preço, Preço Original, Média Oficial, Número de Cotações do Preço, Descartado e Alterado. A descrição de cada uma delas é dada a seguir.

- Código do Item Elementar: código de nove dígitos referente aos insumos que compõem um subitem. Seus primeiros seis dígitos são herdados do código do subitem e os três últimos são referentes ao insumo afim. Por exemplo: o código do subitem “Tomate” é 110371, sendo assim, o Código do Item Elementar para o “Tomate Fresco ou Refrigerado (Kg)” é 110371XXX, onde XXX são os dígitos correspondentes ao insumo “Tomate Fresco ou Refrigerado (Kg)”;
- Descrição do Item: indica se o preço pertence ao item “Hortaliças e Legumes” ou ao item “Frutas”;
- Código do Insumo: código que o insumo possui dentro do sistema utilizado no IBRE para o cálculo do IPC. É equivalente ao Código do Item Elementar, no entanto, não é exibido na

estrutura do índice, sendo usado apenas internamente;

- Descrição do Insumo: apresenta o nome completo do insumo afim;
- Código do Informante: código do local de compra;
- Código do Insumo Informado: código referente ao insumo de um determinado local de compra;
- Data de Referência do Serviço: indica o mês e o ano de referência do índice, informando o período no qual o preço foi utilizado no cálculo do mesmo;
- Município: indica em qual município o preço foi coletado. Este estudo está restrito aos municípios Rio de Janeiro e São Paulo;
- Data de Coleta do Preço: data em que o preço foi coletado. São considerados, neste trabalho, os preços coletados no período de janeiro de 2004 a dezembro de 2011;
- Preço Original: preço coletado;
- Média Oficial: média dos preços coletados, calculada após qualquer descarte ou alteração feita pelos analistas no momento da crítica dos dados. É o valor utilizado no cálculo do índice;
- Número de Cotações do Preço: quantidade de preços de um determinado insumo utilizados no cálculo do índice;
- Descartado: indica se o preço foi ou não descartado pelos analistas no momento da crítica. Contém 1, se o preço foi descartado, e 0, caso contrário;
- Alterado: indica se o preço foi ou não alterado e, caso tenha sido, exibe o valor para o qual o mesmo foi modificado. Caso contrário, contém 0.

Nas Tabelas 3 e 4, é exibida uma pequena amostra da base de dados para que se possa ter uma ideia da disposição dos mesmos. Por uma questão de confidencialidade, o Código do Item Elementar, o Código do Insumo, o Código do Informante e o Código do Insumo Informado não podem ser divulgados. Portanto, tais colunas tiveram seu conteúdo substituído por números aleatórios, sendo que cada número aleatório gerado corresponde sempre ao mesmo código, seja ele do item elementar, do insumo, do informante ou do insumo informado.

Código	Descrição
0	IPC - TODOS OS ITENS
1	ALIMENTAÇÃO
11	GÊNEROS ALIMENTÍCIOS
1103	HORTALIÇAS E LEGUMES
110301	ABÓBORA
110309	AIPIM / MANDIOCA
110313	ALFACE
110315	ALHO
110319	BATATA-INGLESA
110331	BRÓCOLIS
110333	CEBOLA
110335	CENOURA
110345	COUVE
110357	INHAME
110363	PIMENTÃO
110371	TOMATE
110377	COENTRO
1105	FRUTAS
110503	ABACAXI
110507	BANANA-NANICA (D'ÁGUA)
110511	BANANA-PRATA
110513	BANANA DA TERRA
110527	LARANJA-PERA
110534	MAÇÃ
110537	MAMÃO PAPAYA
110539	MAMÃO FORMOSA
110549	PERA
110553	TANGERINA (MEXERICA)
110555	UVA
110569	MELANCIA
110571	MARACUJÁ
110573	MANGA

Figura 3 Recorte da Estrutura do Índice de Preços do Consumidor (IPC)

Tabela 3 Amostra da Base de Dados - Parte 1

Cód. Item Elem.	Desc. Item	Cód. Ins.	Desc. Ins.	Cód. Inf.	Cód. Ins. Inf.	Mun.
711	Hortaliças e Legumes	727	Abóbora Fresca ou Refrigerada (Kg)	973	8583	RJ
711	Hortaliças e Legumes	727	Abóbora Fresca ou Refrigerada (Kg)	973	8583	RJ
711	Hortaliças e Legumes	727	Abóbora Fresca ou Refrigerada (Kg)	973	8583	RJ
711	Hortaliças e Legumes	727	Abóbora Fresca ou Refrigerada (Kg)	973	8583	RJ
711	Hortaliças e Legumes	727	Abóbora Fresca ou Refrigerada (Kg)	973	8583	RJ
711	Hortaliças e Legumes	727	Abóbora Fresca ou Refrigerada (Kg)	973	8583	RJ
;	;	;	;	;	;	;
711	Hortaliças e Legumes	727	Abóbora Fresca ou Refrigerada (Kg)	973	8583	RJ
711	Hortaliças e Legumes	727	Abóbora Fresca ou Refrigerada (Kg)	973	8583	RJ
711	Hortaliças e Legumes	727	Abóbora Fresca ou Refrigerada (Kg)	973	8583	RJ
;	;	;	;	;	;	;
996	Hortaliças e Legumes	779	Tomate Fresco ou Refrigerado (Kg)	228	5274	RJ
996	Hortaliças e Legumes	779	Tomate Fresco ou Refrigerado (Kg)	228	5274	RJ
996	Hortaliças e Legumes	779	Tomate Fresco ou Refrigerado (Kg)	228	5274	RJ
996	Hortaliças e Legumes	779	Tomate Fresco ou Refrigerado (Kg)	228	5274	RJ
;	;	;	;	;	;	;
996	Hortaliças e Legumes	779	Tomate Fresco ou Refrigerado (Kg)	228	5274	RJ
996	Hortaliças e Legumes	779	Tomate Fresco ou Refrigerado (Kg)	228	5274	RJ
996	Hortaliças e Legumes	779	Tomate Fresco ou Refrigerado (Kg)	228	5274	RJ
;	;	;	;	;	;	;
996	Hortaliças e Legumes	779	Tomate Fresco ou Refrigerado (Kg)	228	5274	RJ
996	Hortaliças e Legumes	779	Tomate Fresco ou Refrigerado (Kg)	228	5274	RJ
996	Hortaliças e Legumes	779	Tomate Fresco ou Refrigerado (Kg)	228	5274	RJ
;	;	;	;	;	;	;
996	Hortaliças e Legumes	779	Tomate Fresco ou Refrigerado (Kg)	228	5274	RJ
996	Hortaliças e Legumes	779	Tomate Fresco ou Refrigerado (Kg)	228	5274	RJ
996	Hortaliças e Legumes	779	Tomate Fresco ou Refrigerado (Kg)	228	5274	RJ
;	;	;	;	;	;	;

Tabela 4 Amostra da Base de Dados - Parte 2

Cód. Item Elem.	Dt. Ref. Serv.	Dt. Col. Preço	Preço Original	Preço Oficial	Núm. Cot. Preço	Descartado	Alterado
711	01/2004	05/01/2004	0.38	0.406667	3	0	0
711	01/2004	12/01/2004	0.39	0.406667	3	0	0
711	01/2004	26/01/2004	0.45	0.406667	3	0	0
711	02/2004	02/02/2004	0.45	0.46	3	0	0
711	02/2004	09/02/2004	0.45	0.46	3	0	0
711	02/2004	16/02/2004	0.48	0.46	3	0	0
;	;	;	;	;	;	;	;
711	05/2008	09/05/2008	0.99	1.09	3	0	0
711	05/2008	19/05/2008	0.99	1.09	3	0	0
711	05/2008	29/05/2008	1.29	1.09	3	0	0
;	;	;	;	;	;	;	;
996	08/2004	24/08/2004	2.66	2.66	1	0	0
996	09/2004	07/09/2004	1.75	2.1	3	0	0
996	09/2004	14/09/2004	2.45	2.1	3	0	0
996	09/2004	21/09/2004	2.1	2.1	3	0	0
;	;	;	;	;	;	;	;
996	03/2008	04/03/2008	1.82	2,006667	3	0	0
996	03/2008	14/03/2008	2,1	2,006667	3	0	0
996	03/2008	24/03/2008	3,15	2,006667	3	0	2,1
;	;	;	;	;	;	;	;
996	12/2011	04/12/2011	2,28	3,113333	3	0	0
996	12/2011	14/12/2011	2,86	3,113333	3	0	0
996	12/2011	24/12/2011	4,2	3,113333	3	0	0
;	;	;	;	;	;	;	;

## 4 METODOLOGIA

Os métodos para detecção de *outliers* considerados neste estudo, além do método utilizado atualmente no IBRE pelos analistas de preços, são: variações do Método do IBRE, Método do Boxplot, Método do Boxplot SIQR, Método do Boxplot Ajustado, Método de Cercas Resistentes, Método do Quartil, Método do Quartil Modificado, Método do Desvio Mediano Absoluto e Algoritmo de Tukey.

Todos os métodos aqui explorados envolvem a construção de um intervalo de tolerância. Valores pertencentes a este intervalo são considerados aceitáveis, enquanto valores menores que o limite inferior ou maiores que o limite superior deste intervalo são considerados *outliers*.

Este trabalho foi elaborado no software estatístico *R* (R Core Team, 2013). Todos os métodos aqui estudados foram implementados nesta linguagem, com exceção do Método do Boxplot e do Método do Boxplot Ajustado, que já estão disponíveis nos pacotes *graphics*<sup>7</sup> e *robustbase*<sup>8</sup> do *R*, respectivamente. Os códigos estão disponíveis no Apêndice B.

Na aplicação dos métodos, os dados utilizados são os relativos de preços obtidos por meio da Equação 3, agrupados por insumo e por município. Isto é, os métodos são aplicados nos relativos de preços referentes a um mesmo insumo e pertencentes a um determinado município, como, por exemplo, os relativos de preços referentes ao insumo “Tomate Fresco ou Refrigerado (Kg)” do Rio de Janeiro. O código utilizado para o cálculo dos relativos está disponível no Apêndice A.

Nas Tabelas 5 e 6, é exibida uma amostra dos relativos de preços gerados pela Equação 3. Estas tabelas são formadas pelas colunas Código do Item Elementar, Descrição do Item, Código do Insumo, Descrição do Insumo, Código do Informante, Código do Insumo Informado e Município (Tabela 5), e Código do Item Elementar, Data de Referência do Serviço, Média Oficial e Número de Cotações do Preço, cujas descrições foram dadas na Seção 3, e pelas colunas Média Original, Número de Preços Descartados, Número de Preços Alterados e Relativo (Tabela 6), cujas descrições são dadas a seguir:

- Média Original: média dos preços coletados, calculada antes de qualquer descarte ou alteração feita pelos analistas no momento da crítica dos dados;
- Número de Preços Descartados: quantidade de preços coletados que foram descartados pelos analistas;
- Número de Preços Alterados: quantidade de preços coletados que foram alterados pelos analistas;

<sup>7</sup><http://svitsrv25.epfl.ch/R-doc/library/graphics/html/boxplot.html> em 15 de janeiro de 2014

<sup>8</sup><http://svitsrv25.epfl.ch/R-doc/library/robustbase/html/adjbox.html> em 15 de janeiro de 2014



- Relativo: relativo gerado pela Equação 3.

Vale ressaltar que o valor exibido na coluna Média Original é diferente do da coluna Média Oficial sempre que um ou mais preços são descartados ou alterados.

Conforme definido na Seção 3, os relativos de preços obtidos por meio da Equação 3 levam em conta qualquer descarte ou alteração feita pelos analistas de preços no momento da crítica dos mesmos. Este fato é ilustrado na Tabela 6: pode-se observar que algum (uns) dos três preços coletados para o insumo “Pepino Fresco ou Refrigerado Comum (Kg)”, referentes a março de 2004, foi alterado. Portanto, o relativo referente a abril de 2004 é calculado considerando a média dos preços coletados em março após a alteração feita pelos analistas de preços (1,26), ao invés da média calculada antes da alteração (1,59):  $\frac{1,68}{1,26} = 1,33$ .

De forma a considerar não só o comportamento atual dos preços mas também a influência, no presente, do comportamento dos mesmos no passado, os limites inferiores e superiores dos intervalos de tolerância dos métodos do Boxplot, do Boxplot SIQR, do Boxplot Ajustado, de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e do Algoritmo de Tukey são calculados por meio de cinco diferentes conjuntos de dados, especificados a seguir.

1. Mês de Referência (MR): relativos de preços obtidos por meio da Equação 3, referentes a um mesmo insumo e pertencentes a um determinado município, do mês de referência. Isto é, supondo o cálculo do índice referente a fevereiro de 2011, este conjunto de dados é composto pelos relativos de preços de fevereiro de 2011 contra janeiro de 2011 do insumo e município em questão;
2. Mês Anterior (MA): relativos de preços obtidos por meio da Equação 3, referentes a um mesmo insumo e pertencentes a um determinado município, do mês imediatamente anterior ao de referência. Isto é, supondo o cálculo do índice referente a fevereiro de 2011, este conjunto de dados é composto pelos relativos de preços de janeiro de 2011 contra dezembro de 2010 do insumo e município em questão;
3. Dois Meses Anteriores (2MA): relativos de preços obtidos por meio da Equação 3, referentes a um mesmo insumo e pertencentes a um determinado município, dos dois meses imediatamente anteriores ao de referência. Isto é, supondo o cálculo do índice referente a fevereiro de 2011, este conjunto de dados é composto pelos relativos de preços de janeiro de 2011 contra dezembro de 2010 e dezembro de 2010 contra novembro de 2010 do insumo e município em questão;
4. Mesmo Mês dos Anos Anteriores (MMAA): relativos de preços obtidos por meio da Equação 3, referentes a um mesmo insumo e pertencentes a um determinado município, do mesmo

mês dos anos anteriores. Isto é, supondo o cálculo do índice referente a fevereiro de 2011, este conjunto de dados é composto pelos relativos de preços de fevereiro de 2004 contra janeiro de 2004, fevereiro de 2005 contra janeiro de 2005, fevereiro de 2006 contra janeiro de 2006, fevereiro de 2007 contra janeiro de 2007, fevereiro de 2008 contra janeiro de 2008, fevereiro de 2009 contra janeiro de 2009 e fevereiro de 2010 contra janeiro de 2010, do insumo e município em questão;

5. Todos os Anteriores (TA): todos os relativos de preços anteriores, obtidos por meio da Equação 3, referentes a um mesmo insumo e pertencentes a um determinado município. Isto é, supondo o cálculo do índice referente a fevereiro de 2011, este conjunto de dados é composto pelos relativos de preços de fevereiro de 2004 contra janeiro de 2004, março de 2004 contra fevereiro de 2004, . . . , dezembro de 2010 contra novembro de 2010 e janeiro de 2011 contra dezembro de 2010, do insumo e município em questão.

A descrição das fórmulas para o cálculo dos limites inferior e superior do intervalo de tolerância gerado por cada um dos métodos é dada a seguir.

Tabela 5 Relativos Gerados pela Equação 3 - Parte 1

Cód. Item Elem.	Desc. Item	Cód. Ins.	Desc. Ins.	Cód. Inf.	Cód. Ins. Inf.	Mun.
22	Frutas	878	Abacaxi Fresco ou Refrigerado (Unid)	128	7591	RJ
22	Frutas	878	Abacaxi Fresco ou Refrigerado (Unid)	128	7591	RJ
22	Frutas	878	Abacaxi Fresco ou Refrigerado (Unid)	128	7591	RJ
22	Frutas	878	Abacaxi Fresco ou Refrigerado (Unid)	128	7591	RJ
22	Frutas	878	Abacaxi Fresco ou Refrigerado (Unid)	128	7591	RJ
⋮	⋮	⋮	⋮	⋮	⋮	⋮
22	Frutas	878	Abacaxi Fresco ou Refrigerado (Unid)	128	7591	RJ
22	Frutas	878	Abacaxi Fresco ou Refrigerado (Unid)	128	7591	RJ
22	Frutas	878	Abacaxi Fresco ou Refrigerado (Unid)	128	7591	RJ
⋮	⋮	⋮	⋮	⋮	⋮	⋮
82	Hortaliças e Legumes	887	Pepino Fresco ou Refrigerado Comum (Kg)	608	4245	SP
82	Hortaliças e Legumes	887	Pepino Fresco ou Refrigerado Comum (Kg)	608	4245	SP
82	Hortaliças e Legumes	887	Pepino Fresco ou Refrigerado Comum (Kg)	608	4245	SP
82	Hortaliças e Legumes	887	Pepino Fresco ou Refrigerado Comum (Kg)	608	4245	SP
82	Hortaliças e Legumes	887	Pepino Fresco ou Refrigerado Comum (Kg)	608	4245	SP
⋮	⋮	⋮	⋮	⋮	⋮	⋮
82	Hortaliças e Legumes	887	Pepino Fresco ou Refrigerado Comum (Kg)	608	4245	SP
82	Hortaliças e Legumes	887	Pepino Fresco ou Refrigerado Comum (Kg)	608	4245	SP
82	Hortaliças e Legumes	887	Pepino Fresco ou Refrigerado Comum (Kg)	608	4245	SP
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Tabela 6 Relativos Gerados pela Equação 3 - Parte 2

Cód. Item Elem.	Dt. Ref. Serv.	Média Original	Média Oficial	Núm. Cot. Preço	Num. Desc.	Num. Alt.	Relativo
22	01/2004	1,540000	1,540000	3	0	0	-
22	02/2004	1,743333	1,743333	3	0	0	1,13203
22	03/2004	2,103333	2,103333	3	0	0	1,20650
22	04/2004	2,103333	2,103333	3	0	0	1,00000
22	05/2004	2,140000	2,140000	3	0	0	1,01743
∴	∴	∴	∴	∴	∴	∴	∴
22	10/2011	3,576667	3,576667	3	0	0	1,10962
22	11/2011	2,916667	2,916667	3	0	0	0,81547
22	12/2011	3,150000	3,150000	3	0	0	1,08000
∴	∴	∴	∴	∴	∴	∴	∴
82	01/2004	1,356667	1,356667	3	0	0	-
82	02/2004	1,290000	1,290000	3	0	0	0,95086
82	03/2004	1,590000	1,256667	3	0	1	1,23256
82	04/2004	1,676667	1,676667	3	0	0	1,33422
82	05/2004	2,586667	2,586667	3	0	0	1,54274
∴	∴	∴	∴	∴	∴	∴	∴
82	10/2011	2,123333	2,123333	3	0	0	0,87620
82	11/2011	1,323333	1,323333	3	0	0	0,62323
82	12/2011	1,256667	1,256667	3	0	0	0,94962
∴	∴	∴	∴	∴	∴	∴	∴

#### 4.1 Método do IBRE

Uma das atividades de maior destaque no IBRE é o cálculo de índices de preços. Um dos processos mais importantes para esta tarefa é a atividade de crítica executada pelos analistas de preços. Nesse processo, o corpo técnico utiliza ferramentas construídas pela experiência do Instituto na apuração de indicadores.

Para cada produto ou serviço componente do índice, observa-se que os reajustes de preços ocorrem com maior frequência dentro de um intervalo. Com esta percepção, são estabelecidos intervalos para facilitar a identificação de *outliers*.

Para preços administrados, serviços não expostos a choques de oferta, utilizam-se parâmetros de crítica mais restritos para o monitoramento mais eficiente da trajetória de seus preços.

Para alimentos *in natura*, utilizam-se parâmetros de crítica mais amplos capazes de absorver choques de oferta típicos para esse grupo de produtos. Sendo assim, no caso dos itens “Hortaliças e Legumes” e “Frutas”, os parâmetros foram fixados em  $\pm 50\%$ , o que significa que o método atual do IBRE sinaliza como valores extremos as variações com relação ao mês anterior iguais ou inferiores a  $-50\%$  e iguais ou superiores a  $50\%$ .

Sinalizar as variações com relação ao mês anterior iguais ou inferiores a  $-50\%$  e iguais ou superiores a  $50\%$  é equivalente a sinalizar os relativos de preços  $R_t^{j,s,l,k}$  iguais ou inferiores a 0.5 e iguais ou superiores a 1.5, onde  $R_t^{j,s,l,k}$  é o relativo do insumo  $j$ , do subitem  $s$ , no local de compras  $l$ , no município  $k$ , no período de referência  $t$ , obtido através da Equação 3. Sendo assim, o intervalo de tolerância para este método é dado por:

$$[0,5, 1,5] \quad (9)$$

Uma vez sinalizados, cabe aos analistas de preços decidir se os mesmos são realmente *outliers* ou se estão corretos. Caso sejam considerados *outliers*, os analistas devem escolher entre descartá-los ou alterá-los.

Determinados itens do IPC são compostos por produtos muito heterogêneos, como, por exemplo, “Peças para Automóvel”. Neste item, são acompanhados os preços de diversas peças de reposição. Contudo, não há uma amostra robusta para cada tipo de peça cujo preço é pesquisado. A amostra é construída pelo conjunto de todas as peças. Neste caso, não há como estimar um preço médio para cada peça e utilizá-lo como uma alternativa à substituição de um preço *outlier*, questão que leva os analistas a excluírem o preço extremo da amostra de apuração do índice.

Para produtos homogêneos, como o “Tomate”, por exemplo, a amostra é constituída somente por um tipo de produto. Neste caso, há como se estimar um preço médio e utilizá-lo na atividade de crítica quando necessário.

Ao processo de crítica descrito, dá-se o nome de *Crítica Horizontal*. Além deste processo,

há também a chamada *Crítica Vertical*. Nesta última, o preço do produto coletado é comparado aos outros preços pesquisados para o mesmo insumo em outros estabelecimentos comerciais. Esses processos são complementares e essenciais para a precisão do indicador. Resta acrescentar, que a coleta do IPC acompanha a evolução dos preços dos produtos e serviços preservando o local de compra e a descrição do item pesquisado.

Supondo que os preços de um insumo qualquer sejam coletados nos locais de compras A, B e C, de um determinado município, no período compreendido entre janeiro de 2004 a dezembro de 2011, a Figura 4 ilustra as duas críticas feitas pelos analistas de preços.

Locais de Compras	Meses				
	jan/04	...	out/11	nov/11	dez/11
Local de Compras A	1,94	...	2,89	3,19	3,16
Local de Compras B	2,39	...	2,39	2,69	3,59
Local de Compras C	2,32	...	2,14	2,94	3,25

Figura 4 Tipos de Críticas Realizadas no IPC do IBRE

## 4.2 Variações do Método do IBRE

As variações do Método do IBRE consistem em considerar diferentes valores para os parâmetros de crítica dos itens “Hortaliças e Legumes” e “Frutas”. A fórmula geral para o intervalo de tolerância gerado pelas variações do Método do IBRE é dada por:

$$[(1 - p), (1 + p)] \quad (10)$$

onde  $p = 30\%, 35\%, 40\%, 45\%, 46\%, 47\%, 48\%, 49\%, 49,25\%, 49,5\%$  e  $49,75\%$ .

Todos os valores considerados para  $p$  geram intervalos de tolerância menores que o do Método do IBRE. Portanto, todas as variações do Método do IBRE consideradas neste estudo detectam uma quantidade de *outliers* maior ou igual que a do método original.

## 4.3 Método do Boxplot

O intervalo de tolerância gerado pelo Método do Boxplot, segundo Tukey (1977), é dado por:

$$[Q_1 - 1,5 \times IQR, Q_3 + 1,5 \times IQR] \quad (11)$$

onde  $Q_1$  e  $Q_3$  são, respectivamente, o primeiro quartil e o terceiro quartil, e  $IQR$  é o intervalo interquartílico, dado por  $Q_3 - Q_1$ .

#### 4.4 Método do Boxplot SIQR

O Método do Boxplot SIQR, introduzido por Kimber (1990), é semelhante ao Método do Boxplot, mas, no lugar do intervalo interquartílico, o Método do Boxplot SIQR utiliza os intervalos semi-interquartílicos inferior ( $SIQR_L$ ) e superior ( $SIQR_U$ ) na construção de seu intervalo de tolerância, dado por:

$$[Q_1 - 3 \times SIQR_L, Q_3 + 3 \times SIQR_U] \quad (12)$$

onde  $Q_1$  é o primeiro quartil,  $Q_3$  é o terceiro quartil,  $SIQR_L$  é o intervalo semi-interquartílico inferior, dado por  $Q_2 - Q_1$ , onde  $Q_2$  é a mediana, e  $SIQR_U$  é o intervalo semi-interquartílico superior, dado por  $Q_3 - Q_2$ . A outra diferença entre os métodos é que o do Boxplot multiplica o intervalo interquartílico pela constante 1.5, enquanto o do Boxplot SIQR multiplica os intervalos semi-interquartílicos inferior e superior pela constante 3.

#### 4.5 Método do Boxplot Ajustado

Para a construção do intervalo de tolerância gerado pelo Método do Boxplot Ajustado, deve-se calcular primeiramente uma medida de assimetria denominada *medcouple* ( $MC$ ), introduzida por Brys, Hubert e Struyf (2004), para, em seguida e de acordo com o valor obtido para esta medida, computar os limites inferior e superior do intervalo. O passo a passo para a construção do mesmo, segundo Hubert e Vandervieren (2008), é exibido a seguir:

1. Ordenar os preços relativos;
2. Calcular a mediana dos relativos ordenados;
3. Dividir os dados em dois grupos: o primeiro formado pelos relativos menores que a mediana, representado pelo índice  $i$ , e o segundo formado pelos relativos maiores que a mediana, representado pelo índice  $j$ ;
4. Para cada par  $(x_i, x_j)$ , onde  $x_i \neq x_j$ , calcular a função  $h(x_i, x_j)$ , dada por:

$$h(x_i, x_j) = \frac{(x_j - Q_2) - (Q_2 - x_i)}{x_j - x_i} \quad (13)$$

onde  $Q_2$  é a mediana.

5. Dos valores resultantes, calcular o  $MC$ , dado por:

$$MC = med_{x_i \leq Q_2 \leq x_j} h(x_i, x_j) \quad (14)$$

O  $MC$  é semelhante ao *Quartile Skewness* ( $QS$ ), introduzido por Bowley (1920), e ao *Octile Skewness* ( $OS$ ), introduzido por Hinkley (1975), dados, respectivamente, por:

$$QS = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} \quad (15)$$

$$OS = \frac{(Q_{0.875} - Q_2) - (Q_2 - Q_{0.125})}{Q_{0.875} - Q_{0.125}} \quad (16)$$

onde  $Q_1$  é o primeiro quartil,  $Q_2$  é a mediana,  $Q_3$  é o terceiro quartil e  $Q_{0.875}$  e  $Q_{0.125}$  são os quantiles de 0,125 e 0,875. O  $MC$  combina o  $QS$  e o  $OS$ , pois tem a sensibilidade do  $OS$  para detectar assimetria e a robustez do  $QS$  para outliers. O  $MC$  varia de -1 a 1. Para uma distribuição enviesada para a direita, o  $MC$  assume valores positivos. Para uma distribuição enviesada para a esquerda, o  $MC$  assume valores negativos. E, para distribuições simétricas, tem-se  $MC = 0$ .

6. De acordo com o valor de  $MC$ , calcular os limites inferior e superior do intervalo de tolerância do Método do Boxplot Ajustado através das seguintes equações:

$$MC \geq 0 \begin{cases} LI = Q_1 - 1,5e^{-4MC} IQR \\ LS = Q_3 + 1,5e^{3MC} IQR \end{cases} \quad (17)$$

$$MC < 0 \begin{cases} LI = Q_1 - 1,5e^{-3MC} IQR \\ LS = Q_3 + 1,5e^{4MC} IQR \end{cases} \quad (18)$$

onde  $Q_1$  e  $Q_3$  são, respectivamente, o primeiro quartil e o terceiro quartil, e  $IQR$  é o intervalo interquartílico, dado por  $Q_3 - Q_1$ .



#### 4.6 Método de Cercas Resistentes

O intervalo de tolerância do Método de Cercas Resistentes, segundo Rais (2008), é dado por:

$$(Q_1 - c_L \times IQR, Q_3 + c_U \times IQR) \quad (19)$$

onde  $Q_1$  e  $Q_3$  são, respectivamente, o primeiro e o terceiro quartil,  $IQR$  é o intervalo interquartil dado por  $Q_3 - Q_1$  e  $c_L$  e  $c_U$  são constantes predeterminadas.

Quando  $c_L = c_U = c = 1,5$ , este método se iguala ao do Boxplot.

#### 4.7 Método do Quartil

Segundo Rais (2008), o intervalo de tolerância para o Método do Quartil é dado por:

$$[Q_2 - c_L \times \max \{SIQR_L, |aQ_2|\}, Q_2 + c_U \times \max \{SIQR_U, |aQ_2|\}] \quad (20)$$

onde  $Q_2$  é a mediana,  $SIQR_L$  é o intervalo semi-interquartil inferior, dado por  $Q_2 - Q_1$ ,  $SIQR_U$  é o intervalo semi-interquartil superior, dado por  $Q_3 - Q_2$  e  $a$ ,  $c_L$  e  $c_U$  são constantes predeterminadas, sendo  $0 \leq a \leq 1$ .

Quando a variabilidade do conjunto de dados é muito pequena, o intervalo de tolerância gerado é bem estreito e sinaliza como *outlier* qualquer observação que desvie ligeiramente da mediana. Portanto, ao invés de considerar apenas os intervalos semi-interquartis inferior e superior, a fórmula do intervalo de tolerância do Método do Quartil identifica o valor máximo entre os intervalos semi-interquartis inferior e superior e o valor absoluto de  $aQ_2$ , impondo um intervalo de tolerância mínimo.

#### 4.8 Método do Quartil Modificado

A fórmula utilizada no cálculo dos limites inferior e superior do intervalo de tolerância gerado pelo Método do Quartil Modificado difere ligeiramente da utilizada no Método do Quartil:

$$[Q_2 - \max \{c_L \times SIQR_L, |aQ_2|\}, Q_2 + \max \{c_U \times SIQR_U, |aQ_2|\}] \quad (21)$$

onde  $Q_2$  é a mediana,  $SIQR_L$  é o intervalo semi-interquartil inferior, dado por  $Q_2 - Q_1$ ,  $SIQR_U$  é o intervalo semi-interquartil superior, dado por  $Q_3 - Q_2$  e  $a$ ,  $c_L$  e  $c_U$  são constantes predeterminadas, sendo  $0 \leq a \leq 1$ .

A diferença consiste em multiplicar apenas os intervalos semi-interquartis inferior e

superior pelas constantes  $c_L$  e  $c_U$ , respectivamente, ao invés das mesmas multiplicarem também  $|aQ_2|$ , na busca do valor máximo.

#### 4.9 Método do Desvio Mediano Absoluto

Segundo Hutton (2008), o intervalo de tolerância para o Método do Desvio Mediano Absoluto é dado por:

$$\left[ Q_2 - c_L \times MAD_{(R_t^{j,s,l,k} - Q_2)}, Q_2 + c_U \times MAD_{(R_t^{j,s,l,k} - Q_2)} \right] \quad (22)$$

onde  $Q_2$  é a mediana,  $R_t^{j,s,l,k}$  é o relativo do insumo  $j$ , do subitem  $s$ , no local de compras  $l$ , no município  $k$ , no período de referência  $t$ , gerado pela Equação 3,  $MAD_{(R_t^{j,s,l,k} - Q_2)}$  é a mediana calculada a partir dos valores absolutos de  $(R_t^{j,s,l,k} - Q_2)$  e  $c_L$  e  $c_U$  são constantes predeterminadas.

#### 4.10 Algoritmo de Tukey

O Algoritmo de Tukey é um método de detecção de outliers em dados racionais, mais especificamente em dados de preços relativos. Segundo Rais (2008), a versão do Algoritmo de Tukey considerada neste estudo é a descrita em United Kingdom's Office of National Statistics (2006):

1. Ordenar os preços relativos em ordem crescente;
2. Remover os preços relativos iguais a 1;
3. Remover 5% das caudas, sendo 2,5% da cauda inferior e 2,5% da cauda superior, e seja  $D$  o conjunto de dados formado pelos relativos restantes;
4. Sejam  $\bar{x}_D$  a média aritmética de  $D$ ,  $\bar{x}_L$  a média aritmética dos preços relativos menores que  $\bar{x}_D$  e  $\bar{x}_U$  a média aritmética dos preços relativos maiores que  $\bar{x}_D$ . E então, sejam  $\Delta\bar{x}_L = \bar{x}_D - \bar{x}_L$  e  $\Delta\bar{x}_U = \bar{x}_U - \bar{x}_D$ .

O intervalo de tolerância deste método é dado por:

$$[\bar{x}_D - c_L \Delta\bar{x}_L, \bar{x}_D + c_U \Delta\bar{x}_U] \quad (23)$$

onde  $c_L$  e  $c_U$  são constantes predeterminadas.

## 5 RESULTADOS

Os métodos descritos na Seção 4 são aplicados para cada um dos insumos dos subitens da Figura 3, pertencentes aos municípios Rio de Janeiro e São Paulo e os resultados com as respectivas análises são apresentados nessa Seção.

Todos os métodos aqui abordados, com exceção do Método do IBRE e suas variações, requerem uma quantidade mínima de informações para serem aplicados. Neste estudo, considera-se que cada um dos conjuntos de dados descritos na Seção 4 deve conter, pelo menos, duas informações. De forma a possibilitar a comparação entre todos os métodos, o Método do IBRE e suas variações também foram aplicados sob esta condição.

Conforme mencionado, na aplicação dos métodos do Quartil, do Quartil Modificado, de Cercas Resistentes e do Desvio Mediano Absoluto, para considerar  $c_L = c_U$ , é necessário que a distribuição dos dados seja simétrica. Já o Algoritmo de Tukey não requer que os dados sejam simétricos.

Neste trabalho, não é feito nenhum estudo sobre a distribuição dos dados e supõe-se que a mesma seja simétrica. Portanto, o Método de Cercas Resistentes, o Método do Quartil, o Método do Quartil Modificado, o Método do Desvio Mediano Absoluto e o Algoritmo de Tukey são aplicados considerando  $c_L = c_U = c$ , onde  $c = \{0; 0,1; 0,2; \dots, 3,8; 3,9; 4\}$ . Quanto menor for o valor de  $c$ , menor será o intervalo de tolerância gerado e maior será a quantidade de *outliers* detectados. Quanto maior for o valor de  $c$ , maior será o intervalo de tolerância gerado e menor será a quantidade de *outliers* detectados. Além disso, o Método do Quartil e o Método do Quartil Modificado são empregados considerando  $a = 0,05$ , pois, segundo Lee et al. (1992), este é um valor razoável na maioria das aplicações.

Para que se possa comparar o desempenho de métodos para detecção de *outliers*, é necessário conhecer os verdadeiros valores extremos antecipadamente. Neste trabalho, tal comparação é feita assumindo que os relativos cujos preços que o formam foram descartados ou alterados pelos analistas do IBRE da FGV são os verdadeiros *outliers*.

O Método do IBRE é bastante correlacionado com os preços alterados ou descartados pelos analistas. Isto porque muitos preços podem ser modificados ou desconsiderados simplesmente pelo fato do relativo formado por eles ser menor que o limite inferior ou maior que o limite superior do intervalo de tolerância do método, o que contribui para o aumento da proporção de acerto do mesmo. Ademais, os preços alterados ou descartados pelos analistas podem não ser, de fato, os verdadeiros *outliers*. Isto é, pode acontecer de preços que foram modificados ou desprezados não serem aberrantes ou pode ocorrer de preços que foram considerados como discrepantes não serem. Sendo assim, a suposição de que os preços alterados ou descartados pelos analistas são os verdadeiros valores extremos pode influenciar os resultados, fazendo com que o mesmo seja favorecido

em comparação com os demais métodos.

Na base de dados utilizada neste estudo, nenhum preço foi descartado e 316 relativos foram alterados por conta de modificações feitas pelos analistas em um ou mais preços que os formam. Destes 316 relativos, 120 pertencem ao município do Rio de Janeiro e 196 ao município de São Paulo.

Desta forma, é possível computar duas medidas através das quais os métodos são avaliados. A primeira é a porcentagem de acerto do método, dada pela divisão da quantidade de *outliers* detectados corretamente pela quantidade de verdadeiros *outliers*. A segunda é o número de falsos positivos produzidos pelo método, dado pela divisão da quantidade total de *outliers* detectados pela quantidade de *outliers* detectados corretamente. Diz-se que um método detecta corretamente um *outlier* quando sinaliza como valor extremo um verdadeiro *outlier*, isto é, um relativo alterado pelos analistas de preços. A porcentagem de acerto do método informa a proporção de verdadeiros *outliers* detectados e o número de falsos positivos produzidos pelo método informa quantos valores precisaram ser sinalizados para um verdadeiro *outlier* ser detectado. Quanto maior for a proporção de acerto do método e menor for a quantidade de falsos positivos produzida, melhor será o desempenho do método.

Nesta Seção, a análise do desempenho dos métodos é feita por partes. Primeiramente, os métodos são divididos em três grupos. O primeiro é formado pelo Método do IBRE e por suas variações. O segundo é constituído pelos Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado. Por fim, os Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e o Algoritmo de Tukey integram o terceiro grupo.

O primeiro grupo é formado por métodos cuja construção dos intervalos de tolerância não depende dos dados. Já o segundo e o terceiro grupos são compostos por métodos cujo cálculo dos limites inferior e superior dos intervalos de tolerância depende dos dados. Os métodos do terceiro grupo dependem, ainda, dos valores atribuídos para as constantes  $a$  (Método do Quartil e Método do Quartil Modificado),  $c_L$  e  $c_U$ .

Para cada grupo, os métodos são analisados e aquele com o melhor desempenho dentre os demais é considerado o vencedor. Então, na Seção 5.1, os resultados do Método do IBRE e os de suas variações são apresentados e comparados entre si. Na Seção 5.2, é feita uma comparação entre os Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado. Na Seção 5.3, os resultados dos Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e do Algoritmo de Tukey são comparados entre si. Por fim, na Seção 5.4, é feita uma comparação entre os métodos vencedores, tornando possível a construção de um *ranking*.

As análises são realizadas para cada um dos municípios considerados neste estudo: Rio de Janeiro e São Paulo. Ao todo, os métodos foram aplicados em 50.116 relativos referentes ao município do Rio de Janeiro e em 55.560 referentes ao município de São Paulo.

## 5.1 Método do IBRE e Variações

A seguir, são apresentados os resultados do Método do IBRE e de suas variações para o município do Rio de Janeiro e de São Paulo.

### 5.1.1 Método do IBRE e Variações para o Rio de Janeiro

Dos 50.116 relativos referentes ao Rio de Janeiro, 2.945 foram classificados como *outliers* pelo Método do IBRE. Destes 2.945, 58 foram alterados pelos analistas de preços e são, portanto, considerados como verdadeiros *outliers*. Sendo assim, o Método do IBRE identificou corretamente 58 *outliers*. Como foram alterados, ao todo, 120 relativos pertencentes ao Rio de Janeiro, a proporção de acerto do Método do IBRE neste município é dada por:  $\frac{58}{120} = 0,48333$ . Já o número de falsos positivos produzidos pelo Método do IBRE no município do Rio de Janeiro, dado pela divisão da quantidade total de *outliers* detectados (2.945) pela quantidade de *outliers* detectados corretamente (58), é igual a:  $\frac{2.945}{58} = 51$ . Ou seja, a cada 51 relativos sinalizados, um *outlier* é detectado corretamente pelo Método do IBRE.

As variações do Método do IBRE com  $p = 49,75\%$ ,  $49,5\%$ ,  $49,25\%$  e  $49\%$  apresentaram desempenhos ligeiramente superiores ao do método original. Com  $p = 49,75\%$ , a quantidade total de *outliers* detectados foi de 2.968. Destes, 59 foram detectados corretamente. Assim, a proporção de acerto desta variação é de 0,49167 e o número de falsos positivos produzidos por ela é igual a 50. Com  $p = 49,5\%$ , a quantidade total de *outliers* detectados foi de 3.004. Destes, 59 foram detectados corretamente. Sendo assim, a proporção de acerto desta variação também é de 0,49167 e o número de falsos positivos produzidos por ela é igual a 51. Com  $p = 49,25\%$ , a quantidade total de *outliers* detectados foi de 3.035. Destes, 60 foram detectados corretamente. Então, a proporção de acerto desta variação é de 0,5 e o número de falsos positivos produzidos por ela é igual a 51. Com  $p = 49\%$ , a proporção de acerto e o número de falsos positivos produzidos por ela são iguais aos de quando  $p = 49,25\%$ . No entanto, a quantidade total de *outliers* identificados quando  $p = 49\%$  é maior: 3.068.

A variação do Método do IBRE com  $p = 49,75\%$  acerta mais que o método original e ainda produz uma quantidade menor de falsos positivos. Já as variações do Método do IBRE com  $p = 49,5\%$ ,  $49,25\%$  e  $49\%$  acertam mais que o método original e produzem a mesma quantidade de falsos positivos que ele.

Todas as demais variações do Método do IBRE apresentaram proporções de acerto superiores à do método original, mas produziram uma quantidade maior de falsos positivos do que ele.

Na Tabela 7, são exibidas a quantidade total de *outliers* detectados, o número de *outliers* identificados corretamente, a proporção de acerto e o número de falsos positivos produzidos pelo

Método do IBRE e por cada uma de suas variações no município do Rio de Janeiro.

Tabela 7 Quantidade Total de *Outliers* Identificados (*Outliers* Id.), Número de *Outliers* Identificados Corretamente (*Outliers* Id. Corretamente), Proporção de Acerto (Prop. de Acerto) e Número de Falsos Positivos Gerados pelo Método do IBRE ( $p = 50\%$ ) e suas Variações ( $p = 30\%, 35\%, 40\%, 45\%, 46\%, 47\%, 48\%, 49\%, 49,25\%, 49,5\%$  e  $49,75\%$ ) no Município do Rio de Janeiro

$p$	<i>Outliers</i> Id.	<i>Outliers</i> Id. Corretamente	Prop. de Acerto	Falsos Positivos
0,5000	2.945	58	0,48333	51
0,4975	2.968	59	0,49167	50
0,4950	3.004	59	0,49167	51
0,4925	3.035	60	0,50000	51
0,4900	3.068	60	0,50000	51
0,4800	3.234	61	0,50833	53
0,4700	3.418	61	0,50833	56
0,4600	3.599	63	0,52500	57
0,4500	3.769	63	0,52500	60
0,4000	4.956	70	0,58333	71
0,3500	6.604	74	0,61667	89
0,3000	8.898	80	0,66667	111

Pode-se observar, por meio da Tabela 7, que quanto menor o valor de  $p$ , maior é a quantidade total de *outliers* detectados, o número de *outliers* identificados corretamente, a proporção de acerto dos métodos e o número de falsos positivos produzidos por eles (exceto quando  $p = 49,75\%$ ).

O mesmo comportamento pode ser observado por meio da Figura 5, onde é exibido o gráfico da proporção de acerto versus a quantidade de falsos positivos produzidos pelo Método do IBRE e por cada uma de suas variações no município do Rio de Janeiro. Os pontos referentes às variações do Método do IBRE estão todos à direita do ponto referente ao método original. Isto indica que todas as variações do Método do IBRE consideradas neste estudo apresentam uma proporção de acerto maior que a do método original. Além disso, todos os pontos referentes às variações do Método do IBRE estão acima do ponto referente ao método original, exceto quando  $p = 49,75\%$ . Isto indica que todas as variações do Método do IBRE consideradas neste estudo produzem uma quantidade maior de falsos positivos do que o método original, exceto quando  $p = 49,75\%$ .

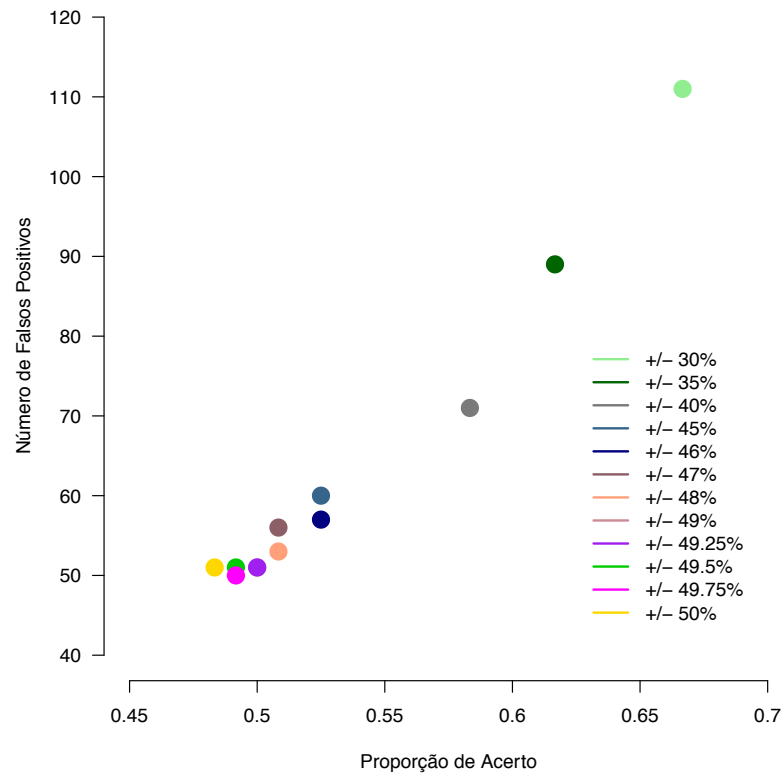


Figura 5 Proporção de Acerto  $\times$  Número de Falsos Positivos para o Método do IBRE e suas Variações no Município do Rio de Janeiro

Portanto, a variação do Método do IBRE com  $p = 49,75\%$  é considerada a vencedora dentre as demais por apresentar uma proporção de acerto maior que a do Método do IBRE com um número menor de falsos positivos.

### 5.1.2 Método do IBRE e Variações para São Paulo

Dos 55.560 relativos referentes a São Paulo, 2.119 foram classificados como *outliers* pelo Método do IBRE. Destes 2.119, 80 foram identificados corretamente. Como foram alterados, ao todo, 196 relativos pertencentes a São Paulo, a proporção de acerto do Método do IBRE neste município é dada por:  $\frac{80}{196} = 0,40816$ . Já o número de falsos positivos produzidos pelo Método do IBRE no município de São Paulo é igual a:  $\frac{2.119}{80} = 26$ .

As variações do Método do IBRE com  $p = 49,75\%$  e  $49,5\%$  detectaram, ao todo, 2.133 e 2.155 *outliers*, respectivamente, e apresentaram proporções de acerto iguais à do método original, mas produzindo um falso positivo a mais que ele.

Todas as demais variações do Método do IBRE apresentaram proporções de acerto superiores à do método original, mas também produziram uma quantidade de falsos positivos maior do

que ele.

Na Tabela 8, são exibidas a quantidade total de *outliers* detectados, o número de *outliers* identificados corretamente, a proporção de acerto e o número de falsos positivos produzidos pelo Método do IBRE e por cada uma de suas variações no município de São Paulo.

Tabela 8 Quantidade Total de *Outliers* Identificados (*Outliers* Id.), Número de *Outliers* Identificados Corretamente (*Outliers* Id. Corretamente), Proporção de Acerto (Prop. de Acerto) e Número de Falsos Positivos Gerados pelo Método do IBRE ( $p = 50\%$ ) e suas Variações ( $p = 30\%, 35\%, 40\%, 45\%, 46\%, 47\%, 48\%, 49\%, 49,25\%, 49,5\%$  e  $49,75\%$ ) no Município de São Paulo

$p$	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
0,5000	2119	80	0,40816	26
0,4975	2133	80	0,40816	27
0,4950	2155	80	0,40816	27
0,4925	2182	82	0,41837	27
0,4900	2206	82	0,41837	27
0,4800	2317	86	0,43878	27
0,4700	2444	87	0,44388	28
0,4600	2581	89	0,45408	29
0,4500	2716	90	0,45918	30
0,4000	3775	93	0,47449	41
0,3500	5042	104	0,53061	48
0,3000	7176	110	0,56122	65

Novamente, pode-se observar, por meio da Tabela 7, que quanto menor o valor de  $p$ , maiores são a quantidade total de *outliers* detectados, o número de *outliers* identificados corretamente, a proporção de acerto dos métodos e o número de falsos positivos produzidos por eles.

O mesmo comportamento pode ser observado por meio da Figura 6, onde é exibido o gráfico da proporção de acerto versus a quantidade de falsos positivos produzidos pelo Método do IBRE e por cada uma de suas variações no município de São Paulo. Os pontos referentes às variações do Método do IBRE estão todos à direita do ponto referente ao método original, exceto quando  $p = 49,75\%$  e  $49,5\%$ . Isto indica que todas as variações do Método do IBRE consideradas neste estudo apresentam uma proporção de acerto maior que a do método original, menos quando  $p = 49,75\%$  e  $49,5\%$ . No entanto, todos os pontos referentes às variações do Método do IBRE estão acima do ponto referente ao método original. Isto indica que todas as variações do Método do IBRE consideradas neste estudo produzem uma quantidade maior de falsos positivos do que o método original.



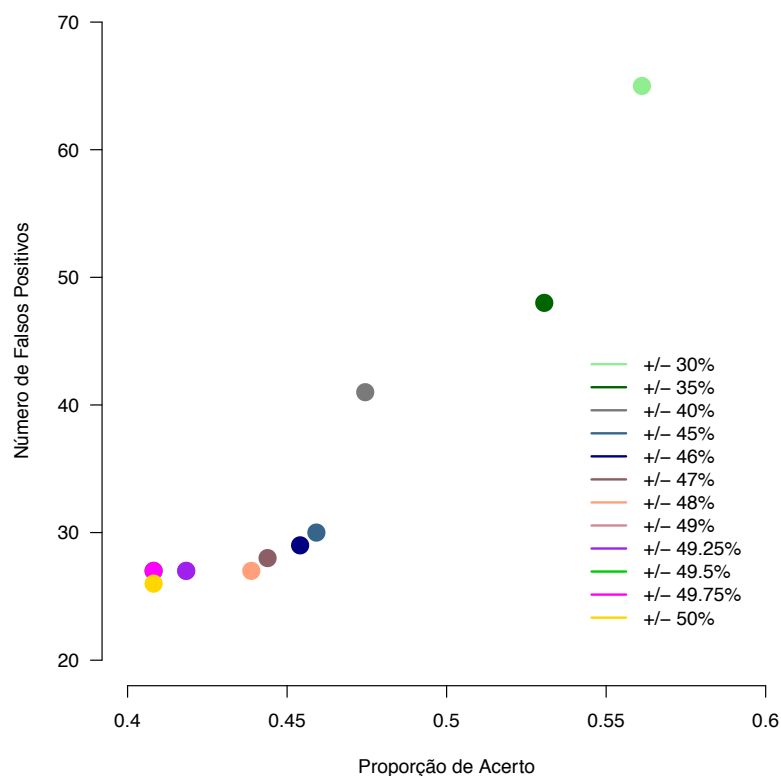


Figura 6 Proporção de Acerto  $\times$  Número de Falsos Positivos para o Método do IBRE e suas Variações no Município de São Paulo

Para o município de São Paulo, nenhuma das variações do Método do IBRE superou o próprio. A que mais se aproximou foi a com  $p = 49.75\%$ , com a mesma proporção de acerto do método original e com um falso positivo a mais. Portanto, o Método do IBRE é considerado o vencedor.

## 5.2 Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado

A seguir, são apresentados os resultados dos Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado para o município do Rio de Janeiro e de São Paulo.

Todos os três métodos foram aplicados para cada um dos cinco conjuntos de dados descritos na Seção 4. Então, aqui, a análise dos resultados é feita em duas etapas. Na primeira, para cada um dos conjuntos de dados considerados, verifica-se qual método apresentou o melhor desempenho. Na segunda, os métodos selecionados são comparados entre si a fim de se encontrar o vencedor deste segundo grupo de métodos.

### 5.2.1 Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado para o Rio de Janeiro

Na Tabela 9, são exibidas a quantidade total de *outliers* detectados, o número de *outliers* identificados corretamente, a proporção de acerto e o número de falsos positivos produzidos pelos Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado no município do Rio de Janeiro. Ao boxplot que utiliza o intervalo interquartílico na construção dos limites inferior e superior do intervalo de tolerância, dá-se o apelido de “Padrão”, ao Boxplot SIQR, de “SIQR”, e ao Boxplot Ajustado, de “Ajustado”. As siglas entre parênteses indicam o conjunto de dados ao qual os métodos foram aplicados: MR - Mês de Referência, MA - Mês Imediatamente Anterior, 2MA - 2 Meses Imediatamente Anteriores, MMAA - Mesmo Mês dos Anos Anteriores e TA - Todas os Anteriores.

Tabela 9 Quantidade Total de *Outliers* Identificados (*Outliers* Id.), Número de *Outliers* Identificados Corretamente (*Outliers* Id. Corretamente), Proporção de Acerto (Prop. de Acerto) e Número de Falsos Positivos Gerados pelos Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado no Município do Rio de Janeiro

Boxplot	Out. Id.	Out. Id. Corretamente	Prop. de Acerto	Falsos Positivos
Padrão (MR)	3.387	36	0,3000	94
SIQR (MR)	3.917	31	0,2583	126
Ajustado (MR)	4.256	29	0,2417	147
Padrão (MA)	11.102	64	0,5333	173
SIQR (MA)	12.369	64	0,5333	193
Ajustado (MA)	12.689	64	0,5333	198
Padrão (2MA)	7.748	57	0,4750	136
SIQR (2MA)	9.182	57	0,4750	161
Ajustado (2MA)	9.806	55	0,4583	178
Padrão (MMAA)	4.868	36	0,3000	135
SIQR (MMAA)	5.678	30	0,2500	189
Ajustado (MMAA)	6.379	28	0,2333	228
Padrão (TA)	3.305	48	0,4000	69
SIQR (TA)	3.570	44	0,3667	81
Ajustado (TA)	4.206	46	0,3833	91

Por meio da Tabela 9, pode-se observar que, dentre os métodos que utilizam o conjunto de dados formado pelo relativos pertencentes ao mês de referência do cálculo do índice para a construção dos limites inferior e superior de seus intervalos de tolerância, o Método do Boxplot apresentou o melhor desempenho. A quantidade total de *outliers* identificados por este método foi de 3.387, dos quais 36 foram identificados corretamente. Sendo assim, a proporção de acerto do

Método do Boxplot no Rio de Janeiro, considerando o conjunto de dados MR, é de 0,3 e o número de falsos positivos produzidos por ele neste município é igual a 94.

Tendo em vista os métodos que utilizam o conjunto de dados formado pelo relativos pertencentes ao mês imediatamente anterior ao de referência do cálculo do índice, o que apresentou o melhor desempenho foi, novamente, o Método do Boxplot. A quantidade total de *outliers* identificados por este método foi de 11.102, dos quais 64 foram identificados corretamente. Sendo assim, a proporção de acerto do Método do Boxplot no Rio de Janeiro, considerando o conjunto de dados MA, é de 0,53333 e o número de falsos positivos produzidos por ele neste município é igual a 173.

Dentre os métodos que utilizam o conjunto de dados formado pelo relativos pertencentes aos dois meses imediatamente anteriores ao de referência do cálculo do índice, o Método do Boxplot apresentou, mais uma vez, o melhor desempenho. A quantidade total de *outliers* identificados por este método foi de 7.748, dos quais 57 foram identificados corretamente. Sendo assim, a proporção de acerto do Método do Boxplot no Rio de Janeiro, considerando o conjunto de dados 2MA, é de 0,475 e o número de falsos positivos produzidos por ele neste município é igual a 136.

Ao considerar os métodos que utilizam o conjunto de dados formado pelo relativos pertencentes ao mesmo mês dos anos anteriores, o Método do Boxplot apresentou, de novo, o melhor desempenho. A quantidade total de *outliers* identificados por este método foi de 4.868, dos quais 36 foram identificados corretamente. Sendo assim, a proporção de acerto do Método do Boxplot no Rio de Janeiro, considerando o conjunto de dados MMAA, é de 0,3 e o número de falsos positivos produzidos por ele neste município é igual a 135.

Para finalizar, entre os métodos que utilizam o conjunto de dados formado pelo relativos pertencentes a todos os meses dos anos anteriores, o método que apresentou o melhor desempenho foi, outra vez, o Método do Boxplot. A quantidade total de *outliers* identificados por este método foi de 3.305, dos quais 48 foram identificados corretamente. Sendo assim, a proporção de acerto do Método do Boxplot no Rio de Janeiro, considerando o conjunto de dados TA, é de 0,4 e o número de falsos positivos produzidos por ele neste município é igual a 69.

Resumindo, independente do conjunto de dados ao qual foi aplicado, o Método do Boxplot apresentou o melhor desempenho dentre os Métodos do Boxplot SIQR e do Boxplot Ajustado. A Tabela 9 é, então, resumida na Tabela 10, onde são exibidas a quantidade total de *outliers* detectados, o número de *outliers* identificados corretamente, a proporção de acerto e o número de falsos positivos produzidos pelo Método do Boxplot quando aplicado a cada um dos conjuntos de dados considerados neste estudo.

Na Tabela 10, os métodos foram ordenados em ordem decrescente pela proporção de acerto e pelo número de falsos positivos produzidos em ordem crescente. Desta forma, fica claro que o Método do Boxplot quando aplicado ao conjunto de dados TA apresenta um desempenho melhor do que quando aplicado aos conjuntos de dados MR e MMAA. Isto porque, quando aplicado a

estes conjuntos de dados, o Método do Boxplot apresenta proporções de acerto menores e números de falsos positivos maiores do que quando aplicado ao conjunto de dados TA. Quando aplicado aos conjuntos de dados MA e 2MA, o Método do Boxplot apresenta proporções de acerto maiores do que quando aplicado ao conjunto de dados TA, mas produz quantidades bem maiores de falsos positivos que não compensam o acréscimo nas proporções.

Tabela 10 Quantidade Total de *Outliers* Identificados (*Outliers* Id.), Número de *Outliers* Identificados Corretamente (*Outliers* Id. Corretamente), Proporção de Acerto (Prop. de Acerto) e Número de Falsos Positivos Gerados pelo Método do Boxplot quando Aplicado aos Conjuntos de Dados MR, MA, 2MA, TA e MMAA no Município do Rio de Janeiro

Boxplot	Out. Id.	Out. Id. Corretamente	Prop. de Acerto	Falsos Positivos
Padrão (MA)	11.102	64	0,5333	173
Padrão (2MA)	7.748	57	0,4750	136
Padrão (TA)	3.305	48	0,4000	69
Padrão (MR)	3.387	36	0,3000	94
Padrão (MMAA)	4.868	36	0,3000	135

Na Figura 7, é exibido o gráfico da proporção de acerto versus a quantidade de falsos positivos produzidos pelos Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado, quando aplicados a cada um dos conjuntos de dados considerados neste trabalho, para o município do Rio de Janeiro.

Ao observá-la, pode-se chegar na mesma conclusão. O desempenho de todos os métodos cujos pontos estão à esquerda do referente ao Método do Boxplot aplicado ao conjunto de dados TA são inferiores ao dele. Isto porque tais métodos, além de apresentarem proporções de acerto inferiores à do Método do Boxplot aplicado ao conjunto de dados TA, produzem mais falsos positivos do que ele. Já os métodos cujos pontos estão à direita do ponto referente ao Método do Boxplot aplicado ao conjunto de dados TA possuem proporções de acerto superiores à dele, mas produzem uma quantidade muito maior de falsos positivos que não compensam o acréscimo na proporção. Portanto, o Método do Boxplot aplicado ao conjunto de dados TA é considerado o vencedor dentre o Método do Boxplot aplicado aos conjuntos de dados MR, MA, 2MA e MMAA e os Métodos do Boxplot SIQR e Boxplot Ajustado aplicados aos conjuntos de dados MR, MA, 2MA, MMAA e TA.

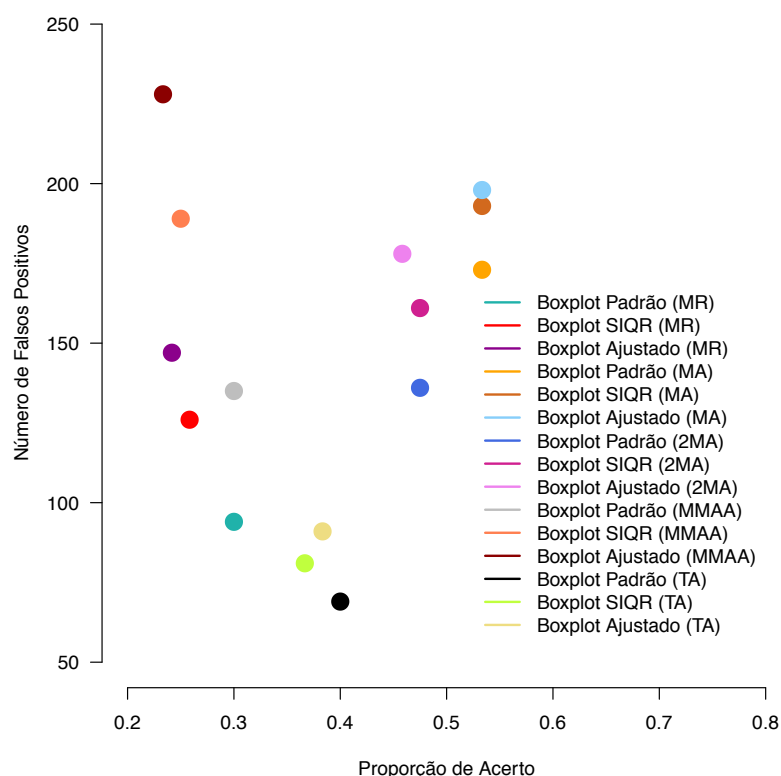


Figura 7 Proporção de Acerto x Número de Falsos Positivos para os Métodos do Boxplot, Boxplot SIQR e Boxplot Ajustado para o Município do Rio de Janeiro

### 5.2.2 Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado para São Paulo

Na Tabela 11, são exibidas a quantidade total de *outliers* detectados, o número de *outliers* identificados corretamente, a proporção de acerto e o número de falsos positivos produzidos pelos Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado no município de São Paulo.

Por meio da Tabela 11, pode-se observar que, dentre os métodos que utilizam o conjunto de dados formado pelo relativos pertencentes ao mês de referência do cálculo do índice para a construção dos limites inferior e superior de seus intervalos de tolerância, o Método do Boxplot apresentou o melhor desempenho. A quantidade total de *outliers* identificados por este método foi de 5.015, dos quais 48 foram identificados corretamente. Sendo assim, a proporção de acerto do Método do Boxplot em São Paulo, considerando o conjunto de dados MR, é de 0,2449 e o número de falsos positivos produzidos por ele neste município é igual a 104.

Tabela 11 Quantidade Total de *Outliers* Identificados (*Outliers* Id.), Número de *Outliers* Identificados Corretamente (*Outliers* Id. Corretamente), Proporção de Acerto (Prop. de Acerto) e Número de Falsos Positivos Gerados pelos Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado no Município de São Paulo

Boxplot	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
Padrão (MR)	5.015	48	0,2449	104
SIQR (MR)	5.911	47	0,2398	126
Ajustado (MR)	5.679	30	0,1530	189
Padrão (MA)	10.628	112	0,5714	95
SIQR (MA)	12.657	104	0,5306	122
Ajustado (MA)	12.436	100	0,5102	124
Padrão (2MA)	9.518	91	0,4642	105
SIQR (2MA)	11.598	101	0,5153	115
Ajustado (2MA)	11.297	88	0,4489	128
Padrão (MMAA)	5.044	68	0,3469	74
SIQR (MMAA)	6.542	70	0,3571	93
Ajustado (MMAA)	7.038	60	0,3061	117
Padrão (TA)	3.673	81	0,4132	45
SIQR (TA)	4.102	79	0,4030	52
Ajustado (TA)	4.905	81	0,4132	61

Ao considerar os métodos que utilizam o conjunto de dados formado pelo relativos pertencentes ao mês imediatamente anterior ao de referência do cálculo do índice, o método que apresentou o melhor desempenho foi, novamente, o Método do Boxplot. A quantidade total de *outliers* identificados por este método foi de 10.628, dos quais 112 foram identificados corretamente. Sendo assim, a proporção de acerto do Método do Boxplot em São Paulo, considerando o conjunto de dados MA, é de 0,5714 e o número de falsos positivos produzidos por ele neste município é igual a 95.

Levando em conta os métodos que utilizam o conjunto de dados formado pelo relativos pertencentes aos dois meses imediatamente anteriores ao de referência do cálculo do índice, o Método do Boxplot apresentou, mais uma vez, o melhor desempenho. A quantidade total de *outliers* identificados por este método foi de 9.518, dos quais 91 foram identificados corretamente. Sendo assim, a proporção de acerto do Método do Boxplot em São Paulo, considerando o conjunto de dados 2MA, é de 0,4642 e o número de falsos positivos produzidos por ele neste município é igual a 105. O Método do Boxplot SIQR detectou, ao todo, 11.598 *outliers*, dos quais 101 foram detectados corretamente. Embora este método tenha detectado corretamente dez *outliers* a mais que o Método

do Boxplot, ele também produziu dez falsos positivos a mais que o Método do Boxplot. Portanto, o Método do Boxplot foi considerado o melhor dentre estes dois.

Em meio aos métodos que utilizam o conjunto de dados formado pelo relativos pertencentes ao mesmo mês dos anos anteriores, o Método do Boxplot apresentou, de novo, o melhor desempenho. A quantidade total de *outliers* identificados por este método foi de 5.044, dos quais 68 foram identificados corretamente. Sendo assim, a proporção de acerto do Método do Boxplot em São Paulo, considerando o conjunto de dados MMAA, é de 0,3469 e o número de falsos positivos produzidos por ele neste município é igual a 74.

Para finalizar, dentre os métodos que utilizam o conjunto de dados formado pelo relativos pertencentes a todos os meses dos anos anteriores, o método que apresentou o melhor desempenho foi outra vez o Método do Boxplot. A quantidade total de *outliers* identificados por este método foi de 3.673, dos quais 81 foram identificados corretamente. Sendo assim, a proporção de acerto do Método do Boxplot em São Paulo, considerando o conjunto de dados TA, é de 0,41327 e o número de falsos positivos produzidos por ele neste município é igual a 45.

Resumindo, independente do conjunto de dados ao qual foi aplicado, o Método do Boxplot apresentou o melhor desempenho dentre os Métodos do Boxplot SIQR e do Boxplot Ajustado. A Tabela 11 é, então, resumida na Tabela 12, onde são exibidas a quantidade total de *outliers* detectados, o número de *outliers* identificados corretamente, a proporção de acerto e o número de falsos positivos produzidos pelo Método do Boxplot quando aplicado a cada um dos conjuntos de dados considerados neste estudo.

Na Tabela 12, os métodos foram ordenados em ordem decrescente pela proporção de acerto. Desta forma, fica claro que o Método do Boxplot quando aplicado ao conjunto de dados TA apresenta um desempenho melhor do que quando aplicado aos conjuntos de dados MR e MMAA. Isto porque, quando aplicado a estes conjuntos de dados, o Método do Boxplot apresenta proporções de acerto menores e números de falsos positivos maiores do que quando aplicado ao conjunto de dados TA. Quando aplicado aos conjuntos de dados MA e 2MA, o Método do Boxplot apresenta proporções de acerto maiores do que quando aplicado ao conjunto de dados TA, mas produz quantidades bem maiores de falsos positivos que não compensam o acréscimo nas proporções.

Tabela 12 Quantidade Total de *Outliers* Identificados (*Outliers* Id.), Número de *Outliers* Identificados Corretamente (*Outliers* Id. Corretamente), Proporção de Acerto (Prop. de Acerto) e Número de Falsos Positivos Gerados pelo Método do Boxplot quando Aplicado aos Conjuntos de Dados MR, MA, 2MA, TA e MMAA no Município de São Paulo

Boxplot	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
Padrão (MA)	10.628	112	0,5714	95
Padrão (2MA)	9.518	91	0,4642	105
Padrão (TA)	3.673	81	0,4132	45
Padrão (MMAA)	5.044	68	0,3469	74
Padrão (MR)	5.015	48	0,2449	104

Na Figura 8, é exibido o gráfico da proporção de acerto versus a quantidade de falsos positivos produzidos pelos Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado, quando aplicados a cada um dos conjuntos de dados considerados neste trabalho, para o município de São Paulo.

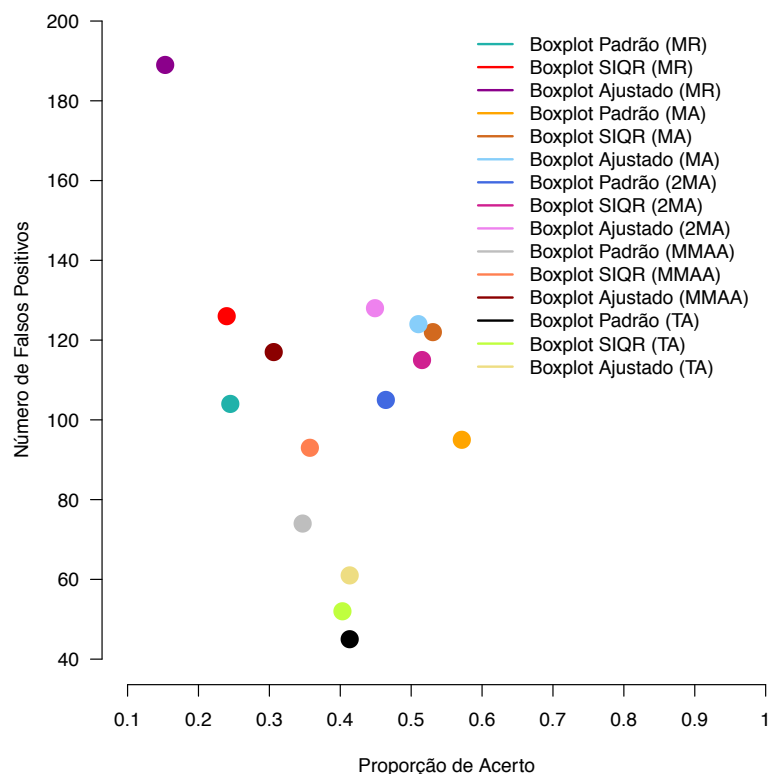


Figura 8 Proporção de Acerto  $\times$  Número de Falsos Positivos para os Métodos do Boxplot, Boxplot SIQR e Boxplot Ajustado para o Município de São Paulo

Ao observá-la, pode-se chegar à conclusão análoga à do Rio de Janeiro. O desempenho



de todos os métodos cujos pontos estão à esquerda do referente ao Método do Boxplot aplicado ao conjunto de dados TA são inferiores ao dele. Isto porque tais métodos, além de apresentarem proporções de acerto inferiores à do Método do Boxplot aplicado ao conjunto de dados TA, produzem mais falsos positivos do que ele. Já os métodos cujos pontos estão à direita do ponto referente ao Método do Boxplot aplicado ao conjunto de dados TA possuem proporções de acerto superiores à dele, mas produzem uma quantidade muito maior de falsos positivos que não compensam o acréscimo na proporção. Portanto, o Método do Boxplot aplicado ao conjunto de dados TA é considerado o melhor dentre o Método do Boxplot aplicado aos conjuntos de dados MR, MA, 2MA e MMAA e os Métodos do Boxplot SIQR e Boxplot Ajustado aplicados aos conjuntos de dados MR, MA, 2MA, MMAA e TA.

### **5.3 Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e do Algoritmo de Tukey**

A seguir, são apresentados os resultados dos Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e do Algoritmo de Tukey para o município do Rio de Janeiro e de São Paulo.

Assim como os métodos do segundo grupo, os limites inferior e superior dos intervalos de tolerância dos métodos do terceiro grupo são calculados com base nos conjuntos de dados descritos na Seção 4. Sendo assim, para cada um dos métodos em estudo, é construído um gráfico da proporção de acerto contra o número de falsos positivos gerados pelo método em questão quando aplicado a cada um dos conjuntos de dados considerados.

Além disso, os métodos são aplicados, supondo simetria, considerando  $c_L = c_U = c$ , com  $c = \{0; 0,1; 0,2; \dots, 3,8; 3,9; 4\}$ . Para cada um dos quarenta e um valores considerados para  $c$ , tem-se um intervalo de tolerância. Ou seja, para um mesmo método, são gerados quarenta e um intervalos de tolerância, cada um deles com uma proporção de acerto e um número de falsos positivos produzido. Portanto, aqui, ao invés de ser representado por um único ponto, o desempenho do método é representado por uma curva.

Novamente, os resultados são analisados em duas etapas. Na primeira, constata-se em qual conjunto de dados os métodos apresentam o melhor desempenho. Isto é feito por meio da análise das curvas geradas pela aplicação do método em cada um dos conjuntos de dados considerados, segundo o seguinte critério: aquela que permanecer abaixo das demais para a maioria dos valores do eixo  $x$  é considerada a vencedora, pois apresenta as mesmas proporções de acerto que as demais e quantidades menores de falsos positivos. Na segunda etapa, as curvas vencedoras são comparadas entre si, de forma a eleger o melhor método dentre o de Cercas Resistentes, o do Quartil, o do Quartil Modificado, o do Desvio Mediano Absoluto e o Algoritmo de Tukey. Por fim, são revelados

os valores de  $c$  responsáveis pelo desempenho apresentado pelo método eleito.

Aqui, todos os resultados são exibidos por meio de figuras. No Apêndice C, estão disponíveis as tabelas com os resultados numéricos usadas para gerar os gráficos a seguir.

### **5.3.1 Métodos do Quartil, do Quartil Modificado, de Cercas Resistentes, do Desvio Mediano Absoluto e do Algoritmo de Tukey para o Rio de Janeiro**

Nas Figuras 9, 10, 11, 12 e 13, são exibidos, respectivamente, os gráficos das proporções de acerto contra os números de falsos positivos gerados para o município do Rio de Janeiro pelos Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e do Algoritmo de Tukey quando aplicados aos conjuntos de dados MR, MA, 2MA, MMAA e TA, para cada um dos valores de  $c$  considerados neste estudo.

Ao observá-las, percebe-se que o desempenho de todos os métodos no município do Rio de Janeiro é melhor quando os mesmos são aplicados ao conjunto de dados TA. Isto é, ao utilizar todos os relativos anteriores aos do mês de referência do cálculo do índice na construção dos limites inferior e superior de seus intervalos de tolerância, todos os métodos produzem quantidades menores de falsos positivos, independente da proporção de acerto gerada.

Este comportamento é claro para os Métodos de Cercas Resistentes (Figura 9), do Quartil Modificado (Figura 11), do Desvio Mediano Absoluto (Figura 12) e para o Algoritmo de Tukey (Figura 13). Já o Método do Quartil (Figura 10) também apresenta um bom desempenho quando aplicado aos conjuntos de dados MA e 2MA. Ou seja, ao fazer uso dos relativos dos meses imediatamente anterior ou dos relativos dos dois meses imediatamente anteriores ao mês de referência do cálculo do índice, o Método do Quartil apresenta um desempenho tão bom quanto o que apresenta ao ser aplicado ao conjunto de dados TA. No entanto, ao considerar valores mais elevados para a proporção de acerto, entre 0,7 e 1, o Método do Quartil gera as menores quantidades de falsos positivos quando aplicado ao conjunto de dados TA.

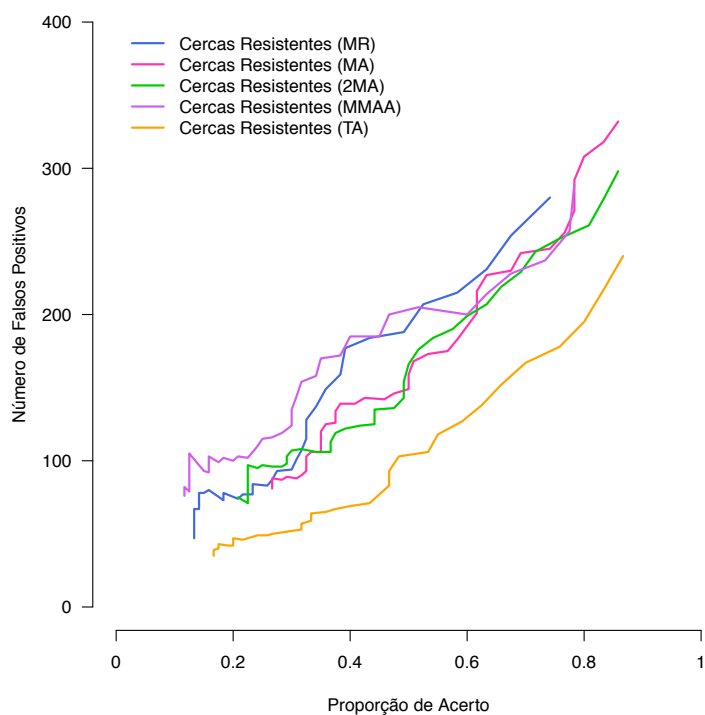


Figura 9 Desempenho do Método de Cercas Resistentes no Município do Rio de Janeiro.

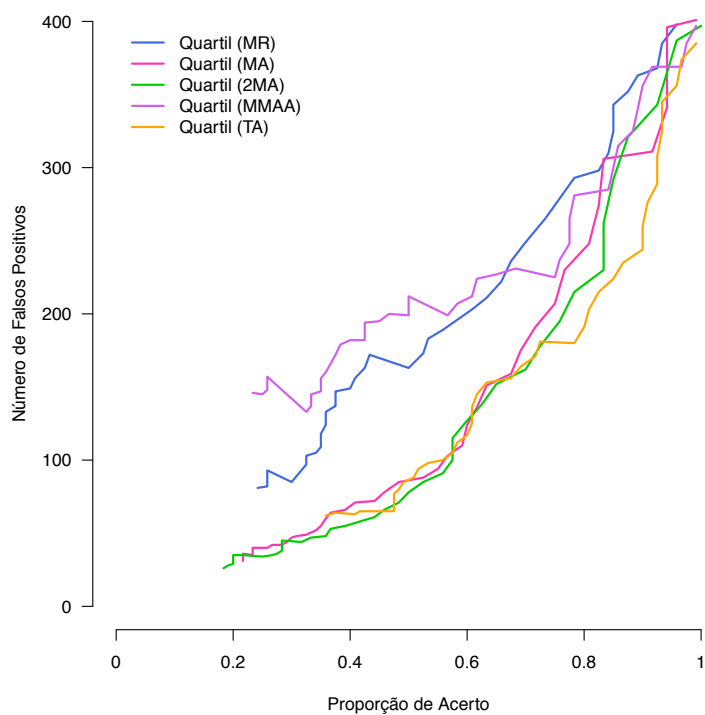


Figura 10 Desempenho do Método do Quartil no Município do Rio de Janeiro.

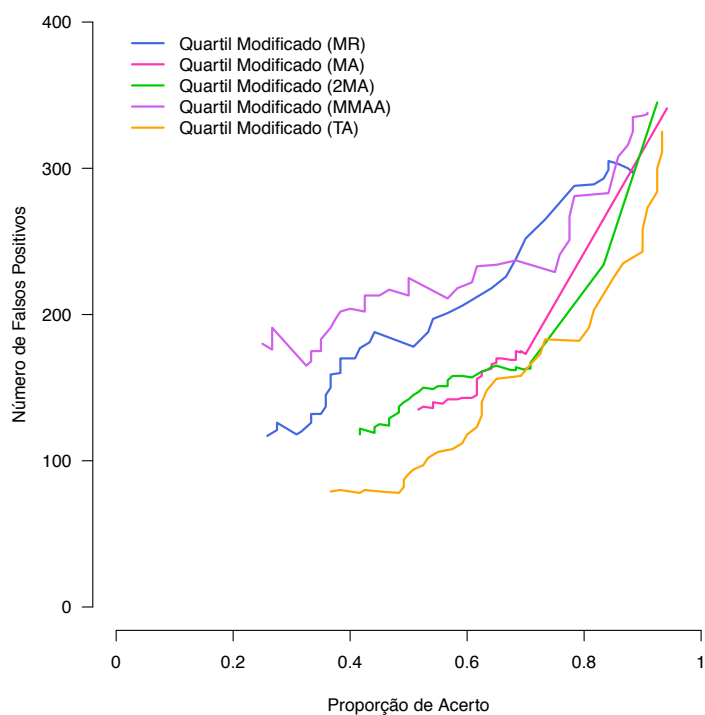


Figura 11 Desempenho do Método do Quartil Modificado no Município do Rio de Janeiro.

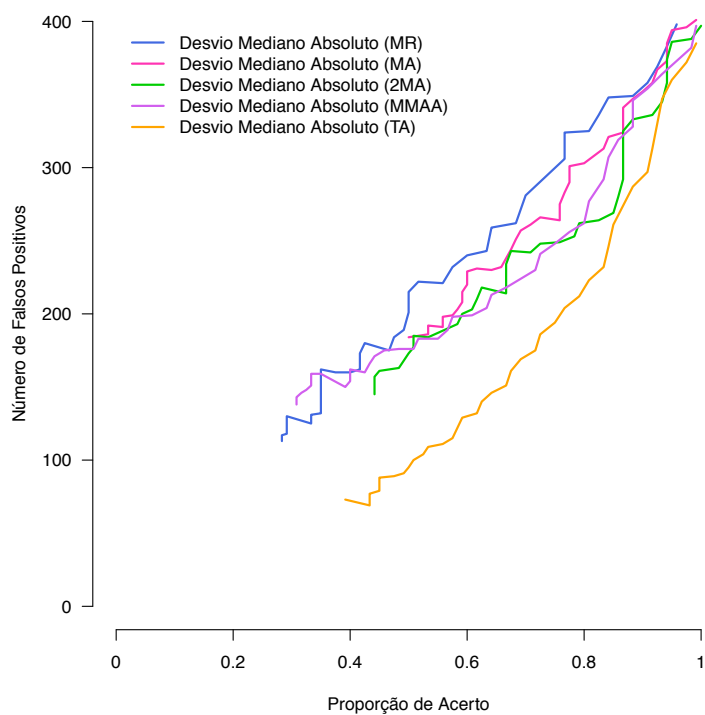


Figura 12 Desempenho do Método do Desvio Mediano Absoluto no Município do Rio de Janeiro.

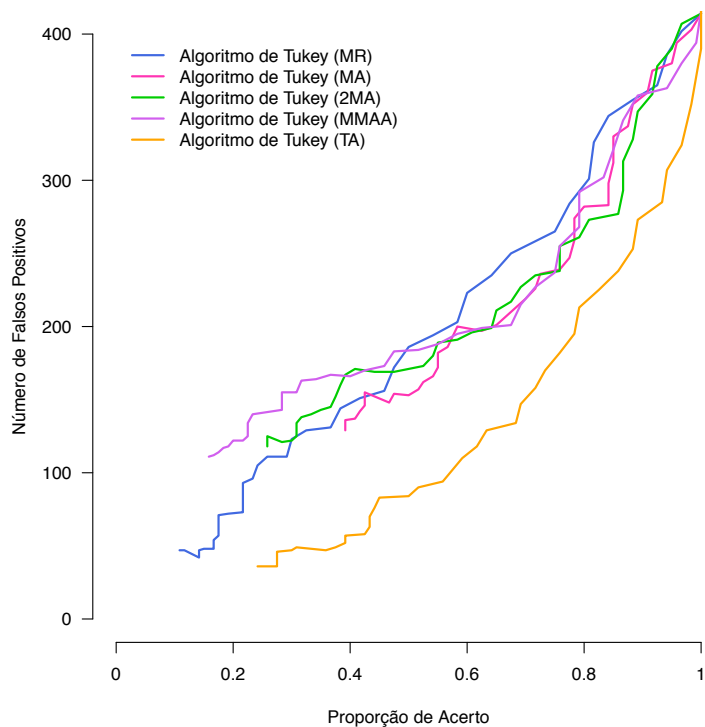


Figura 13 Desempenho do Algoritmo de Tukey no Município do Rio de Janeiro.

Na Figura 14, é exibido o gráfico das proporções de acerto contra os números de falsos positivos gerados para o município do Rio de Janeiro pelos Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e do Algoritmo de Tukey quando aplicados ao conjunto de dados vencedor: o TA.

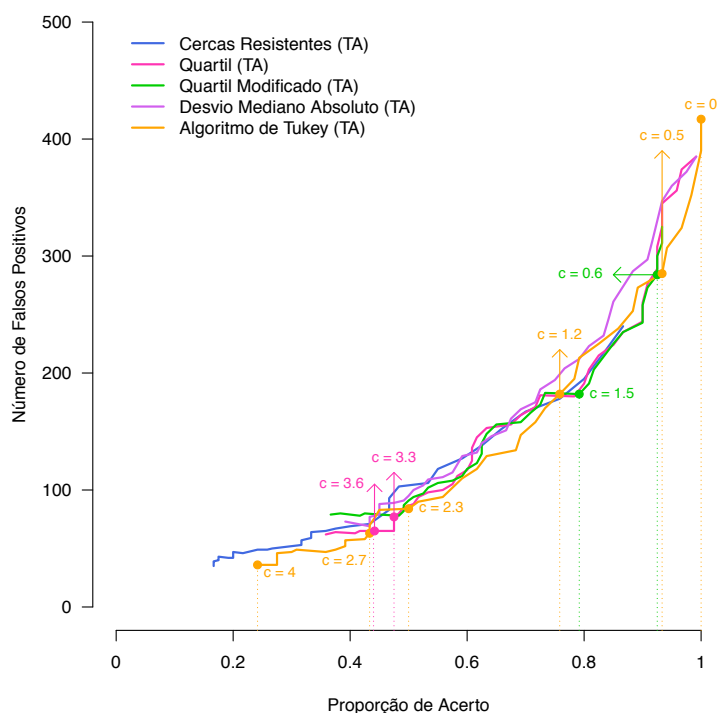


Figura 14 Desempenho dos Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e do Algoritmo de Tukey no município do Rio de Janeiro quando aplicados ao conjunto de dados TA.

Ao analisar o comportamento das curvas, observa-se que o método ideal para gerar proporções de acerto na faixa de 0,24 a 0,43 é o Algoritmo de Tukey. Ao considerar para  $c$  valores entre 2,7 e 4, este método alcança proporções de acerto no intervalo mencionado produzindo quantidades menores de falsos positivos que os demais métodos.

Para gerar proporções de acerto na faixa de 0,44 a 0,49, o Método do Quartil é considerado o mais adequado. Ao utilizar para  $c$  valores entre 3,1 e 3,6, este método atinge proporções de acerto no intervalo mencionado fornecendo quantidades menores de falsos positivos que os outros métodos.

De forma a produzir proporções de acerto na faixa de 0,50 a 0,75, o Algoritmo de Tukey é o método mais apropriado. Ao considerar para  $c$  valores entre 1,2 e 2,3, este método alcança proporções no intervalo mencionado produzindo quantidades menores de falsos positivos que os demais métodos.

Para obter proporções de acerto na faixa de 0,79 a 0,92, o Método do Quartil Modificado é considerado o melhor. Ao utilizar para  $c$  valores entre 0,6 e 1,5, este método atinge proporções de acerto no intervalo mencionado fornecendo quantidades menores de falsos positivos.

Por fim, de maneira a conceber proporções de acerto na faixa de 0,93 a 1, o Algoritmo de Tukey é o método ideal. Ao utilizar para  $c$  valores entre 0 e 0,5, este método alcança proporções de

acerto no intervalo mencionado produzindo quantidades menores de falsos positivos que os demais métodos.

Os resultados da análise do desempenho dos métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Aboluto e do Algoritmo de Tukey no município do Rio de Janeiro quando aplicados ao conjunto de dados TA são resumidos na tabela abaixo.

**Tabela 13** Resultados da Análise do Desempenho dos Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Aboluto e do Algoritmo de Tukey no Município do Rio de Janeiro quando Aplicados ao Conjunto de Dados TA

Proporção de Acerto	Constante $c$	Método Escolhido
0,24 a 0,43	2,7 a 4,0	Algoritmo de Tukey
0,44 a 0,49	3,1 a 3,6	Método do Quartil
0,50 a 0,75	1,2 a 2,3	Algoritmo de Tukey
0,79 a 0,92	0,6 a 1,5	Método do Quartil Modificado
0,93 a 1,00	0,0 a 1,0	Algoritmo de Tukey

Então, conforme Tabela 13, a curva referente ao Algoritmo de Tukey é que permanece abaixo das demais para a maioria das proporções de acerto observadas. Portanto, conclui-se que, para o município do Rio de Janeiro, este método é o melhor dentre os Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado e do Desvio Mediano Absoluto.

### **5.3.2 Métodos do Quartil, do Quartil Modificado, de Cercas Resistentes, do Desvio Mediano Absoluto e do Algoritmo de Tukey para São Paulo**

Nas Figuras 15, 16, 17, 18 e 19, são exibidos, respectivamente, os gráficos das proporções de acerto contra os números de falsos positivos gerados para o município de São Paulo pelos Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e do Algoritmo de Tukey quando aplicados aos conjuntos de dados MR, MA, 2MA, MMAA e TA, para cada um dos valores de  $c$  considerados neste estudo.

Ao examiná-las, verifica-se que o desempenho de todos os métodos no município de São Paulo também é melhor quando os mesmos são aplicados ao conjunto de dados TA.

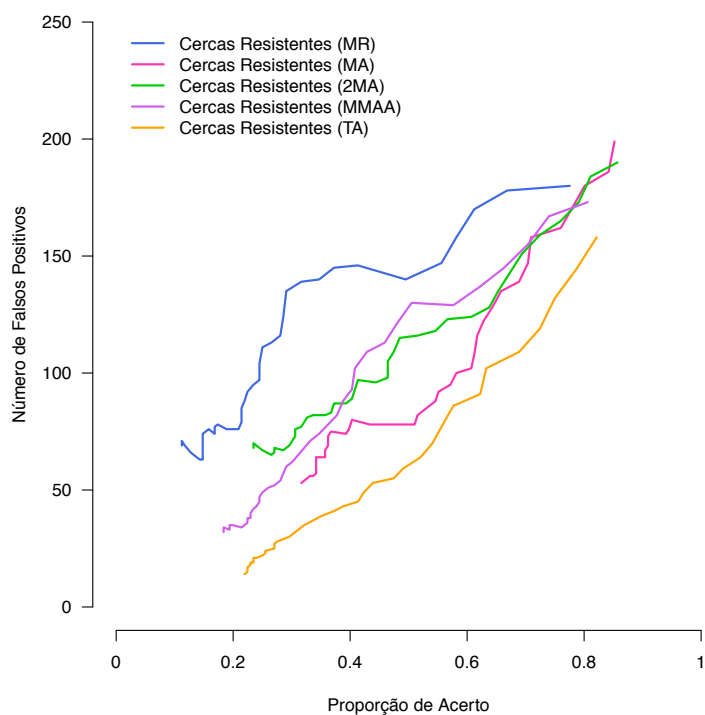


Figura 15 Desempenho do Método de Cercas Resistentes no Município de São Paulo.

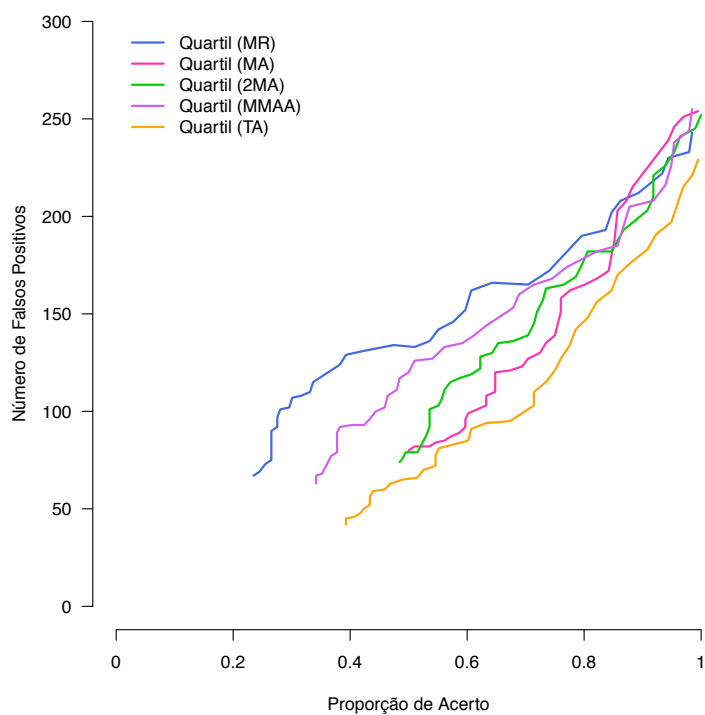


Figura 16 Desempenho do Método do Quartil no Município de São Paulo.



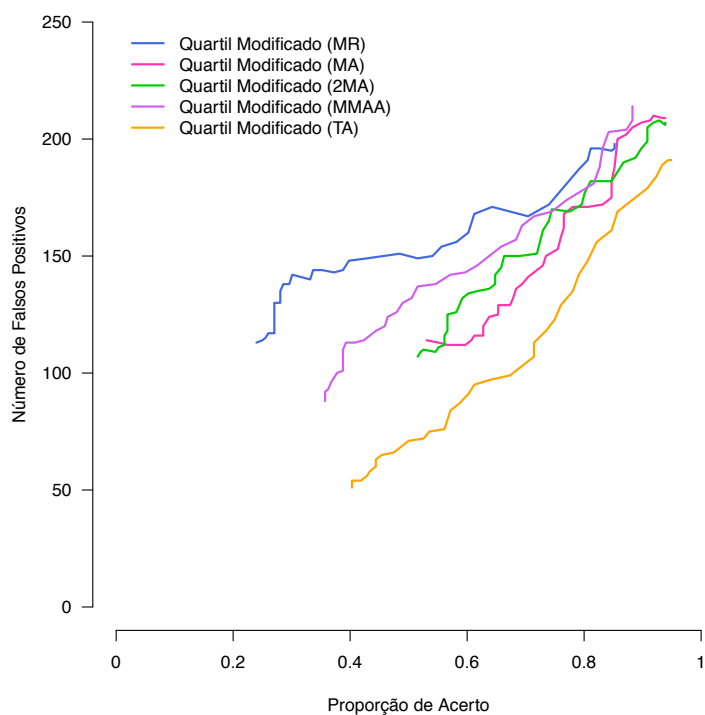


Figura 17 Desempenho do Método do Quartil Modificado no Município de São Paulo.

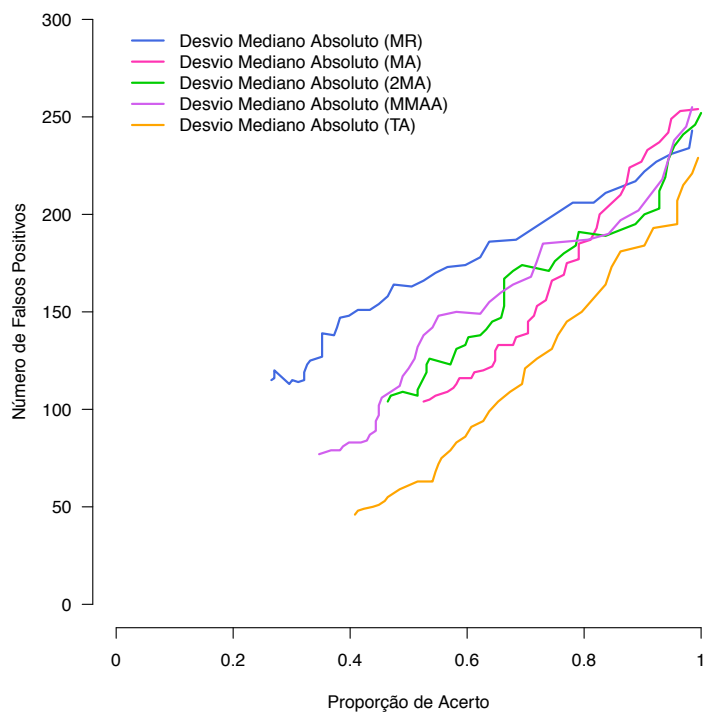


Figura 18 Desempenho do Método do Desvio Mediano Absoluto no Município de São Paulo.

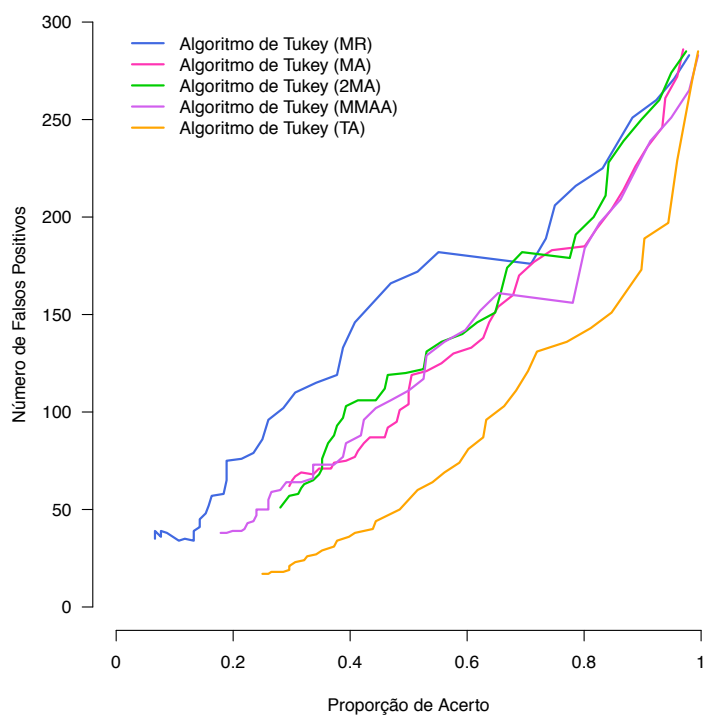


Figura 19 Desempenho do Algoritmo de Tukey no Município de São Paulo.

Na Figura 20, é exibido o gráfico das proporções de acerto contra os números de falsos positivos gerados para o município de São Paulo pelos Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e do Algoritmo de Tukey quando aplicados ao conjunto de dados vencedor: o TA.

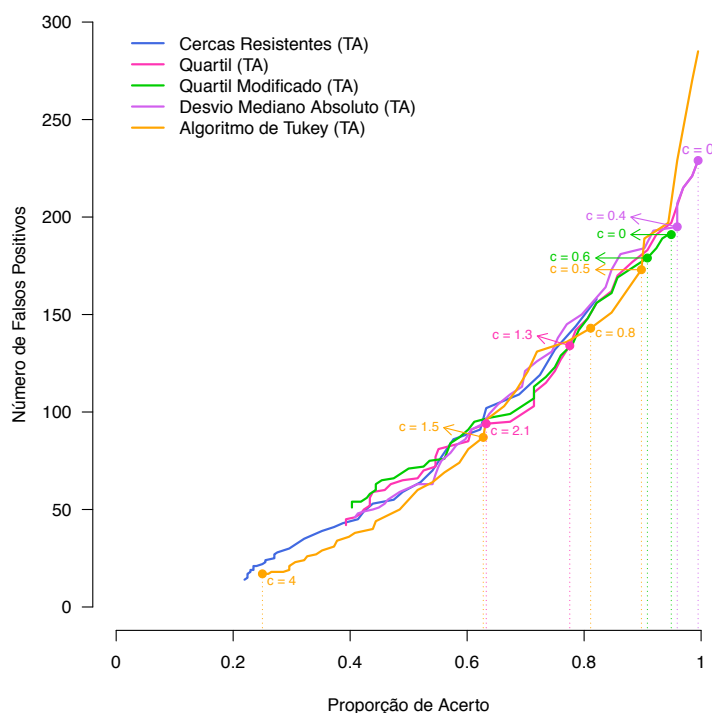


Figura 20 Desempenho dos Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e do Algoritmo de Tukey no município de São Paulo quando aplicados ao conjunto de dados TA.

Ao analisar o comportamento das curvas, observa-se que o método ideal para gerar proporções de acerto na faixa de 0,25 a 0,62 é o Algoritmo de Tukey. Ao considerar para  $c$  valores entre 1,5 e 4, este método alcança proporções de acerto no intervalo mencionado produzindo quantidades menores de falsos positivos que os demais métodos.

Para fornecer proporções de acerto na faixa de 0,63 a 0,78, o Método do Quartil é considerado o mais adequado. ao utilizar para  $c$  valores entre 1,3 e 2,1, este método atinge proporções de acerto no intervalo mencionado fornecendo quantidades menores de falsos positivos que os outros métodos.

De forma a adquirir proporções de acerto na faixa de 0,81 a 0,89, o Algoritmo de Tukey é o método mais apropriado. Ao considerar para  $c$  valores entre 0,5 e 0,8, este método alcança proporções no intervalo mencionado produzindo quantidades menores de falsos positivos que os demais métodos.

Para gerar proporções de acerto na faixa de 0,90 a 0,94, o Método do Quartil Modificado é considerado o melhor. Ao utilizar para  $c$  valores entre 0 e 0,6, este método atinge proporções de acerto no intervalo mencionado fornecendo quantidades menores de falsos positivos.

Por fim, para conquistar proporções de acerto na faixa de 0,95 a 0,99, o Método do Desvio Mediano Absoluto é o ideal. Ao utilizar para  $c$  valores entre 0 e 0,4, este método alcança proporções

de acerto no intervalo mencionado produzindo quantidades menores de falsos positivos que os demais métodos.

Os resultados da análise do desempenho dos métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e do Algoritmo de Tukey no município de São Paulo quando aplicados ao conjunto de dados TA são resumidos na tabela abaixo.

Tabela 14 Resultados da Análise do Desempenho dos Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e do Algoritmo de Tukey no Município de São Paulo quando Aplicados ao Conjunto de Dados TA

Proporção de Acerto	Constante $c$	Método Escolhido
0,25 a 0,62	1,5 a 4,0	Algoritmo de Tukey
0,63 a 0,78	1,3 a 2,1	Método do Quartil
0,81 a 0,89	0,5 a 0,8	Algoritmo de Tukey
0,90 a 0,94	0,0 a 0,6	Método do Quartil Modificado
0,95 a 1,00	0,0 a 0,4	Método do Desvio Mediano Absoluto

Sendo assim, conforme Tabela 14, a curva referente ao Algoritmo de Tukey é que permanece abaixo das demais para a maioria das proporções e acerto observadas. Portanto, conclui-se que, para o município de São Paulo, este método é o melhor dentre os Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado e do Desvio Mediano Absoluto.

## 5.4 Resultado Geral

A seguir, é construído um ranking referente ao desempenho de cada um dos métodos aplicados para o município do Rio de Janeiro e outro para São Paulo.

### 5.4.1 Resultado Geral para o Rio de Janeiro

Em terceiro lugar, com proporção de acerto igual a 0,4 e 69 falsos positivos produzidos, ficou o Método do Boxplot aplicado ao conjunto de dados TA.

Em segundo lugar, com proporção de acerto igual a 0,4333 e 63 falsos positivos produzidos, ficou o Algoritmo de Tukey aplicado ao conjunto de dados TA, com  $c = 2,7$ .

Em primeiro lugar, com proporção de acerto igual a 0,4916 e 50 falsos positivos produzidos, ficou a variação do Método do IBRE com  $p = 49,75\%$ .

### 5.4.2 Resultado Geral para São Paulo

Em terceiro lugar, com proporção de acerto igual a 0,4132 e 45 falsos positivos produzidos, ficou o Método do Boxplot aplicado ao conjunto de dados TA.

Em segundo lugar, com proporção de acerto igual a 0,4438 e 44 falsos positivos produzidos, ficou o Algoritmo de Tukey aplicado ao conjunto de dados TA, com  $c = 2,4$ .

Em primeiro lugar, com proporção de acerto igual a 0,4081 e 26 falsos positivos produzidos, ficou o método utilizado atualmente no IBRE, o qual considera parâmetros fixos de crítica iguais a  $\pm 50\%$ .

## 6 CONCLUSÕES

Neste trabalho, é feita uma comparação entre os seguintes métodos para detecção de *ouliers*: Método do IBRE, variações do Método do IBRE, Método do Boxplot, Método do Boxplot SIQR, Método do Boxplot Ajustado, Método de Cercas Resistentes, Método do Quartil, Método do Quartil Modificado, Método do Desvio Mediano Absoluto e Algoritmo de Tukey. Esta comparação é feita por meio da porcentagem de acerto dos métodos e do número de falsos positivos produzidos pelos mesmos. Para tal, os preços descartados ou alterados pelos analistas são considerados como sendo os verdadeiros *ouliers*.

A base de dados utilizada neste estudo é composta pelos preços coletados para os insumos dos subitens pertencentes aos itens “Hortaliças e Legumes” e “Frutas”, que fazem parte do subgrupo “Gêneros Alimentícios” que, por sua vez, está dentro do grupo “Alimentação”, integrantes da estrutura da amostra de bens e serviços do IPC. A coleta dos preços foi realizada no período compreendido entre janeiro de 2004 e dezembro de 2011, nos municípios Rio de Janeiro e São Paulo. Nesta base de dados, nenhum preço foi descartado e 316 relativos foram alterados por conta de modificações feitas pelos analistas em um ou mais preços que o formam, sendo 120 pertencentes ao Rio de Janeiro e 196 pertencentes a São Paulo.

Neste estudo, foi feita também uma análise do conjunto de dados aos quais os métodos do Boxplot, do Boxplot SIQR, do Boxplot Ajustado, de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e o Algoritmo de Tukey foram aplicados. Constatou-se que todos os métodos obtiveram um melhor desempenho quando aplicados ao conjunto de dados formado por todos os relativos de preços anteriores aos do mês de referência do cálculo do índice, denominado “Todos os Anteriores” e representado pela sigla “TA”. Este resultado é razoável pois o conjunto de dados TA contém mais informações, isto é, o desempenho dos métodos foi melhor porque tal conjunto é mais rico.

Para serem analisados, os métodos foram divididos em três grupos. O primeiro é formado pelo Método do IBRE e por suas variações. O segundo é constituído pelos Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado. E o terceiro é integrado pelos Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e do Algoritmo de Tukey.

Para cada grupo, um método foi eleito o vencedor por apresentar o melhor desempenho dentre os demais. Em seguida, os métodos vencedores foram comparados entre si. Esta comparação deu origem a um ranking.

Para o município do Rio de Janeiro, a variação do Método do IBRE com  $p = 49,75\%$  apresentou um desempenho ligeiramente superior ao do método original, ocupando, assim, a primeira posição do *ranking*. Em segundo lugar, ficou o Algoritmo de Tukey com  $c = 2,7$ . Por fim, a terceira posição foi ocupada pelo Método do Boxplot.

Para o município de São Paulo, o Método do IBRE apresentou o melhor desempenho dentre os demais, ocupando, então, a primeira posição do *ranking*. A segunda posição foi ocupada pelo Algoritmo de Tukey com  $c = 2,4$ . Para finalizar, em terceiro lugar, ficou o Método do Boxplot.

Conforme mencionado, a suposição de que os preços alterados ou descartados pelos análises são os verdadeiros valores extremos pode influenciar os resultados, fazendo com que o mesmo seja favorecido em comparação com os demais métodos. Portanto, em trabalhos futuros, espera-se testar os métodos aqui abordados em dados obtidos por simulação ou que constituam bases largamente utilizadas na literatura, para que este pressuposto não interfira nos resultados.

Ainda que o método utilizado atualmente no IBRE tenha apresentado o melhor desempenho dentre os demais, este estudo foi importante pois testou métodos que utilizam os preços observados/coletados para a construção de seus intervalos de tolerância, o que faz com que seus limites inferior e superior sejam diferentes a cada fechamento do índice, ao invés de serem sempre fixos. Este trabalho possibilitou ainda o uso de informações passadas na construção dos intervalos de tolerância, permitindo que comportamentos observados anteriormente fossem levados em consideração no presente. Sendo assim, a obtenção de parâmetros móveis de crítica é a maior contribuição deste estudo.

Além disso, é importante chamar atenção para as proporções de acerto alcançadas pelo Método do IBRE tanto para o Rio de Janeiro (0,48333) quanto para São paulo (0,40816). Ambas são inferiores a 50%, o que mostra que mais da metade dos *outliers* não foram detectados pelo método atual, o que talvez seja um indício de que tal método precisa ser revisado.

Um próximo passo para o aproveitamento deste trabalho no IBRE seria a aplicação dos métodos aqui propostos para os demais subitens da estrutura de bens e serviços do IPC. Para o seu aproveitamento nas demais instituições, um passo inicial seria definir uma amostra de dados aos quais os métodos possam ser aplicados.

Em paralelo, como neste estudo não é feita nenhuma transformação nos dados e os métodos são aplicados considerando  $c_L = c_U = c$ , pode-se estudar a distribuição das informações, de forma a concluir se as mesmas são ou não simétricas. Caso sejam assimétricas, sugere-se a aplicação do Método de Cercas Resistentes, do Método do Quartil, do Método do Quartil Modificado e do Método do Desvio Mediano Absoluto em dados transformados ou a aplicação dos mesmos e do Algoritmo de Tukey considerando  $c_L \neq c_U$ .

Para finalizar, tem-se que a comparação entre os métodos para detecção de *outliers* foi baseada na análise dos gráficos, obtidos por meio dos valores da proporção de acerto dos métodos e da quantidade de falsos positivos gerada por eles. Em trabalhos futuros, poderia ser considerada a determinação dos parâmetros dos métodos usando um processo de treinamento para calibração dos testes. Além disso, outro aspecto a ser considerado é a identificação de critérios mais objetivos para a seleção do melhor método para detecção de *outliers*.





## Referências

AUCREMANNE, L. et al. **A study of belgian inflation, relative prices and nominal rigidities using new robust measures of skewness and tail weight.** In: Hubert, M., Pison, G., Struyf, A., Van Aelst, S. (Eds.), **Theory and Applications of Recent Robust Methods, Series: Statistics for Industry and Technology.** [S.l.]: Birkhauser, Basel, 2004. 13–25 p.

BARNETT, V.; LEWIS, T. **Outliers in Statistical Data.** [S.l.]: John Wiley & Sons, 1994.

BOWLEY, A. L. **Elements of Statistics.** New York: Charles Scribner's Sons, 1920.

BOX, G.; COX, D. An analysis of transformations. **Journal of the Royal Statistical Society–Series B**, v. 26, n. 2, p. 211–252, 1964.

BRYs, G.; HUBERT, M.; STRUYF, A. A robust measure of skewness. **Journal of Computational and Graphical Statistics**, v. 13, n. 4, p. 996–1017, Dec. 2004.

DIEWERT, W. E. The consumer price index and index number theory: A survey. Vancouver, Canada, Feb. 2001.

GRUBBS, F. Procedures for detecting outlying observations in samples. **Technometrics**, v. 11, p. 1–21, 1969.

GUMBEL, E. Discussion on rejection of outliers by anscombe. **F. J. Technometrics**, v. 2, p. 165–166., 1960.

HAMPEL, F. R. The influence curve and its role in robust estimation. **Journal of the American Statistical Association**, v. 69, n. 346, p. 383–393, 1974.

HAWKINS, D. M. **Identification of Outliers.** New York: Chapman and Hall, London, 1980.

HIDIROGLOU, M.; BERTHELOT, J. Statistical editing and imputation for periodic business surveys. **Survey Methodology**, v. 12, n. 1, p. 73–83, June 1986.

HINKLEY, D. V. On power transformations to symmetry. **Biometrika**, v. 62, n. 1, p. 101–111, Apr. 1975.

HOAGLIN, D.; MOSTELLER, F.; TUKEY, J. **Exploring Data Tables, Trends and Shapes**. New York: Wiley, 1995. 463-478 p.

HUBERT, M.; VANDERVIEREN, E. An adjusted boxplot for skewed distributions. **Computational Statistics and Data Analysis**, Department of Mathematics - Leuven Statistics Research Center, K.U.Leuven, Celestijnenlaan 200B, B-3001 Leuven, Belgium, Department of Mathematics & Computer Science, University of Antwerp, Middelheimlaan 1, B-2020 Antwerp, Belgium, v. 62, n. 12, p. 5186–5201, Aug. 2008.

HUTTON, H. Dynamic outlier detection in price index surveys. **Statistical Society of Canada Proceedings**, Statistics Canada, 2008.

INSTITUTO BRASILEIRO DE ECONOMIA. **Metdologia IGP-DI**. [S.l.], 2013.

KIMBER, A. C. Exploratory data analysis for possibly censored data from skewed distributions. **Applied Statistics**, Guildford, UK, v. 39, n. 1, p. 21–30, 1990.

LEE, H. et al. **Outliers in Sample Surveys**. [S.l.]: [Ottawa, Ont.] : Statistics Canada, Business Survey Methods Division, 1992.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2013. Disponível em: <<http://www.R-project.org/>>.

RAIS, S. Outlier detection for statistics canada's consumer price index. **Business Survey Methods Division, Statistics Canada**, 2007.

RAIS, S. Outlier detection for the consumer price index. **Statistical Society of Canada Proceedings**, Industrial Organization, Finance, and Prices Section, BSMD, 2008.

SAÏDI, A.; RUBIN-BLEUER, S. Detection of outliers in the canadian consumer price index.

**Business Survey Methods Division, Statistics Canada**, May 2005.

TAN, P.; STEINBACH, M.; KUMAR, V. **Introduction to Data Mining**. [S.l.]: Addison-Wesley, 2006.

THOMPSON, K.; SIGMAN, S. Statistical methods for developing ratio edit tolerances for economic data. **Journal of Official Statistics**, v. 15, n. 4, 1999.

TUKEY, J. W. **Exploratory Data Analysis**. Princeton University and Bell Telephone Laboratories: Reading, Massachusetts: Addison-Wesley Publishing Company, 1977. 39-49 p.

UNITED KINGDOM'S OFFICE OF NATIONAL STATISTICS. **Consumer Price Indices Technical Manual**v. [S.l.], 2006.



**ANEXO - Estrutura Completa de Bens e Serviços do IPC**

<b>Código</b>	<b>Descrição</b>
0	IPC - TODOS OS ITENS
1	ALIMENTAÇÃO
11	GÊNEROS ALIMENTÍCIOS
1101	ARROZ E FEIJÃO
110101	ARROZ
110107	FEIJÃO-CARIOCA
110111	FEIJÃO-PRETO
110115	FEIJÃO-MULATINHO
1103	HORTALIÇAS E LEGUMES
110301	ABÓBORA
110309	APIIM / MANDIOCA
110313	ALFACE
110315	ALHO
110319	BATATA-INGLESA
110331	BRÓCOLIS
110333	CEBOLA
110335	CENOURA
110345	COUVE
110357	INHAME
110363	PIMENTÃO
110371	TOMATE
110377	COENTRO
1105	FRUTAS
110503	ABACAXI
110507	BANANA-NANICA (D'ÁGUA)
110511	BANANA-PRATA
110513	BANANA DA TERRA
110527	LARANJA-PERA
110534	MAÇÃ
110537	MAMÃO PAPAYA
110539	MAMÃO FORMOSA
110549	PERA
110553	TANGERINA (MEXERICA)
110555	UVA
110569	MELANCIA
110571	MARACUJÁ
110573	MANGA
1107	MASSAS E FARINHAS
110703	FARINHA DE MANDIOCA

Código	Descrição
110707	FARINHA DE TRIGO
110711	FARINHA E FLOCOS DE CEREAIS
110715	MACARRÃO
110717	MACARRÃO INSTANTÂNEO
110737	FUBÁ DE MILHO
1109	PANIFICADOS E BISCOITOS
110901	BISCOITOS
110911	PÃO FRANCÊS
110913	PÃO DE FORMA
110915	PÃO DE OUTROS TIPOS
110917	BOLO PRONTO
110921	PÃO DOCE
1111	ADOÇANTES
111101	AÇÚCAR REFINADO
111103	AÇÚCAR CRISTAL
1113	DOCES E CHOCOLATES
111303	BOMBONS E CHOCOLATES
111329	BALAS E DOCES
1115	LATICÍNIOS
111503	IOGURTE NATURAL OU COM POLPA DE FRUTA
111505	LEITE CONDENSADO
111507	LEITE EM PÓ
111511	LEITE TIPO LONGA VIDA
111513	MANTEIGA
111515	QUEIJO MINAS
111517	QUEIJO MUÇARELA
111519	QUEIJO PARMESÃO
111521	QUEIJO PRATO
111522	QUEIJO COALHO
111525	REQUEIJÃO
111527	SORVETE E PICOLÉ
111539	BEBIDAS LÁCTEAS
1117	AVES E OVOS
111701	FRANGO EM PEDAÇOS
111703	FRANGO INTEIRO
111707	OVOS
1119	CARNES BOVINAS
111901	ACÉM
111903	ALCATRA
111907	PATINHO
111909	CHÃO DE DENTRO
111911	CONTRAFILÉ

Código	Descrição
111913	COSTELA BOVINA
111915	PICANHA
111917	FÍGADO BOVINO
111919	FILÉ-MIGNON
111921	LAGARTO
111927	CARNE MOÍDA
111928	MAMINHA
111929	MÚSCULO
111931	PÁ / PALETA
111932	PEITO BOVINO
1121	CARNES SUÍNAS
112101	CARRÉ / BISTECA
112103	COSTELA SUÍNA
112107	LOMBO SUÍNO
112109	PERNIL SUÍNO
112112	BACON / TOUCINHO
1123	CARNES DE OUTROS ANIMAIS
112303	PALETA DE CARNEIRO
112305	COSTELA DE CARNEIRO
1125	PESCADOS FRESCOS
112507	CAMARÃO
112510	CAVALA
112515	CORVINA
112517	FILÉ DE PEIXE DE ÁGUA DOCE NÃO ESPECIFICADO
112519	NAMORADO
112529	SARDINHA
112536	PESCADA
112554	CAÇÃO
112559	MERLUZA
1127	CARNES E PEIXES INDUSTRIALIZADOS
112701	ATUM
112702	APRESUNTADO
112703	BACALHAU
112711	CARNE SECA
112715	CARNE SALGADA DE BOVINO
112719	LINGUIÇA
112723	MORTADELA
112731	PRESUNTO
112733	SALSICHA E SALSICHÃO
112737	SARDINHA EM CONSERVA
112761	PATÊ
1129	ÓLEOS E GORDURAS

Código	Descrição
112903	AZEITE
112907	MARGARINA
112911	ÓLEO DE SOJA
1131	CONDIMENTOS
113111	LEITE DE COCO
113113	MAIONESE
113115	EXTRATO DE TOMATE
113153	TEMPEROS PRONTOS
1133	VEGETAIS EM CONSERVA
113301	AZEITONA EM CONSERVA
113305	ERVILHA EM CONSERVA
113309	PALMITO EM CONSERVA
113355	MILHO EM CONSERVA
1135	BEBIDAS NÃO ALCOÓLICAS
113505	CAFÉ EM PÓ
113517	REFRIGERANTES E ÁGUA MINERAL
113521	SUCOS DE FRUTA
113529	POLPA DE FRUTA CONGELADA
113553	REFRESCO DE FRUTA EM PÓ
113557	ERVA MATE
113573	BEBIDAS DE SOJA
1136	BEBIDAS ALCOÓLICAS
113601	CERVEJA
113603	BEBIDAS DESTILADAS
113605	VINHO
1137	OUTROS GÊNEROS ALIMENTÍCIOS
113715	MILHO DE PIPOCA
113719	SOPA DESIDRATADA
113755	BATATA FRITA
1139	ALIMENTOS PRONTOS CONGELADOS
113901	ALIMENTOS PREPARADOS E CONGELADOS DE CARNE BOVINA
113921	ALIMENTOS PREPARADOS E CONGELADOS DE AVES
113941	MASSAS PREPARADAS E CONGELADAS
1141	GÊNEROS ALIMENTÍCIOS DIET / LIGHT
114104	IOGURTE DIET / LIGHT
114110	REFRIGERANTE DIET / LIGHT
114116	PÃO DE FORMA DIET / LIGHT
12	ALIMENTAÇÃO FORA
1201	RESTAURANTES
120101	REFEIÇÕES EM BARES E RESTAURANTES
1203	BARES E LANCHONETES
120301	DOCES E SALGADOS
120309	SANDUÍCHES



Código	Descrição
120311	SORVETES FORA DE CASA
120314	CAFÉ DA MANHÃ
1207	BEBIDAS ALCOÓLICAS E NÃO ALCOÓLICAS
120701	SUCOS DE FRUTAS FORA DE CASA
120703	CAFEZINHO
120705	REFRIGERANTES E ÁGUA MINERAL FORA DE CASA
120707	CERVEJAS E CHOPPS
120709	OUTRAS BEBIDAS ALCOÓLICAS
2	HABITAÇÃO
21	ALUGUEL E ENCARGOS
2101	ALUGUEL E CONDOMÍNIO
210101	ALUGUEL RESIDENCIAL
210103	CONDOMÍNIO RESIDENCIAL
22	SERVIÇOS PÚBLICOS DE RESIDÊNCIA
2201	SERVIÇOS PÚBLICOS DE RESIDÊNCIA
220101	TARIFA DE ELETRICIDADE RESIDENCIAL
220103	GÁS DE BUJÃO
220105	TARIFA DE GÁS ENCANADO
220111	TAXA DE ÁGUA E ESGOTO RESIDENCIAL
23	MOBILIÁRIO
2301	MÓVEIS
230111	MÓVEIS PARA RESIDÊNCIA
2307	COLCHÕES
230703	COLCHÕES E COLCHONETES
24	ROUPAS DE CAMA, MESA E BANHO
2401	ROUPAS DE CAMA, MESA E BANHO
240103	LENÇOL E FRONHA
240105	TOALHAS EM GERAL
25	ELETRODOMÉSTICOS E EQUIPAMENTOS
2501	ELETRODOMÉSTICOS
250101	AR CONDICIONADO
250123	GELADEIRA E FREEZER
250129	MÁQUINA DE LAVAR ROUPAS
250131	LIQUIDIFICADOR
250143	VENTILADOR E CIRCULADOR DE AR
250145	FOGÃO
250147	FORNO ELÉTRICO E DE MICRO-ONDAS
2503	EQUIPAMENTOS ELETRÔNICOS
250301	APARELHO DE SOM
250305	VIDEO GAME
250317	APARELHO DE TV
250319	COMPUTADOR E PERIFÉRICOS

<b>Código</b>	<b>Descrição</b>
250326	APARELHO TELEFÔNICO CELULAR
250329	APARELHO DE DVD E BLU-RAY
250331	MÁQUINA DE FOTOGRAFAR E FILMAR
<b>26</b>	<b>UTENSÍLIOS DIVERSOS</b>
2601	ARTIGOS DE DECORAÇÃO
260101	CORTINA
260115	TAPETE
2603	LOUÇAS E PANEAS
260305	COPO DE VIDRO OU PLÁSTICO
260309	PANEAS EM GERAL
260317	TALHERES
260319	PRATOS DE LOUÇA
2605	OUTROS UTENSÍLIOS DOMÉSTICOS
260511	GARRAFA TÉRMICA
260525	PILHA
<b>27</b>	<b>ARTIGOS DE CONSERVAÇÃO E REPARO</b>
2701	MATERIAL PARA LIMPEZA
270101	ÁGUA SANITÁRIA
270105	AMACIANTE
270111	DESINFETANTE
270113	DETERGENTE
270119	INSETICIDA
270135	SABÃO EM BARRA
270137	SABÃO EM PÓ
2703	MATERIAL PARA PINTURA
270303	MATERIAL DE PINTURA
2705	MATERIAL HIDRÁULICO
270501	MATERIAL HIDRÁULICO
2707	MATERIAL ELÉTRICO
270703	LÂMPADA
270705	MATERIAL ELÉTRICO - DIVERSOS
2711	MATERIAL PARA REPAROS DE RESIDÊNCIA
271101	MATERIAL PARA REPAROS DE RESIDÊNCIA
<b>28</b>	<b>SERVIÇOS DE RESIDÊNCIA</b>
2801	EMPREGADOS DOMÉSTICOS
280101	EMPREGADA DOMÉSTICA DIARISTA
280103	EMPREGADA DOMÉSTICA MENSALISTA
280105	LAVADEIRA E PASSADEIRA
2803	OUTROS SERVIÇOS DE RESIDÊNCIA
280307	CONCERTO DE ELETRODOMÉSTICOS
2805	MÃO DE OBRA PARA REPAROS EM RESIDÊNCIA
280501	MÃO DE OBRA PARA REPAROS EM RESIDÊNCIA

Código	Descrição
3	VESTUÁRIO
31	ROUPAS
3101	ROUPAS MASCULINAS
310101	BERMUDA MASCULINA
310103	TERNO E BLAZER MASCULINO
310105	CALÇA COMPRIDA MASCULINA
310111	CAMISA MASCULINA
310115	AGASALHO MASCULINO
310117	ROUPA ÍNTIMA MASCULINA
3103	ROUPAS FEMININAS
310301	BERMUDA FEMININA
310307	BLUSA FEMININA
310309	CALÇA COMPRIDA FEMININA
310313	ROUPA ÍNTIMA FEMININA
310319	AGASALHO FEMININO
310331	VESTIDO E SAIA
3105	ROUPAS INFANTIS
310501	AGASALHO INFANTIL
310503	BERMUDA INFANTIL
310505	BLUSA DE MALHA INFANTIL
310507	CALÇA INFANTIL
310513	ROUPA INFANTIL PARA PRÁTICA ESPORTIVA
310517	VESTIDO E SAIA INFANTIL
310525	ROUPA ÍNTIMA INFANTIL
32	CALÇADOS
3201	CALÇADOS MASCULINOS
320101	SANDÁLIA MASCULINA
320105	SAPATO MASCULINO
320107	TÊNIS MASCULINO
3203	CALÇADOS FEMININOS
320303	SANDÁLIA FEMININA
320305	TÊNIS FEMININO
320307	SAPATO FEMININO
3205	CALÇADOS INFANTIS
320505	SANDÁLIA INFANTIL
320509	SAPATO INFANTIL
320511	TÊNIS INFANTIL
33	ACESSÓRIOS DO VESTUÁRIO
3301	RELÓGIOS E BIJUTERIAS
330101	BIJUTERIAS EM GERAL
330105	RELÓGIO
3303	CINTOS E BOLSAS

Código	Descrição
330301	CINTO E BOLSA
35	SERVIÇOS DO VESTUÁRIO
3501	SERVIÇOS DE CONFEÇÃO
350101	ALFAIATE E COSTUREIRA
3503	CONSERVAÇÃO DO VESTUÁRIO
350301	TINTURARIA E LAVANDERIA
4	SAÚDE E CUIDADOS PESSOAIS
41	SERVIÇOS DE SAÚDE
4101	HOSPITAIS E LABORATÓRIOS
410103	EXAME DE LABORATÓRIO
410105	EXAMES RADIOLÓGICOS E DIAGNÓSTICOS POR IMAGEM
410107	DIÁRIA HOSPITALAR
4103	MÉDICO, DENTISTA E OUTROS
410301	DENTISTA
410305	MÉDICO
410307	PSICÓLOGO
410309	PLANO E SEGURO DE SAÚDE
410319	OUTROS PROFISSIONAIS DE SAÚDE
42	PRODUTOS MÉDICO-ODONTOLÓGICOS
4201	MEDICAMENTOS EM GERAL
420105	GASTROPROTETOR
420107	PSICOTRÓPICO E ANOREXÍGENO
420111	ANALGÉSICO E ANTITÉRMICO
420114	ANTIINFLAMATÓRIO E ANTIBIÓTICO
420115	ANTIGRIPE E ANTITUSSÍGENO
420116	ANTIALÉRGICO E BRONCODILATADOR
420117	VASODILATADOR PARA PRESSÃO ARTERIAL
420118	CALMANTE E ANTIDEPRESSIVO
420123	ANTICONCEPCIONAL
420125	DERMATOLÓGICO
420126	VITAMINA E FORTIFICANTE
420131	ANTIMICÓTICO E PARASITICIDA
420133	REMÉDIO PARA DIABETE
420135	REMÉDIO PARA OSTEOPOROSE
420137	REMÉDIOS OFTAMOLÓGICOS
4203	APARELHOS MÉDICO-ODONTOLÓGICOS
420301	APARELHO DENTÁRIO
420303	ÓCULOS E LENTES
4205	PRODUTOS FARMACÊUTICOS
420501	ALGODÃO E COTONETE
420509	PRESERVATIVO E LUBRIFICANTE
43	CUIDADOS PESSOAIS

<b>Código</b>	<b>Descrição</b>
4301	ARTIGOS DE HIGIENE E CUIDADO PESSOAL
430101	ABSORVENTE HIGIÊNICO
430105	ARTIGO DE MAQUILAGEM
430109	CREME DENTAL
430113	PROTETORES PARA A PELE
430115	DESODORANTE
430119	ESMALTE PARA UNHA
430121	FRALDA DESCARTÁVEL
430125	PAPEL HIGIÊNICO
430127	PERFUME
430131	SABONETE
430133	SHAMPOO, CONDICIONADOR E CREME
4303	SERVIÇOS DE CUIDADOS PESSOAIS
430301	BARBEARIA
430303	SALÃO DE BELEZA
5	EDUCAÇÃO, LEITURA E RECREAÇÃO
51	EDUCAÇÃO
5101	CURSOS FORMAIS
510101	CURSO DE ENSINO FUNDAMENTAL
510103	CURSO DE ENSINO MÉDIO
510105	CURSO DE EDUCAÇÃO INFANTIL (PRÉ-ESCOLAR)
510107	CURSO DE ENSINO SUPERIOR
510153	CURSO DE EDUCAÇÃO INFANTIL (CRECHE)
510155	CURSO DE PÓS-GRADUAÇÃO
5103	CURSOS NÃO FORMAIS
510303	CURSO DE INFORMÁTICA
510304	CURSO PREPARATÓRIO - VESTIBULAR/CONCURSO
510305	CURSO DE LÍNGUA ESTRANGEIRA
510307	CURSO DE MÚSICA E DANÇA
510319	CURSOS DIVERSOS
5105	MATERIAL ESCOLAR E LIVROS EM GERAL
510503	MATERIAL ESCOLAR (EXCLUSIVE LIVROS)
510513	LIVROS NÃO DIDÁTICOS
510515	LIVROS DIDÁTICOS
52	LEITURA
5201	JORNAIS E REVISTAS
520101	JORNAL
520105	REVISTA EM GERAL
53	RECREAÇÃO
5301	BRINQUEDOS E ARTIGOS ESPORTIVOS
530103	BICICLETA
530107	BONECA

<b>Código</b>	<b>Descrição</b>
530113	JOGOS PARA RECREAÇÃO
530129	ARTIGOS ESPORTIVOS
5303	INSTRUMENTOS MUSICAIS
530317	INSTRUMENTO MUSICAL
5305	FOTOGRAFIA
530505	REVELAÇÃO DE FOTO
5307	DVD / CD
530755	DVD / CD
5309	SALAS DE ESPETÁCULO
530903	CINEMA
530905	SHOW MUSICAL
530907	TEATRO
5311	CULTURA FÍSICA
531103	ACADEMIA DE GINÁSTICA
5313	CLUBES E PARQUES
531301	CLUBE DE RECREAÇÃO
531303	FUTEBOL E OUTROS EVENTOS ESPORTIVOS
5315	PASSEIOS E FÉRIAS
531501	HOTEL
531505	PASSAGEM AÉREA
531509	EXCURSÃO E TOUR
<b>6</b>	<b>TRANSPORTES</b>
<b>61</b>	<b>TRANSPORTE PÚBLICO</b>
6101	TRANSPORTE PÚBLICO URBANO
610101	TARIFA DE BARCO E AEROBARCO
610103	TARIFA DE METRÔ
610105	TARIFA DE ÔNIBUS URBANO
610107	TARIFA DE TÁXI
610109	TRANSPORTE ESCOLAR
610111	TARIFA DE TREM URBANO
610115	TARIFA DE TRANSPORTE DE VAN E SIMILARES
6103	TRANSPORTE PÚBLICO INTERURBANO
610303	TARIFA DE ÔNIBUS INTERURBANO
<b>62</b>	<b>TRANSPORTE PRÓPRIO</b>
<b>6201</b>	<b>VEÍCULOS</b>
620101	AUTOMÓVEL NOVO
620103	AUTOMÓVEL USADO
6203	PEÇAS E ACESSÓRIOS
620309	ACESSÓRIOS E PEÇAS PARA AUTOMÓVEIS
620319	PNEU PARA AUTOMÓVEL
6205	COMBUSTÍVEIS E LUBRIFICANTES
620501	ETANOL

<b>Código</b>	<b>Descrição</b>
620503	GASOLINA
620505	ÓLEO LUBRIFICANTE
620507	ÓLEO DIESEL
620509	GÁS NATURAL VEICULAR
6207	SERVIÇOS DE OFICINA
620707	SERVIÇO DE REPARO EM AUTOMÓVEL
6209	OUTROS GASTOS COM VEÍCULOS
620903	LICENCIAMENTO - IPVA
620905	ESTACIONAMENTO E GARAGEM
620907	LAVAGEM E LUBRIFICAÇÃO
620909	PEDÁGIO
620911	SEGURO FACULTATIVO PARA VEÍCULO
620913	SEGURO OBRIGATÓRIO PARA VEÍCULO
<b>7</b>	<b>DESPESAS DIVERSAS</b>
71	FUMO
7103	FUMO
710301	CIGARROS
72	OUTRAS DESPESAS DIVERSAS
7201	CORREIO E TELEFONE PÚBLICO
720101	CARTÃO DE TELEFONE
720105	TARIFA POSTAL
720115	ACESSO À INTERNET EM LOJA
7203	LOTÉRIAS
720301	BILHETE LOTÉRICO
720305	JOGO LOTÉRICO
7205	DESPESAS COM ANIMAIS DOMÉSTICOS
720501	ALIMENTOS PARA ANIMAIS DOMÉSTICOS
720503	CLÍNICA VETERINÁRIA
7209	SERVIÇOS DIVERSOS
720901	CARTÓRIO
720905	SERVIÇO RELIGIOSO E FUNERÁRIO
<b>8</b>	<b>COMUNICAÇÃO</b>
81	SERVIÇOS DE COMUNICAÇÃO
8101	SERVIÇOS DE COMUNICAÇÃO
810101	TARIFA DE TELEFONE RESIDENCIAL
810103	PACOTES DE TELEFONIA FIXA E INTERNET
810105	TARIFA DE TELEFONE MÓVEL
810107	MENSALIDADE PARA TV POR ASSINATURA
810109	MENSALIDADE PARA INTERNET

## APÊNDICE A - Código para Cálculo dos Relativos

```

rel = function(tabela) {

  rel = c("-")

  for (i in 2:dim(tabela)[1]) {

    print(paste("i = ", i))

    if ((tabela$cod_ins_inf[i] == tabela$cod_ins_inf[i-1]) && (((
      tabela$ano_dcol[i] == tabela$ano_dcol[i-1]) && (as.numeric(
        tabela$mes_dcol[i]) == as.numeric(tabela$mes_dcol[i-1]) + 1)) | ((as.
        numeric(tabela$ano_dcol[i]) == as.numeric(tabela$ano_dcol[i-1]) + 1) &&
        (as.numeric(tabela$mes_dcol[i]) == 1) && (as.numeric(tabela$mes_dcol[i
        -1]) == 12)))) {

      rel = c(rel , round(tabela$media_original[i] / tabela$media_oficial[i-1],
        5))

    }

    else {

      rel = c(rel , "-")

    }

  }

  resultado = data.frame(tabela , rel)

  return(resultado)

}

```



## APÊNDICE B - Código da Implementação dos Métodos

### # Método do IBRE e Variações

```

aplica_metodos_ibre_e_variacoes = function(tabela , regiao , percorre) {

  out_metodo_ibre = c()

  out_metodo_ibre_49.75 = c()

  out_metodo_ibre_49.5 = c()

  out_metodo_ibre_49.25 = c()

  out_metodo_ibre_49 = c()

  out_metodo_ibre_48 = c()

  out_metodo_ibre_47 = c()

  out_metodo_ibre_46 = c()

  out_metodo_ibre_45 = c()

  out_metodo_ibre_40 = c()

  out_metodo_ibre_35 = c()

  out_metodo_ibre_30 = c()

  outlier = c()

  cod_ins = c()

  cod_inf = c()

  cod_ins_inf = c()

  regiao_uf = c()

  ano_dcol = c()

  mes_dcol = c()

```

```

rel = c()

for (z in 1:nrow(percorre)) {

  print(paste("z = ", z))

  i = percorre[z, 1]

  print(paste("i = ", i))

  j = percorre[z, 2]

  print(paste("j = ", j))

  k = percorre[z, 3]

  print(paste("k = ", k))

  temp1 = tabela[tabela$regiao_uf == regiao & tabela$ano_dcol == i &
    tabela$mes_dcol == j & tabela$cod_ins == k,]$rel

  temp2 = as.numeric(levels(temp1)[as.numeric(temp1)])

  temp3 = temp2[which(is.na(temp2) == FALSE)]

  temp4 = tabela[tabela$regiao_uf == regiao & tabela$ano_dcol == i &
    tabela$mes_dcol == j & tabela$cod_ins == k,]

  print(paste("tamanho temp3 = ", length(temp3)))

  if (length(temp3) >= 2) {

    qtdade_obs = length(temp3)

    metodo_ibre = rep(0, qtdade_obs)

    metodo_ibre[which(temp3 <= 0.5 | temp3 >= 1.5)] = 1

    metodo_ibre_49.75 = rep(0, qtdade_obs)

    metodo_ibre_49.75[which(temp3 <= 0.5025 | temp3 >= 1.4975)] = 1

    metodo_ibre_49.5 = rep(0, qtdade_obs)

```

```

metodo_ibre_49.5[which(temp3 <= 0.5050 | temp3 >= 1.495)] = 1

metodo_ibre_49.25 = rep(0, qtdade_obs)

metodo_ibre_49.25[which(temp3 <= 0.5075 | temp3 >= 1.4925)] = 1

metodo_ibre_49 = rep(0, qtdade_obs)

metodo_ibre_49[which(temp3 <= 0.51 | temp3 >= 1.49)] = 1

metodo_ibre_48 = rep(0, qtdade_obs)

metodo_ibre_48[which(temp3 <= 0.52 | temp3 >= 1.48)] = 1

metodo_ibre_47 = rep(0, qtdade_obs)

metodo_ibre_47[which(temp3 <= 0.53 | temp3 >= 1.47)] = 1

metodo_ibre_46 = rep(0, qtdade_obs)

metodo_ibre_46[which(temp3 <= 0.54 | temp3 >= 1.46)] = 1

metodo_ibre_45 = rep(0, qtdade_obs)

metodo_ibre_45[which(temp3 <= 0.55 | temp3 >= 1.45)] = 1

metodo_ibre_40 = rep(0, qtdade_obs)

metodo_ibre_40[which(temp3 <= 0.60 | temp3 >= 1.40)] = 1

metodo_ibre_35 = rep(0, qtdade_obs)

metodo_ibre_35[which(temp3 <= 0.65 | temp3 >= 1.35)] = 1

metodo_ibre_30 = rep(0, qtdade_obs)

metodo_ibre_30[which(temp3 <= 0.70 | temp3 >= 1.30)] = 1

out_metodo_ibre = c(out_metodo_ibre, metodo_ibre)

out_metodo_ibre_49.75 = c(out_metodo_ibre_49.75, metodo_ibre_49.75)

out_metodo_ibre_49.5 = c(out_metodo_ibre_49.5, metodo_ibre_49.5)

```

```

out_metodo_ibre_49.25 = c(out_metodo_ibre_49.25, metodo_ibre_49.25)

out_metodo_ibre_49 = c(out_metodo_ibre_49, metodo_ibre_49)

out_metodo_ibre_48 = c(out_metodo_ibre_48, metodo_ibre_48)

out_metodo_ibre_47 = c(out_metodo_ibre_47, metodo_ibre_47)

out_metodo_ibre_46 = c(out_metodo_ibre_46, metodo_ibre_46)

out_metodo_ibre_45 = c(out_metodo_ibre_45, metodo_ibre_45)

out_metodo_ibre_40 = c(out_metodo_ibre_40, metodo_ibre_40)

out_metodo_ibre_35 = c(out_metodo_ibre_35, metodo_ibre_35)

out_metodo_ibre_30 = c(out_metodo_ibre_30, metodo_ibre_30)

sem_traco = temp4[temp4$rel != "-",]

alt_ou_desc = sem_traco$num_palt + sem_traco$num_pdesc

alt_ou_desc[alt_ou_desc > 0] = 1

outlier = c(outlier, alt_ou_desc)

cod_ins = c(cod_ins, sem_traco$cod_ins)

cod_inf = c(cod_inf, sem_traco$cod_inf)

cod_ins_inf = c(cod_ins_inf, sem_traco$cod_ins_inf)

regiao_uf = c(regiao_uf, sem_traco$regiao_uf)

ano_dcol = c(ano_dcol, sem_traco$ano_dcol)

mes_dcol = c(mes_dcol, sem_traco$mes_dcol)

rel = c(rel, as.numeric(levels(sem_traco$rel)[as.numeric(sem_traco$rel)
]))

}

```

```
}  
  
resultado = data.frame(cod_ins ,  
  
    cod_inf ,  
  
    cod_ins_inf ,  
  
    regioao_uf ,  
  
    ano_dcol ,  
  
    mes_dcol ,  
  
    rel ,  
  
    out_metodo_ibre ,  
  
    out_metodo_ibre_49.75 ,  
  
    out_metodo_ibre_49.5 ,  
  
    out_metodo_ibre_49.25 ,  
  
    out_metodo_ibre_49 ,  
  
    out_metodo_ibre_48 ,  
  
    out_metodo_ibre_47 ,  
  
    out_metodo_ibre_46 ,  
  
    out_metodo_ibre_45 ,  
  
    out_metodo_ibre_40 ,  
  
    out_metodo_ibre_35 ,  
  
    out_metodo_ibre_30 ,  
  
    outlier)  
  
return(resultado)
```

```
}
```

# Métodos do Boxplot, do Boxplot SIQR e do Boxplot Ajustado

```
aplica_metodos_boxplots = function(tabela , regioao , percorre) {

  out_metodo_boxplot_padrao = c()

  out_metodo_boxplot_siqr = c()

  out_metodo_boxplot_ajustado = c()

  out_metodo_boxplot_padrao_mes_anterior = c()

  out_metodo_boxplot_siqr_mes_anterior = c()

  out_metodo_boxplot_ajustado_mes_anterior = c()

  out_metodo_boxplot_padrao_dois_anteriores = c()

  out_metodo_boxplot_siqr_dois_anteriores = c()

  out_metodo_boxplot_ajustado_dois_anteriores = c()

  out_metodo_boxplot_padrao_mesmo_mes_anos_anteriores = c()

  out_metodo_boxplot_siqr_mesmo_mes_anos_anteriores = c()

  out_metodo_boxplot_ajustado_mesmo_mes_anos_anteriores = c()

  out_metodo_boxplot_padrao_anteriores = c()

  out_metodo_boxplot_siqr_anteriores = c()

  out_metodo_boxplot_ajustado_anteriores = c()

  outlier = c()

  cod_ins = c()

  cod_inf = c()

  cod_ins_inf = c()

}
```

```
regiao_uf = c()

ano_dcol = c()

mes_dcol = c()

rel = c()

ano = c()

mes = c()

ins = c()

vet_li_boxplot_padrao = c()

vet_ls_boxplot_padrao = c()

vet_li_boxplot_siqr = c()

vet_ls_boxplot_siqr = c()

vet_li_boxplot_ajustado = c()

vet_ls_boxplot_ajustado = c()

vet_li_boxplot_padrao_mes_anterior = c()

vet_ls_boxplot_padrao_mes_anterior = c()

vet_li_boxplot_siqr_mes_anterior = c()

vet_ls_boxplot_siqr_mes_anterior = c()

vet_li_boxplot_ajustado_mes_anterior = c()

vet_ls_boxplot_ajustado_mes_anterior = c()

vet_li_boxplot_padrao_dois_anteriores = c()

vet_ls_boxplot_padrao_dois_anteriores = c()

vet_li_boxplot_siqr_dois_anteriores = c()
```

```

vet_ls_boxplot_siqr_dois_anteriores = c()

vet_li_boxplot_ajustado_dois_anteriores = c()

vet_ls_boxplot_ajustado_dois_anteriores = c()

vet_li_boxplot_padrao_mesmo_mes_anos_anteriores = c()

vet_ls_boxplot_padrao_mesmo_mes_anos_anteriores = c()

vet_li_boxplot_siqr_mesmo_mes_anos_anteriores = c()

vet_ls_boxplot_siqr_mesmo_mes_anos_anteriores = c()

vet_li_boxplot_ajustado_mesmo_mes_anos_anteriores = c()

vet_ls_boxplot_ajustado_mesmo_mes_anos_anteriores = c()

vet_li_boxplot_padrao_anteriores = c()

vet_ls_boxplot_padrao_anteriores = c()

vet_li_boxplot_siqr_anteriores = c()

vet_ls_boxplot_siqr_anteriores = c()

vet_li_boxplot_ajustado_anteriores = c()

vet_ls_boxplot_ajustado_anteriores = c()

for (z in 1:nrow(percorre)) {

  print(paste("z = ", z))

  i = percorre[z, 1]

  print(paste("i = ", i))

  j = percorre[z, 2]

  print(paste("j = ", j))

  k = percorre[z, 3]

```



```

print(paste("k = ", k))

temp1 = tabela[tabela$regiao_uf == regiao & tabela$ano_dcol == i &
  tabela$mes_dcol == j & tabela$cod_ins == k,]$rel # Pega os relativos do
  ano/êms atual

temp2 = as.numeric(levels(temp1)[as.numeric(temp1)]) # Transforma em
  únmero, gerando NA para os que ão "-"

temp3 = temp2[which(is.na(temp2) == FALSE)] # Exclui os NAs

temp4 = tabela[tabela$regiao_uf == regiao & tabela$ano_dcol == i &
  tabela$mes_dcol == j & tabela$cod_ins == k,] # Tabela do ano/êms atual

if (j == 1) {

  temp5 = tabela[tabela$regiao_uf == regiao & tabela$ano_dcol == i - 1 &
    tabela$mes_dcol == 12 & tabela$cod_ins == k,]$rel # Pega os relativos
    do êms anterior

  temp17_1 = tabela[tabela$regiao_uf == regiao & tabela$ano_dcol == i - 1
    & tabela$mes_dcol == 11 & tabela$cod_ins == k,]

  temp17_2 = tabela[tabela$regiao_uf == regiao & tabela$ano_dcol == i - 1
    & tabela$mes_dcol == 12 & tabela$cod_ins == k,]

  temp17 = rbind(temp17_1, temp17_2)$rel

}

else if (j == 2) {

  temp5 = tabela[tabela$regiao_uf == regiao & tabela$ano_dcol == i &
    tabela$mes_dcol == j - 1 & tabela$cod_ins == k,]$rel # Pega os
    relativos do êms anterior

  temp17_1 = tabela[tabela$regiao_uf == regiao & tabela$ano_dcol == i - 1
    & tabela$mes_dcol == 12 & tabela$cod_ins == k,]

  temp17_2 = tabela[tabela$regiao_uf == regiao & tabela$ano_dcol == i &
    tabela$mes_dcol == j - 1 & tabela$cod_ins == k,]

  temp17 = rbind(temp17_1, temp17_2)$rel

```

```

}

else {

  temp5 = tabela[tabela$regiao_uf == regiao & tabela$ano_dcol == i &
    tabela$mes_dcol == j - 1 & tabela$cod_ins == k,]$rel # Pega os
    relativos do êms anterior

  temp17 = tabela[tabela$regiao_uf == regiao & tabela$ano_dcol == i &
    tabela$mes_dcol == c(j - 2, j - 1) & tabela$cod_ins == k,]$rel

}

temp6 = as.numeric(levels(temp5)[as.numeric(temp5)]) # Transforma em
  únmero, gerando NA para os que ão "-"

temp7 = temp6[which(is.na(temp6) == FALSE)] # Exclui os NAs

temp18 = as.numeric(levels(temp17)[as.numeric(temp17)]) # Transforma em
  únmero, gerando NA para os que ão "-"

temp19 = temp18[which(is.na(temp18) == FALSE)] # Exclui os NAs

temp8 = tabela[tabela$regiao_uf == regiao & tabela$ano_dcol < i &
  tabela$mes_dcol == j & tabela$cod_ins == k,]$rel # Pega os relativos do
  mesmo êms dos anos anteriores

temp9 = as.numeric(levels(temp8)[as.numeric(temp8)]) # Transforma em
  únmero, gerando NA para os que ão "-"

temp10 = temp9[which(is.na(temp9) == FALSE)] # Exclui os NAs

temp11 = tabela[tabela$regiao_uf == regiao & tabela$ano_dcol < i &
  tabela$cod_ins == k,] # Tabela com os relativos dos anos anteriores

temp12 = tabela[tabela$regiao_uf == regiao & tabela$ano_dcol == i &
  tabela$mes_dcol < j & tabela$cod_ins == k,] # Tabela com os relativos
  dos meses anteriores do mesmo ano

temp13 = rbind(temp11, temp12) # Tabela todos os relativos anteriores

temp14 = rbind(temp11, temp12)$rel # Pega todos os relativos anteriores

```

```

temp15 = as.numeric(levels(temp14)[as.numeric(temp14)]) # Transforma em
    número, gerando NA para os que são "-"

temp16 = temp15[which(is.na(temp15) == FALSE)] # Exclui os NAs

print(paste("tamanho temp3 = ", length(temp3)))

if (length(temp3) >= 2) {

    qtdade_obs = length(temp3)

    metodo_boxplot_padrao = rep(0, qtdade_obs)

    li_boxplot_padrao = quantile(temp3, 0.25, na.rm = TRUE) - 1.5 * IQR(
        temp3)

    ls_boxplot_padrao = quantile(temp3, 0.75, na.rm = TRUE) + 1.5 * IQR(
        temp3)

    metodo_boxplot_padrao[which(temp3 < li_boxplot_padrao | temp3 >
        ls_boxplot_padrao)] = 1

    metodo_boxplot_siqr = rep(0, qtdade_obs)

    li_boxplot_siqr = quantile(temp3, 0.25, na.rm = TRUE) - 3 * (median(
        temp3) - quantile(temp3, 0.25, na.rm = TRUE))

    ls_boxplot_siqr = quantile(temp3, 0.75, na.rm = TRUE) + 3 * (quantile(
        temp3, 0.75, na.rm = TRUE) - median(temp3))

    metodo_boxplot_siqr[which(temp3 < li_boxplot_siqr | temp3 >
        ls_boxplot_siqr)] = 1

    boxplot_ajustado = adjbox(temp3)

    metodo_boxplot_ajustado = rep(0, qtdade_obs)

    li_boxplot_ajustado = boxplot_ajustado$fence[1]

    ls_boxplot_ajustado = boxplot_ajustado$fence[2]

    metodo_boxplot_ajustado[which(temp3 < li_boxplot_ajustado | temp3 >
        ls_boxplot_ajustado)] = 1

```

```

metodo_boxplot_padrao_mes_anterior = rep(0, qtdade_obs)

metodo_boxplot_siqr_mes_anterior = rep(0, qtdade_obs)

metodo_boxplot_ajustado_mes_anterior = rep(0, qtdade_obs)

if (length(temp7) >= 2) {

  li_boxplot_padrao_mes_anterior = quantile(temp7, 0.25, na.rm = TRUE)
    - 1.5 * IQR(temp7)

  ls_boxplot_padrao_mes_anterior = quantile(temp7, 0.75, na.rm = TRUE) +
    1.5 * IQR(temp7)

  metodo_boxplot_padrao_mes_anterior[which(temp3 <
    li_boxplot_padrao_mes_anterior | temp3 >
    ls_boxplot_padrao_mes_anterior)] = 1

  li_boxplot_siqr_mes_anterior = quantile(temp7, 0.25, na.rm = TRUE) - 3
    * (median(temp7) - quantile(temp7, 0.25, na.rm = TRUE))

  ls_boxplot_siqr_mes_anterior = quantile(temp7, 0.75, na.rm = TRUE) + 3
    * (quantile(temp7, 0.75, na.rm = TRUE) - median(temp7))

  metodo_boxplot_siqr_mes_anterior[which(temp3 <
    li_boxplot_siqr_mes_anterior | temp3 > ls_boxplot_siqr_mes_anterior
    )] = 1

  boxplot_ajustado_mes_anterior = adjbox(temp7)

  li_boxplot_ajustado_mes_anterior = boxplot_ajustado_mes_anterior$fence
    [1]

  ls_boxplot_ajustado_mes_anterior = boxplot_ajustado_mes_anterior$fence
    [2]

  metodo_boxplot_ajustado_mes_anterior[which(temp3 <
    li_boxplot_ajustado_mes_anterior | temp3 >
    ls_boxplot_ajustado_mes_anterior)] = 1

}

else {

```

```

metodo_boxplot_padrao_mes_anterior = metodo_boxplot_padrao

metodo_boxplot_siqr_mes_anterior = metodo_boxplot_siqr

metodo_boxplot_ajustado_mes_anterior = metodo_boxplot_ajustado

li_boxplot_padrao_mes_anterior = li_boxplot_padrao

ls_boxplot_padrao_mes_anterior = ls_boxplot_padrao

li_boxplot_siqr_mes_anterior = li_boxplot_siqr

ls_boxplot_siqr_mes_anterior = ls_boxplot_siqr

li_boxplot_ajustado_mes_anterior = li_boxplot_ajustado

ls_boxplot_ajustado_mes_anterior = ls_boxplot_ajustado

}

metodo_boxplot_padrao_mesmo_mes_anos_anteriores = rep(0, qtdade_obs)

metodo_boxplot_siqr_mesmo_mes_anos_anteriores = rep(0, qtdade_obs)

metodo_boxplot_ajustado_mesmo_mes_anos_anteriores = rep(0, qtdade_obs)

if (length(temp10) >= 2) {

  li_boxplot_padrao_mesmo_mes_anos_anteriores = quantile(temp10, 0.25,
    na.rm = TRUE) - 1.5 * IQR(temp10)

  ls_boxplot_padrao_mesmo_mes_anos_anteriores = quantile(temp10, 0.75,
    na.rm = TRUE) + 1.5 * IQR(temp10)

  metodo_boxplot_padrao_mesmo_mes_anos_anteriores[which(temp3 <
    li_boxplot_padrao_mesmo_mes_anos_anteriores | temp3 >
    ls_boxplot_padrao_mesmo_mes_anos_anteriores)] = 1

  li_boxplot_siqr_mesmo_mes_anos_anteriores = quantile(temp10, 0.25, na.
    rm = TRUE) - 3 * (median(temp10) - quantile(temp10, 0.25, na.rm =
    TRUE))

  ls_boxplot_siqr_mesmo_mes_anos_anteriores = quantile(temp10, 0.75, na.
    rm = TRUE) + 3 * (quantile(temp10, 0.75, na.rm = TRUE) - median(

```

```

temp10))

metodo_boxplot_siqr_mesmo_mes_anos_anteriores[ which(temp3 <
  li_boxplot_siqr_mesmo_mes_anos_anteriores | temp3 >
  ls_boxplot_siqr_mesmo_mes_anos_anteriores)] = 1

boxplot_ajustado_mesmo_mes_anos_anteriores = adjbox(temp10)

li_boxplot_ajustado_mesmo_mes_anos_anteriores =
  boxplot_ajustado_mesmo_mes_anos_anteriores$fence[1]

ls_boxplot_ajustado_mesmo_mes_anos_anteriores =
  boxplot_ajustado_mesmo_mes_anos_anteriores$fence[2]

metodo_boxplot_ajustado_mesmo_mes_anos_anteriores[ which(temp3 <
  li_boxplot_ajustado_mesmo_mes_anos_anteriores | temp3 >
  ls_boxplot_ajustado_mesmo_mes_anos_anteriores)] = 1

}

else {

metodo_boxplot_padrao_mesmo_mes_anos_anteriores = metodo_boxplot_padrao

metodo_boxplot_siqr_mesmo_mes_anos_anteriores = metodo_boxplot_siqr

metodo_boxplot_ajustado_mesmo_mes_anos_anteriores =
  metodo_boxplot_ajustado

li_boxplot_padrao_mesmo_mes_anos_anteriores = li_boxplot_padrao

ls_boxplot_padrao_mesmo_mes_anos_anteriores = ls_boxplot_padrao

li_boxplot_siqr_mesmo_mes_anos_anteriores = li_boxplot_siqr

ls_boxplot_siqr_mesmo_mes_anos_anteriores = ls_boxplot_siqr

li_boxplot_ajustado_mesmo_mes_anos_anteriores = li_boxplot_ajustado

ls_boxplot_ajustado_mesmo_mes_anos_anteriores = ls_boxplot_ajustado

}

metodo_boxplot_padrao_anteriores = rep(0, qtdade_obs)

```

```

metodo_boxplot_siqr_anteriores = rep(0, qtdade_obs)

metodo_boxplot_ajustado_anteriores = rep(0, qtdade_obs)

if (length(temp16) >= 2) {

  li_boxplot_padrao_anteriores = quantile(temp16, 0.25, na.rm = TRUE) -
    1.5 * IQR(temp16)

  ls_boxplot_padrao_anteriores = quantile(temp16, 0.75, na.rm = TRUE) +
    1.5 * IQR(temp16)

  metodo_boxplot_padrao_anteriores[which(temp3 <
    li_boxplot_padrao_anteriores | temp3 > ls_boxplot_padrao_anteriores
  )] = 1

  li_boxplot_siqr_anteriores = quantile(temp16, 0.25, na.rm = TRUE) - 3
    * (median(temp16) - quantile(temp16, 0.25, na.rm = TRUE))

  ls_boxplot_siqr_anteriores = quantile(temp16, 0.75, na.rm = TRUE) + 3
    * (quantile(temp16, 0.75, na.rm = TRUE) - median(temp16))

  metodo_boxplot_siqr_anteriores[which(temp3 <
    li_boxplot_siqr_anteriores | temp3 > ls_boxplot_siqr_anteriores)] =
    1

  boxplot_ajustado_anteriores = adjbox(temp16)

  li_boxplot_ajustado_anteriores = boxplot_ajustado_anteriores$fence[1]

  ls_boxplot_ajustado_anteriores = boxplot_ajustado_anteriores$fence[2]

  metodo_boxplot_ajustado_anteriores[which(temp3 <
    li_boxplot_ajustado_anteriores | temp3 >
    ls_boxplot_ajustado_anteriores)] = 1

}

else {

  metodo_boxplot_padrao_anteriores = metodo_boxplot_padrao

  metodo_boxplot_siqr_anteriores = metodo_boxplot_siqr

```

```

metodo_boxplot_ajustado_anteriores = metodo_boxplot_ajustado

li_boxplot_padrao_anteriores = li_boxplot_padrao

ls_boxplot_padrao_anteriores = ls_boxplot_padrao

li_boxplot_siqr_anteriores = li_boxplot_siqr

ls_boxplot_siqr_anteriores = ls_boxplot_siqr

li_boxplot_ajustado_anteriores = li_boxplot_ajustado

ls_boxplot_ajustado_anteriores = ls_boxplot_ajustado

}

metodo_boxplot_padrao_dois_anteriores = rep(0, qtdade_obs)

metodo_boxplot_siqr_dois_anteriores = rep(0, qtdade_obs)

metodo_boxplot_ajustado_dois_anteriores = rep(0, qtdade_obs)

if (length(temp19) >= 2) {

  li_boxplot_padrao_dois_anteriores = quantile(temp19, 0.25, na.rm =
    TRUE) - 1.5 * IQR(temp19)

  ls_boxplot_padrao_dois_anteriores = quantile(temp19, 0.75, na.rm =
    TRUE) + 1.5 * IQR(temp19)

  metodo_boxplot_padrao_dois_anteriores[which(temp3 <
    li_boxplot_padrao_dois_anteriores | temp3 >
    ls_boxplot_padrao_dois_anteriores)] = 1

  li_boxplot_siqr_dois_anteriores = quantile(temp19, 0.25, na.rm = TRUE)
    - 3 * (median(temp19) - quantile(temp19, 0.25, na.rm = TRUE))

  ls_boxplot_siqr_dois_anteriores = quantile(temp19, 0.75, na.rm = TRUE)
    + 3 * (quantile(temp19, 0.75, na.rm = TRUE) - median(temp19))

  metodo_boxplot_siqr_dois_anteriores[which(temp3 <
    li_boxplot_siqr_dois_anteriores | temp3 >
    ls_boxplot_siqr_dois_anteriores)] = 1

```



```

boxplot_ajustado_dois_anteriores = adjbox(temp19)

li_boxplot_ajustado_dois_anteriores =
  boxplot_ajustado_dois_anteriores$fence[1]

ls_boxplot_ajustado_dois_anteriores =
  boxplot_ajustado_dois_anteriores$fence[2]

metodo_boxplot_ajustado_dois_anteriores[which(temp3 <
  li_boxplot_ajustado_dois_anteriores | temp3 >
  ls_boxplot_ajustado_dois_anteriores)] = 1

}

else {

metodo_boxplot_padrao_dois_anteriores = metodo_boxplot_padrao

metodo_boxplot_siqr_dois_anteriores = metodo_boxplot_siqr

metodo_boxplot_ajustado_dois_anteriores = metodo_boxplot_ajustado

li_boxplot_padrao_dois_anteriores = li_boxplot_padrao

ls_boxplot_padrao_dois_anteriores = ls_boxplot_padrao

li_boxplot_siqr_dois_anteriores = li_boxplot_siqr

ls_boxplot_siqr_dois_anteriores = ls_boxplot_siqr

li_boxplot_ajustado_dois_anteriores = li_boxplot_ajustado

ls_boxplot_ajustado_dois_anteriores = ls_boxplot_ajustado

}

out_metodo_boxplot_padrao = c(out_metodo_boxplot_padrao ,
  metodo_boxplot_padrao)

out_metodo_boxplot_siqr = c(out_metodo_boxplot_siqr , metodo_boxplot_siqr
  )

```

```

out_metodo_boxplot_ajustado = c(out_metodo_boxplot_ajustado ,
                                metodo_boxplot_ajustado)

out_metodo_boxplot_padrao_mes_anterior = c(
    out_metodo_boxplot_padrao_mes_anterior ,
    metodo_boxplot_padrao_mes_anterior)

out_metodo_boxplot_siqr_mes_anterior = c(
    out_metodo_boxplot_siqr_mes_anterior ,
    metodo_boxplot_siqr_mes_anterior)

out_metodo_boxplot_ajustado_mes_anterior = c(
    out_metodo_boxplot_ajustado_mes_anterior ,
    metodo_boxplot_ajustado_mes_anterior)

out_metodo_boxplot_padrao_dois_anteriores = c(
    out_metodo_boxplot_padrao_dois_anteriores ,
    metodo_boxplot_padrao_dois_anteriores)

out_metodo_boxplot_siqr_dois_anteriores = c(
    out_metodo_boxplot_siqr_dois_anteriores ,
    metodo_boxplot_siqr_dois_anteriores)

out_metodo_boxplot_ajustado_dois_anteriores = c(
    out_metodo_boxplot_ajustado_dois_anteriores ,
    metodo_boxplot_ajustado_dois_anteriores)

out_metodo_boxplot_padrao_mesmo_mes_anos_anteriores = c(
    out_metodo_boxplot_padrao_mesmo_mes_anos_anteriores ,
    metodo_boxplot_padrao_mesmo_mes_anos_anteriores)

out_metodo_boxplot_siqr_mesmo_mes_anos_anteriores = c(
    out_metodo_boxplot_siqr_mesmo_mes_anos_anteriores ,
    metodo_boxplot_siqr_mesmo_mes_anos_anteriores)

out_metodo_boxplot_ajustado_mesmo_mes_anos_anteriores = c(
    out_metodo_boxplot_ajustado_mesmo_mes_anos_anteriores ,
    metodo_boxplot_ajustado_mesmo_mes_anos_anteriores)

out_metodo_boxplot_padrao_anteriores = c(
    out_metodo_boxplot_padrao_anteriores ,
    metodo_boxplot_padrao_anteriores)

```

```

out_metodo_boxplot_siqr_anteriores = c(
  out_metodo_boxplot_siqr_anteriores , metodo_boxplot_siqr_anteriores)

out_metodo_boxplot_ajustado_anteriores = c(
  out_metodo_boxplot_ajustado_anteriores ,
  metodo_boxplot_ajustado_anteriores)

sem_traco = temp4[temp4$rel != "-",]

alt_ou_desc = sem_traco$num_palt + sem_traco$num_pdesc

alt_ou_desc[alt_ou_desc > 0] = 1

outlier = c(outlier , alt_ou_desc)

cod_ins = c(cod_ins , sem_traco$cod_ins)

cod_inf = c(cod_inf , sem_traco$cod_inf)

cod_ins_inf = c(cod_ins_inf , sem_traco$cod_ins_inf)

regiao_uf = c(regiao_uf , sem_traco$regiao_uf)

ano_dcol = c(ano_dcol , sem_traco$ano_dcol)

mes_dcol = c(mes_dcol , sem_traco$mes_dcol)

rel = c(rel , as.numeric(levels(sem_traco$rel)[as.numeric(sem_traco$rel)
]))

ano = c(ano , i)

mes = c(mes , j)

ins = c(ins , k)

vet_li_boxplot_padrao = c(vet_li_boxplot_padrao , li_boxplot_padrao)

vet_ls_boxplot_padrao = c(vet_ls_boxplot_padrao , ls_boxplot_padrao)

vet_li_boxplot_siqr = c(vet_li_boxplot_siqr , li_boxplot_siqr)

vet_ls_boxplot_siqr = c(vet_ls_boxplot_siqr , ls_boxplot_siqr)

```

```

vet_li_boxplot_ajustado = c(vet_li_boxplot_ajustado , li_boxplot_ajustado
)

vet_ls_boxplot_ajustado = c(vet_ls_boxplot_ajustado , ls_boxplot_ajustado
)

vet_li_boxplot_padrao_mes_anterior = c(
    vet_li_boxplot_padrao_mes_anterior , li_boxplot_padrao_mes_anterior)

vet_ls_boxplot_padrao_mes_anterior = c(
    vet_ls_boxplot_padrao_mes_anterior , ls_boxplot_padrao_mes_anterior)

vet_li_boxplot_siqr_mes_anterior = c(vet_li_boxplot_siqr_mes_anterior ,
    li_boxplot_siqr_mes_anterior)

vet_ls_boxplot_siqr_mes_anterior = c(vet_ls_boxplot_siqr_mes_anterior ,
    ls_boxplot_siqr_mes_anterior)

vet_li_boxplot_ajustado_mes_anterior = c(
    vet_li_boxplot_ajustado_mes_anterior ,
    li_boxplot_ajustado_mes_anterior)

vet_ls_boxplot_ajustado_mes_anterior = c(
    vet_ls_boxplot_ajustado_mes_anterior ,
    ls_boxplot_ajustado_mes_anterior)

vet_li_boxplot_padrao_dois_anteriores = c(
    vet_li_boxplot_padrao_dois_anteriores ,
    li_boxplot_padrao_dois_anteriores)

vet_ls_boxplot_padrao_dois_anteriores = c(
    vet_ls_boxplot_padrao_dois_anteriores ,
    ls_boxplot_padrao_dois_anteriores)

vet_li_boxplot_siqr_dois_anteriores = c(
    vet_li_boxplot_siqr_dois_anteriores , li_boxplot_siqr_dois_anteriores)

vet_ls_boxplot_siqr_dois_anteriores = c(
    vet_ls_boxplot_siqr_dois_anteriores , ls_boxplot_siqr_dois_anteriores)

vet_li_boxplot_ajustado_dois_anteriores = c(
    vet_li_boxplot_ajustado_dois_anteriores ,
    li_boxplot_ajustado_dois_anteriores)

```

```

vet_ls_boxplot_ajustado_dois_anteriores = c(
  vet_ls_boxplot_ajustado_dois_anteriores ,
  ls_boxplot_ajustado_dois_anteriores )

vet_li_boxplot_padrao_mesmo_mes_anos_anteriores = c(
  vet_li_boxplot_padrao_mesmo_mes_anos_anteriores ,
  li_boxplot_padrao_mesmo_mes_anos_anteriores )

vet_ls_boxplot_padrao_mesmo_mes_anos_anteriores = c(
  vet_ls_boxplot_padrao_mesmo_mes_anos_anteriores ,
  ls_boxplot_padrao_mesmo_mes_anos_anteriores )

vet_li_boxplot_siqr_mesmo_mes_anos_anteriores = c(
  vet_li_boxplot_siqr_mesmo_mes_anos_anteriores ,
  li_boxplot_siqr_mesmo_mes_anos_anteriores )

vet_ls_boxplot_siqr_mesmo_mes_anos_anteriores = c(
  vet_ls_boxplot_siqr_mesmo_mes_anos_anteriores ,
  ls_boxplot_siqr_mesmo_mes_anos_anteriores )

vet_li_boxplot_ajustado_mesmo_mes_anos_anteriores = c(
  vet_li_boxplot_ajustado_mesmo_mes_anos_anteriores ,
  li_boxplot_ajustado_mesmo_mes_anos_anteriores )

vet_ls_boxplot_ajustado_mesmo_mes_anos_anteriores = c(
  vet_ls_boxplot_ajustado_mesmo_mes_anos_anteriores ,
  ls_boxplot_ajustado_mesmo_mes_anos_anteriores )

vet_li_boxplot_padrao_anteriores = c(vet_li_boxplot_padrao_anteriores ,
  li_boxplot_padrao_anteriores )

vet_ls_boxplot_padrao_anteriores = c(vet_ls_boxplot_padrao_anteriores ,
  ls_boxplot_padrao_anteriores )

vet_li_boxplot_siqr_anteriores = c(vet_li_boxplot_siqr_anteriores ,
  li_boxplot_siqr_anteriores )

vet_ls_boxplot_siqr_anteriores = c(vet_ls_boxplot_siqr_anteriores ,
  ls_boxplot_siqr_anteriores )

vet_li_boxplot_ajustado_anteriores = c(
  vet_li_boxplot_ajustado_anteriores , li_boxplot_ajustado_anteriores )

```

```

        vet_ls_boxplot_ajustado_anteriores = c(
            vet_ls_boxplot_ajustado_anteriores , ls_boxplot_ajustado_anteriores)

    }

}

resultado = data.frame(cod_ins ,

    cod_inf ,

    cod_ins_inf ,

    regioao_uf ,

    ano_dcol ,

    mes_dcol ,

    rel ,

    out_metodo_boxplot_padrao ,

    out_metodo_boxplot_siqr ,

    out_metodo_boxplot_ajustado ,

    out_metodo_boxplot_padrao_mes_anterior ,

    out_metodo_boxplot_siqr_mes_anterior ,

    out_metodo_boxplot_ajustado_mes_anterior ,

    out_metodo_boxplot_padrao_dois_anteriores ,

    out_metodo_boxplot_siqr_dois_anteriores ,

    out_metodo_boxplot_ajustado_dois_anteriores ,

    out_metodo_boxplot_padrao_mesmo_mes_anos_anteriores ,

    out_metodo_boxplot_siqr_mesmo_mes_anos_anteriores ,

    out_metodo_boxplot_ajustado_mesmo_mes_anos_anteriores ,

```

```

out_metodo_boxplot_padrao_anteriores ,

out_metodo_boxplot_siqr_anteriores ,

out_metodo_boxplot_ajustado_anteriores ,

outlier)

intervalos = data.frame(ano ,

mes ,

ins ,

vet_li_boxplot_padrao ,

vet_ls_boxplot_padrao ,

vet_li_boxplot_siqr ,

vet_ls_boxplot_siqr ,

vet_li_boxplot_ajustado ,

vet_ls_boxplot_ajustado ,

vet_li_boxplot_padrao_mes_anterior ,

vet_ls_boxplot_padrao_mes_anterior ,

vet_li_boxplot_siqr_mes_anterior ,

vet_ls_boxplot_siqr_mes_anterior ,

vet_li_boxplot_ajustado_mes_anterior ,

vet_ls_boxplot_ajustado_mes_anterior ,

vet_ls_boxplot_siqr_dois_anteriores ,

vet_li_boxplot_ajustado_dois_anteriores ,

vet_ls_boxplot_ajustado_dois_anteriores ,

```

```

    vet_li_boxplot_padrao_mesmo_mes_anos_anteriores ,
    vet_ls_boxplot_padrao_mesmo_mes_anos_anteriores ,
    vet_li_boxplot_siqr_mesmo_mes_anos_anteriores ,
    vet_ls_boxplot_siqr_mesmo_mes_anos_anteriores ,
    vet_li_boxplot_ajustado_mesmo_mes_anos_anteriores ,
    vet_ls_boxplot_ajustado_mesmo_mes_anos_anteriores ,
    vet_li_boxplot_padrao_anteriores ,
    vet_ls_boxplot_padrao_anteriores ,
    vet_li_boxplot_siqr_anteriores ,
    vet_ls_boxplot_siqr_anteriores ,
    vet_li_boxplot_ajustado_anteriores ,
    vet_ls_boxplot_ajustado_anteriores)

resultado_geral = list(resultado , intervalos)

return(resultado_geral)

}

```

# Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e Algoritmo de Tukey

```

aplica_metodos_outros_outliers = function(tabela , percorre , a = 0.05, cs) {

    resultado_cs = list()

    length(resultado_cs) = length(cs)

    intervalos_cs = list()

    length(intervalos_cs) = length(cs)

```



```

for (c in 1:length(cs)) {

  metodo_cercas_resistentes = c()

  metodo_quartil = c()

  metodo_quartil_mod = c()

  metodo_mad = c()

  metodo_algoritmo_tukey = c()

  metodo_cercas_resistentes_mes_anterior = c()

  metodo_quartil_mes_anterior = c()

  metodo_quartil_mod_mes_anterior = c()

  metodo_mad_mes_anterior = c()

  metodo_algoritmo_tukey_mes_anterior = c()

  metodo_cercas_resistentes_dois_anteriores = c()

  metodo_quartil_dois_anteriores = c()

  metodo_quartil_mod_dois_anteriores = c()

  metodo_mad_dois_anteriores = c()

  metodo_algoritmo_tukey_dois_anteriores = c()

  metodo_cercas_resistentes_mesmo_mes_anos_anteriores = c()

  metodo_quartil_mesmo_mes_anos_anteriores = c()

  metodo_quartil_mod_mesmo_mes_anos_anteriores = c()

  metodo_mad_mesmo_mes_anos_anteriores = c()

  metodo_algoritmo_tukey_mesmo_mes_anos_anteriores = c()

  metodo_cercas_resistentes_anteriores = c()

```

```
metodo_quartil_anteriores = c()

metodo_quartil_mod_anteriores = c()

metodo_mad_anteriores = c()

metodo_algoritmo_tukey_anteriores = c()

cod_ins = c()

cod_inf = c()

cod_ins_inf = c()

ano_dcol = c()

mes_dcol = c()

rel = c()

outlier = c()

ano = c()

mes = c()

ins = c()

vet_li_cercas_resistentes = c()

vet_ls_cercas_resistentes = c()

vet_li_quartil = c()

vet_ls_quartil = c()

vet_li_quartil_mod = c()

vet_ls_quartil_mod = c()

vet_li_mad = c()

vet_ls_mad = c()
```

```
vet_li_algoritmo_tukey = c()

vet_ls_algoritmo_tukey = c()

vet_li_cercas_resistentes_mes_anterior = c()

vet_ls_cercas_resistentes_mes_anterior = c()

vet_li_quartil_mes_anterior = c()

vet_ls_quartil_mes_anterior = c()

vet_li_quartil_mod_mes_anterior = c()

vet_ls_quartil_mod_mes_anterior = c()

vet_li_mad_mes_anterior = c()

vet_ls_mad_mes_anterior = c()

vet_li_algoritmo_tukey_mes_anterior = c()

vet_ls_algoritmo_tukey_mes_anterior = c()

vet_li_cercas_resistentes_dois_anteriores = c()

vet_ls_cercas_resistentes_dois_anteriores = c()

vet_li_quartil_dois_anteriores = c()

vet_ls_quartil_dois_anteriores = c()

vet_li_quartil_mod_dois_anteriores = c()

vet_ls_quartil_mod_dois_anteriores = c()

vet_li_mad_dois_anteriores = c()

vet_ls_mad_dois_anteriores = c()

vet_li_algoritmo_tukey_dois_anteriores = c()

vet_ls_algoritmo_tukey_dois_anteriores = c()
```

```

vet_li_cercas_resistentes_mesmo_mes_anos_anteriores = c()

vet_ls_cercas_resistentes_mesmo_mes_anos_anteriores = c()

vet_li_quartil_mesmo_mes_anos_anteriores = c()

vet_ls_quartil_mesmo_mes_anos_anteriores = c()

vet_li_quartil_mod_mesmo_mes_anos_anteriores = c()

vet_ls_quartil_mod_mesmo_mes_anos_anteriores = c()

vet_li_mad_mesmo_mes_anos_anteriores = c()

vet_ls_mad_mesmo_mes_anos_anteriores = c()

vet_li_algoritmo_tukey_mesmo_mes_anos_anteriores = c()

vet_ls_algoritmo_tukey_mesmo_mes_anos_anteriores = c()

vet_li_cercas_resistentes_anteriores = c()

vet_ls_cercas_resistentes_anteriores = c()

vet_li_quartil_anteriores = c()

vet_ls_quartil_anteriores = c()

vet_li_quartil_mod_anteriores = c()

vet_ls_quartil_mod_anteriores = c()

vet_li_mad_anteriores = c()

vet_ls_mad_anteriores = c()

vet_li_algoritmo_tukey_anteriores = c()

vet_ls_algoritmo_tukey_anteriores = c()

for (z in 1:nrow(percorre)) {

  print(paste("c = ", cs[c]))

```

```

print(paste("z = ", z))

i = percorre[z, 1]

print(paste("i = ", i))

j = percorre[z, 2]

print(paste("j = ", j))

k = percorre[z, 3]

print(paste("k = ", k))

temp3 = tabela[tabela$ano_dcol == i & tabela$mes_dcol == j &
  tabela$cod_ins == k,]$rel # Pega os relativos do ano/êms atual

temp4 = tabela[tabela$ano_dcol == i & tabela$mes_dcol == j &
  tabela$cod_ins == k,] # Pega a tabela com os relativos do ano/êms
  atual

if (j == 1) {

  temp7 = tabela[tabela$ano_dcol == i - 1 & tabela$mes_dcol == 12 &
    tabela$cod_ins == k,]$rel # Pega os relativos do êms anterior

  temp19_1 = tabela[tabela$ano_dcol == i - 1 & tabela$mes_dcol == 11 &
    tabela$cod_ins == k,]

  temp19_2 = tabela[tabela$ano_dcol == i - 1 & tabela$mes_dcol == 12 &
    tabela$cod_ins == k,]

  temp19 = rbind(temp19_1, temp19_2)$rel

}

else if (j == 2) {

  temp7 = tabela[tabela$ano_dcol == i & tabela$mes_dcol == j - 1 &
    tabela$cod_ins == k,]$rel # Pega os relativos do êms anterior

  temp19_1 = tabela[tabela$ano_dcol == i - 1 & tabela$mes_dcol == 12 &
    tabela$cod_ins == k,]

```

```

temp19_2 = tabela[tabela$ano_dcol == i & tabela$mes_dcol == j - 1 &
  tabela$cod_ins == k,]

temp19 = rbind(temp19_1, temp19_2)$rel
}

else {

  temp7 = tabela[tabela$ano_dcol == i & tabela$mes_dcol == j - 1 &
    tabela$cod_ins == k,]$rel # Pega os relativos do êms anterior

  temp19 = tabela[tabela$ano_dcol == i & tabela$mes_dcol == c(j - 2, j
    - 1) & tabela$cod_ins == k,]$rel

}

temp10 = tabela[tabela$ano_dcol < i & tabela$mes_dcol == j &
  tabela$cod_ins == k,]$rel # Pega os relativos do mesmo êms dos anos
  anteriores

temp11 = tabela[tabela$ano_dcol < i & tabela$cod_ins == k,] # Tabela
  com os relativos dos anos anteriores

temp12 = tabela[tabela$ano_dcol == i & tabela$mes_dcol < j &
  tabela$cod_ins == k,] # Tabela com os relativos dos meses
  anteriores do mesmo ano

temp13 = rbind(temp11, temp12) # Tabela todos os relativos anteriores

temp16 = rbind(temp11, temp12)$rel # Pega todos os relativos
  anteriores

if (length(temp3) >= 2) {

  qtdade_obs = length(temp3)

  metodo_cercas_resistentes = rep(0, qtdade_obs)

  li_cercas_resistentes = quantile(temp3, 0.25, na.rm = TRUE) - cs[c]
    * (quantile(temp3, 0.75, na.rm = TRUE) - quantile(temp3, 0.25, na
      .rm = TRUE))

```

```

ls_cercas_resistentes = quantile(temp3, 0.75, na.rm = TRUE) + cs[c]
  * (quantile(temp3, 0.75, na.rm = TRUE) - quantile(temp3, 0.25, na
    .rm = TRUE))

metodo_cercas_resistentes[which(temp3 < li_cercas_resistentes |
  temp3 > ls_cercas_resistentes)] = 1

metodo_quartil = rep(0, qtdade_obs)

li_quartil = median(temp3) - cs[c] * max((median(temp3) - quantile(
  temp3, 0.25, na.rm = TRUE)), abs(a * median(temp3)))

ls_quartil = median(temp3) + cs[c] * max((quantile(temp3, 0.75, na.
  rm = TRUE) - median(temp3)), abs(a * median(temp3)))

metodo_quartil[which(temp3 < li_quartil | temp3 > ls_quartil)] = 1

metodo_quartil_mod = rep(0, qtdade_obs)

li_quartil_mod = median(temp3) - max(cs[c] * (median(temp3) -
  quantile(temp3, 0.25, na.rm = TRUE)), abs(a * median(temp3)))

ls_quartil_mod = median(temp3) + max(cs[c] * (quantile(temp3, 0.75,
  na.rm = TRUE) - median(temp3)), abs(a * median(temp3)))

metodo_quartil_mod[which(temp3 < li_quartil_mod | temp3 >
  ls_quartil_mod)] = 1

metodo_mad = rep(0, qtdade_obs)

mediana = rep(median(temp3), qtdade_obs)

dif = abs(temp3 - mediana)

mad_dif = median(dif)

li_mad = median(temp3) - cs[c] * mad_dif

ls_mad = median(temp3) + cs[c] * mad_dif

metodo_mad[which(temp3 < li_mad | temp3 > ls_mad)] = 1

metodo_algoritmo_tukey = rep(0, qtdade_obs)

```

```

etapa1 = temp3[order(temp3)]

pos_1 = which(etapa1 == 1)

if (length(pos_1) == 0) {

  etapa2 = etapa1

  etapa3 = etapa2[c((arredonda(0.025 * length(etapa2)) + 1) : (
    length(etapa2) - arredonda(0.025 * length(etapa2))))]

  if ((length(etapa3[etapa3 < median(etapa3)]) == 0) | (length(
    etapa3[etapa3 > median(etapa3)]) == 0)) {

    li_algoritmo_tukey = "-"

    ls_algoritmo_tukey = "-"

    metodo_algoritmo_tukey = rep(9, qtdade_obs)

  }

  else {

    xd = mean(etapa3)

    xl = mean(etapa3[etapa3 < median(etapa3)])

    xu = mean(etapa3[etapa3 > median(etapa3)])

    delta_xl = xd - xl

    delta_xu = xu - xd

    li_algoritmo_tukey = xd - cs[c] * delta_xl

    ls_algoritmo_tukey = xd + cs[c] * delta_xu

    metodo_algoritmo_tukey[which(temp3 < li_algoritmo_tukey | temp3
      > ls_algoritmo_tukey)] = 1

  }

}

```



```

else if ((length(pos_1) > 0) & (length(pos_1) < length(temp3))) {

  etapa2 = etapa1[-c(pos_1)]

  if (length(etapa2) == 1) {

    li_algoritmo_tukey = "-"

    ls_algoritmo_tukey = "-"

    metodo_algoritmo_tukey = rep(9, qtdade_obs)

  }

  else {

    etapa3 = etapa2[c((arredonda(0.025 * length(etapa2)) + 1) : (
      length(etapa2) - arredonda(0.025 * length(etapa2)))))]

    if ((length(etapa3[etapa3 < median(etapa3)]) == 0) | (length(
      etapa3[etapa3 > median(etapa3)]) == 0)) {

      li_algoritmo_tukey = "-"

      ls_algoritmo_tukey = "-"

      metodo_algoritmo_tukey = rep(9, qtdade_obs)

    }

    else {

      xd = mean(etapa3)

      xl = mean(etapa3[etapa3 < median(etapa3)])

      xu = mean(etapa3[etapa3 > median(etapa3)])

      delta_xl = xd - xl

      delta_xu = xu - xd

      li_algoritmo_tukey = xd - cs[c] * delta_xl

```

```

ls_algoritmo_tukey = xd + cs[c] * delta_xu

metodo_algoritmo_tukey[which(temp3 < li_algoritmo_tukey |
  temp3 > ls_algoritmo_tukey)] = 1

}

}

}

else {

  li_algoritmo_tukey = "-"

  ls_algoritmo_tukey = "-"

  metodo_algoritmo_tukey = rep(9, qtdade_obs)

}

metodo_cercas_resistentes_mes_anterior = rep(0, qtdade_obs)

metodo_quartil_mes_anterior = rep(0, qtdade_obs)

metodo_quartil_mod_mes_anterior = rep(0, qtdade_obs)

metodo_mad_mes_anterior = rep(0, qtdade_obs)

metodo_algoritmo_tukey_mes_anterior = rep(0, qtdade_obs)

if (length(temp7) >= 2) {

  li_cercas_resistentes_mes_anterior = quantile(temp7, 0.25, na.rm =
    TRUE) - cs[c] * (quantile(temp7, 0.75, na.rm = TRUE) -
    quantile(temp7, 0.25, na.rm = TRUE))

  ls_cercas_resistentes_mes_anterior = quantile(temp7, 0.75, na.rm =
    TRUE) + cs[c] * (quantile(temp7, 0.75, na.rm = TRUE) -
    quantile(temp7, 0.25, na.rm = TRUE))

  metodo_cercas_resistentes_mes_anterior[which(temp3 <
    li_cercas_resistentes_mes_anterior | temp3 >

```

```

ls_cercas_resistentes_mes_anterior]] = 1

li_quartil_mes_anterior = median(temp7) - cs[c] * max((median(
  temp7) - quantile(temp7, 0.25, na.rm = TRUE)), abs(a * median(
  temp7)))

ls_quartil_mes_anterior = median(temp7) + cs[c] * max((quantile(
  temp7, 0.75, na.rm = TRUE) - median(temp7)), abs(a * median(
  temp7)))

metodo_quartil_mes_anterior[which(temp3 < li_quartil_mes_anterior
  | temp3 > ls_quartil_mes_anterior)] = 1

li_quartil_mod_mes_anterior = median(temp7) - max(cs[c] * (median(
  temp7) - quantile(temp7, 0.25, na.rm = TRUE)), abs(a * median(
  temp7)))

ls_quartil_mod_mes_anterior = median(temp7) + max(cs[c] * (
  quantile(temp7, 0.75, na.rm = TRUE) - median(temp7)), abs(a *
  median(temp7)))

metodo_quartil_mod_mes_anterior[which(temp3 <
  li_quartil_mod_mes_anterior | temp3 >
  ls_quartil_mod_mes_anterior)] = 1

mediana = rep(median(temp7), length(temp7))

dif = abs(temp7 - mediana)

mad_dif = median(dif)

li_mad_mes_anterior = median(temp7) - cs[c] * mad_dif

ls_mad_mes_anterior = median(temp7) + cs[c] * mad_dif

metodo_mad_mes_anterior[which(temp3 < li_mad_mes_anterior | temp3
  > ls_mad_mes_anterior)] = 1

etapa1 = temp7[order(temp7)]

pos_1 = which(etapa1 == 1)

if (length(pos_1) == 0) {

```

```

etapa2 = etapa1

etapa3 = etapa2[c((arredonda(0.025 * length(etapa2)) + 1) : (
  length(etapa2) - arredonda(0.025 * length(etapa2))))]

if ((length(etapa3[etapa3 < median(etapa3)]) == 0) | (length(
  etapa3[etapa3 > median(etapa3)]) == 0)) {

  li_algoritmo_tukey_mes_anterior = "-"

  ls_algoritmo_tukey_mes_anterior = "-"

  metodo_algoritmo_tukey_mes_anterior = rep(9, qtdade_obs)

}

else {

  xd = mean(etapa3)

  xl = mean(etapa3[etapa3 < median(etapa3)])

  xu = mean(etapa3[etapa3 > median(etapa3)])

  delta_xl = xd - xl

  delta_xu = xu - xd

  li_algoritmo_tukey_mes_anterior = xd - cs[c] * delta_xl

  ls_algoritmo_tukey_mes_anterior = xd + cs[c] * delta_xu

  metodo_algoritmo_tukey_mes_anterior[which(temp3 <
    li_algoritmo_tukey_mes_anterior | temp3 >
    ls_algoritmo_tukey_mes_anterior)] = 1

}

}

else if ((length(pos_1) > 0) & (length(pos_1) < length(temp7))) {

  etapa2 = etapa1[-c(pos_1)]

```

```

if (length(etapa2) == 1) {

  li_algoritmo_tukey_mes_anterior = "-"

  ls_algoritmo_tukey_mes_anterior = "-"

  metodo_algoritmo_tukey_mes_anterior = rep(9, qtdade_obs)

}

else {

  etapa3 = etapa2[c((arredonda(0.025 * length(etapa2)) + 1) : (
    length(etapa2) - arredonda(0.025 * length(etapa2))))]

  if ((length(etapa3[etapa3 < median(etapa3)]) == 0) | (length(
    etapa3[etapa3 > median(etapa3)]) == 0)) {

    li_algoritmo_tukey_mes_anterior = "-"

    ls_algoritmo_tukey_mes_anterior = "-"

    metodo_algoritmo_tukey_mes_anterior = rep(9, qtdade_obs)

  }

  else {

    xd = mean(etapa3)

    xl = mean(etapa3[etapa3 < median(etapa3)])

    xu = mean(etapa3[etapa3 > median(etapa3)])

    delta_xl = xd - xl

    delta_xu = xu - xd

    li_algoritmo_tukey_mes_anterior = xd - cs[c] * delta_xl

    ls_algoritmo_tukey_mes_anterior = xd + cs[c] * delta_xu

    metodo_algoritmo_tukey_mes_anterior[which(temp3 <
      li_algoritmo_tukey_mes_anterior | temp3 >

```

```

        ls_algoritmo_tukey_mes_anterior)] = 1

    }

}

}

else {

    li_algoritmo_tukey_mes_anterior = "-"

    ls_algoritmo_tukey_mes_anterior = "-"

    metodo_algoritmo_tukey_mes_anterior = rep(9, qtdade_obs)

}

}

else {

    metodo_cercas_resistentes_mes_anterior = metodo_cercas_resistentes

    metodo_quartil_mes_anterior = metodo_quartil

    metodo_quartil_mod_mes_anterior = metodo_quartil_mod

    metodo_mad_mes_anterior = metodo_mad

    metodo_algoritmo_tukey_mes_anterior = metodo_algoritmo_tukey

    li_cercas_resistentes_mes_anterior = li_cercas_resistentes

    ls_cercas_resistentes_mes_anterior = ls_cercas_resistentes

    li_quartil_mes_anterior = li_quartil

    ls_quartil_mes_anterior = ls_quartil

    li_quartil_mod_mes_anterior = li_quartil_mod

    ls_quartil_mod_mes_anterior = ls_quartil_mod

```

```

li_mad_mes_anterior = li_mad

ls_mad_mes_anterior = ls_mad

li_algoritmo_tukey_mes_anterior = li_algoritmo_tukey

ls_algoritmo_tukey_mes_anterior = ls_algoritmo_tukey

}

metodo_cercas_resistentes_mesmo_mes_anos_anteriores = rep(0,
  qtdade_obs)

metodo_quartil_mesmo_mes_anos_anteriores = rep(0, qtdade_obs)

metodo_quartil_mod_mesmo_mes_anos_anteriores = rep(0, qtdade_obs)

metodo_mad_mesmo_mes_anos_anteriores = rep(0, qtdade_obs)

metodo_algoritmo_tukey_mesmo_mes_anos_anteriores = rep(0, qtdade_obs
  )

if (length(temp10) >= 2) {

  li_cercas_resistentes_mesmo_mes_anos_anteriores = quantile(temp10,
    0.25, na.rm = TRUE) - cs[c] * (quantile(temp10, 0.75, na.rm =
    TRUE) - quantile(temp10, 0.25, na.rm = TRUE))

  ls_cercas_resistentes_mesmo_mes_anos_anteriores = quantile(temp10,
    0.75, na.rm = TRUE) + cs[c] * (quantile(temp10, 0.75, na.rm =
    TRUE) - quantile(temp10, 0.25, na.rm = TRUE))

  metodo_cercas_resistentes_mesmo_mes_anos_anteriores[which(temp3 <
    li_cercas_resistentes_mesmo_mes_anos_anteriores | temp3 >
    ls_cercas_resistentes_mesmo_mes_anos_anteriores)] = 1

  li_quartil_mesmo_mes_anos_anteriores = median(temp10) - cs[c] *
    max((median(temp10) - quantile(temp10, 0.25, na.rm = TRUE)),
    abs(a * median(temp10)))

  ls_quartil_mesmo_mes_anos_anteriores = median(temp10) + cs[c] *
    max((quantile(temp10, 0.75, na.rm = TRUE) - median(temp10)),
    abs(a * median(temp10)))

```

```

metodo_quartil_mesmo_mes_anos_anteriores[which(temp3 <
  li_quartil_mesmo_mes_anos_anteriores | temp3 >
  ls_quartil_mesmo_mes_anos_anteriores)] = 1

li_quartil_mod_mesmo_mes_anos_anteriores = median(temp10) - max(cs
  [c] * (median(temp10) - quantile(temp10, 0.25, na.rm = TRUE)),
  abs(a * median(temp10)))

ls_quartil_mod_mesmo_mes_anos_anteriores = median(temp10) + max(cs
  [c] * (quantile(temp10, 0.75, na.rm = TRUE) - median(temp10)),
  abs(a * median(temp10)))

metodo_quartil_mod_mesmo_mes_anos_anteriores[which(temp3 <
  li_quartil_mod_mesmo_mes_anos_anteriores | temp3 >
  ls_quartil_mod_mesmo_mes_anos_anteriores)] = 1

mediana = rep(median(temp10), length(temp10))

dif = abs(temp10 - mediana)

mad_dif = median(dif)

li_mad_mesmo_mes_anos_anteriores = median(temp10) - cs[c] *
  mad_dif

ls_mad_mesmo_mes_anos_anteriores = median(temp10) + cs[c] *
  mad_dif

metodo_mad_mesmo_mes_anos_anteriores[which(temp3 <
  li_mad_mesmo_mes_anos_anteriores | temp3 >
  ls_mad_mesmo_mes_anos_anteriores)] = 1

etapa1 = temp10[order(temp10)]

pos_1 = which(etapa1 == 1)

if (length(pos_1) == 0) {

  etapa2 = etapa1

  etapa3 = etapa2[c((arredonda(0.025 * length(etapa2)) + 1) : (
    length(etapa2) - arredonda(0.025 * length(etapa2)))))]

```



```

if ((length(etapa3[etapa3 < median(etapa3)]) == 0) | (length(
  etapa3[etapa3 > median(etapa3)]) == 0)) {

  li_algoritmo_tukey_mesmo_mes_anos_anteriores = "-"

  ls_algoritmo_tukey_mesmo_mes_anos_anteriores = "-"

  metodo_algoritmo_tukey_mesmo_mes_anos_anteriores = rep(9,
    qtdade_obs)

}

else {

  xd = mean(etapa3)

  xl = mean(etapa3[etapa3 < median(etapa3)])

  xu = mean(etapa3[etapa3 > median(etapa3)])

  delta_xl = xd - xl

  delta_xu = xu - xd

  li_algoritmo_tukey_mesmo_mes_anos_anteriores = xd - cs[c] *
    delta_xl

  ls_algoritmo_tukey_mesmo_mes_anos_anteriores = xd + cs[c] *
    delta_xu

  metodo_algoritmo_tukey_mesmo_mes_anos_anteriores[which(temp3 <
    li_algoritmo_tukey_mesmo_mes_anos_anteriores | temp3 >
    ls_algoritmo_tukey_mesmo_mes_anos_anteriores)] = 1

}

}

else if ((length(pos_1) > 0) & (length(pos_1) < length(temp10))) {

  etapa2 = etapa1[-c(pos_1)]

  if (length(etapa2) == 1) {

```

```

li_algoritmo_tukey_mesmo_mes_anos_anteriores = "-"

ls_algoritmo_tukey_mesmo_mes_anos_anteriores = "-"

metodo_algoritmo_tukey_mesmo_mes_anos_anteriores = rep(9,
  qtdade_obs)

}

else {

  etapa3 = etapa2[c((arredonda(0.025 * length(etapa2)) + 1) : (
    length(etapa2) - arredonda(0.025 * length(etapa2))))]

  if ((length(etapa3[etapa3 < median(etapa3)]) == 0) | (length(
    etapa3[etapa3 > median(etapa3)]) == 0)) {

    li_algoritmo_tukey_mesmo_mes_anos_anteriores = "-"

    ls_algoritmo_tukey_mesmo_mes_anos_anteriores = "-"

    metodo_algoritmo_tukey_mesmo_mes_anos_anteriores = rep(9,
      qtdade_obs)

  }

  else {

    xd = mean(etapa3)

    xl = mean(etapa3[etapa3 < median(etapa3)])

    xu = mean(etapa3[etapa3 > median(etapa3)])

    delta_xl = xd - xl

    delta_xu = xu - xd

    li_algoritmo_tukey_mesmo_mes_anos_anteriores = xd - cs[c] *
      delta_xl

    ls_algoritmo_tukey_mesmo_mes_anos_anteriores = xd + cs[c] *
      delta_xu
  }
}

```

```

        metodo_algoritmo_tukey_mesmo_mes_anos_anteriores[ which(temp3
        < li_algoritmo_tukey_mesmo_mes_anos_anteriores | temp3 >
        ls_algoritmo_tukey_mesmo_mes_anos_anteriores)] = 1

    }

}

}

else {

    li_algoritmo_tukey_mesmo_mes_anos_anteriores = "-"

    ls_algoritmo_tukey_mesmo_mes_anos_anteriores = "-"

    metodo_algoritmo_tukey_mesmo_mes_anos_anteriores = rep(9,
        qtdade_obs)

}

}

else {

    metodo_cercas_resistentes_mesmo_mes_anos_anteriores =
        metodo_cercas_resistentes

    metodo_quartil_mesmo_mes_anos_anteriores = metodo_quartil

    metodo_quartil_mod_mesmo_mes_anos_anteriores = metodo_quartil_mod

    metodo_mad_mesmo_mes_anos_anteriores = metodo_mad

    metodo_algoritmo_tukey_mesmo_mes_anos_anteriores =
        metodo_algoritmo_tukey

    li_cercas_resistentes_mesmo_mes_anos_anteriores =
        li_cercas_resistentes

    ls_cercas_resistentes_mesmo_mes_anos_anteriores =
        ls_cercas_resistentes

    li_quartil_mesmo_mes_anos_anteriores = li_quartil

```

```

ls_quartil_mesmo_mes_anos_anteriores = ls_quartil

li_quartil_mod_mesmo_mes_anos_anteriores = li_quartil_mod

ls_quartil_mod_mesmo_mes_anos_anteriores = ls_quartil_mod

li_mad_mesmo_mes_anos_anteriores = li_mad

ls_mad_mesmo_mes_anos_anteriores = ls_mad

li_algoritmo_tukey_mesmo_mes_anos_anteriores = li_algoritmo_tukey

ls_algoritmo_tukey_mesmo_mes_anos_anteriores = ls_algoritmo_tukey
}

metodo_cercas_resistentes_anteriores = rep(0, qtdade_obs)

metodo_quartil_anteriores = rep(0, qtdade_obs)

metodo_quartil_mod_anteriores = rep(0, qtdade_obs)

metodo_mad_anteriores = rep(0, qtdade_obs)

metodo_algoritmo_tukey_anteriores = rep(0, qtdade_obs)

if (length(temp16) >= 2) {

  li_cercas_resistentes_anteriores = quantile(temp16, 0.25, na.rm =
    TRUE) - cs[c] * (quantile(temp16, 0.75, na.rm = TRUE) -
    quantile(temp16, 0.25, na.rm = TRUE))

  ls_cercas_resistentes_anteriores = quantile(temp16, 0.75, na.rm =
    TRUE) + cs[c] * (quantile(temp16, 0.75, na.rm = TRUE) -
    quantile(temp16, 0.25, na.rm = TRUE))

  metodo_cercas_resistentes_anteriores[which(temp3 <
    li_cercas_resistentes_anteriores | temp3 >
    ls_cercas_resistentes_anteriores)] = 1

  li_quartil_anteriores = median(temp16) - cs[c] * max((median(
    temp16) - quantile(temp16, 0.25, na.rm = TRUE)), abs(a * median
    (temp16)))

```

```

ls_quartil_anteriores = median(temp16) + cs[c] * max((quantile(
  temp16, 0.75, na.rm = TRUE) - median(temp16)), abs(a * median(
  temp16)))

metodo_quartil_anteriores[which(temp3 < li_quartil_anteriores |
  temp3 > ls_quartil_anteriores)] = 1

li_quartil_mod_anteriores = median(temp16) - max(cs[c] * (median(
  temp16) - quantile(temp16, 0.25, na.rm = TRUE)), abs(a * median(
  temp16)))

ls_quartil_mod_anteriores = median(temp16) + max(cs[c] * (quantile(
  temp16, 0.75, na.rm = TRUE) - median(temp16)), abs(a * median(
  temp16)))

metodo_quartil_mod_anteriores[which(temp3 <
  li_quartil_mod_anteriores | temp3 > ls_quartil_mod_anteriores)]
  = 1

mediana = rep(median(temp16), length(temp16))

dif_mes = abs(temp16 - mediana)

mad_dif = median(dif_mes)

li_mad_anteriores = median(temp16) - cs[c] * mad_dif

ls_mad_anteriores = median(temp16) + cs[c] * mad_dif

metodo_mad_anteriores[which(temp3 < li_mad_anteriores | temp3 >
  ls_mad_anteriores)] = 1

etapa1 = temp16[order(temp16)]

pos_1 = which(etapa1 == 1)

if (length(pos_1) == 0) {

  etapa2 = etapa1

  etapa3 = etapa2[c((arredonda(0.025 * length(etapa2)) + 1) : (
    length(etapa2) - arredonda(0.025 * length(etapa2)))))]

```

```

if ((length(etapa3[etapa3 < median(etapa3)]) == 0) | (length(
  etapa3[etapa3 > median(etapa3)]) == 0)) {

  li_algoritmo_tukey_anteriores = "-"

  ls_algoritmo_tukey_anteriores = "-"

  metodo_algoritmo_tukey_anteriores = rep(9, qtdade_obs)

}

else {

  xd = mean(etapa3)

  xl = mean(etapa3[etapa3 < median(etapa3)])

  xu = mean(etapa3[etapa3 > median(etapa3)])

  delta_xl = xd - xl

  delta_xu = xu - xd

  li_algoritmo_tukey_anteriores = xd - cs[c] * delta_xl

  ls_algoritmo_tukey_anteriores = xd + cs[c] * delta_xu

  metodo_algoritmo_tukey_anteriores[which(temp3 <
    li_algoritmo_tukey_anteriores | temp3 >
    ls_algoritmo_tukey_anteriores)] = 1

}

}

else if ((length(pos_1) > 0) & (length(pos_1) < length(temp16))) {

  etapa2 = etapa1[-c(pos_1)]

  if (length(etapa2) == 1) {

    li_algoritmo_tukey_anteriores = "-"

    ls_algoritmo_tukey_anteriores = "-"

```

```

    metodo_algoritmo_tukey_anteriores = rep(9, qtdade_obs)

}

else {

    etapa3 = etapa2[c((arredonda(0.025 * length(etapa2)) + 1) : (
        length(etapa2) - arredonda(0.025 * length(etapa2)))))]

    if ((length(etapa3[etapa3 < median(etapa3)]) == 0) | (length(
        etapa3[etapa3 > median(etapa3)]) == 0)) {

        li_algoritmo_tukey_anteriores = "-"

        ls_algoritmo_tukey_anteriores = "-"

        metodo_algoritmo_tukey_anteriores = rep(9, qtdade_obs)

    }

    else {

        xd = mean(etapa3)

        xl = mean(etapa3[etapa3 < median(etapa3)])

        xu = mean(etapa3[etapa3 > median(etapa3)])

        delta_xl = xd - xl

        delta_xu = xu - xd

        li_algoritmo_tukey_anteriores = xd - cs[c] * delta_xl

        ls_algoritmo_tukey_anteriores = xd + cs[c] * delta_xu

        metodo_algoritmo_tukey_anteriores[which(temp3 <
            li_algoritmo_tukey_anteriores | temp3 >
            ls_algoritmo_tukey_anteriores)] = 1

    }

}

```

```

    }

    else {

        li_algoritmo_tukey_anteriores = "-"

        ls_algoritmo_tukey_anteriores = "-"

        metodo_algoritmo_tukey_anteriores = rep(9, qtdade_obs)

    }

}

else {

    metodo_cercas_resistentes_anteriores = metodo_cercas_resistentes

    metodo_quartil_anteriores = metodo_quartil

    metodo_quartil_mod_anteriores = metodo_quartil_mod

    metodo_mad_anteriores = metodo_mad

    metodo_algoritmo_tukey_anteriores = metodo_algoritmo_tukey

    li_cercas_resistentes_anteriores = li_cercas_resistentes

    ls_cercas_resistentes_anteriores = ls_cercas_resistentes

    li_quartil_anteriores = li_quartil

    ls_quartil_anteriores = ls_quartil

    li_quartil_mod_anteriores = li_quartil_mod

    ls_quartil_mod_anteriores = ls_quartil_mod

    li_mad_anteriores = li_mad

    ls_mad_anteriores = ls_mad

    li_algoritmo_tukey_anteriores = li_algoritmo_tukey

```



```

ls_algoritmo_tukey_anteriores = ls_algoritmo_tukey

}

metodo_cercas_resistentes_dois_anteriores = rep(0, qtdade_obs)

metodo_quartil_dois_anteriores = rep(0, qtdade_obs)

metodo_quartil_mod_dois_anteriores = rep(0, qtdade_obs)

metodo_mad_dois_anteriores = rep(0, qtdade_obs)

metodo_algoritmo_tukey_dois_anteriores = rep(0, qtdade_obs)

if (length(temp19) >= 2) {

  li_cercas_resistentes_dois_anteriores = quantile(temp19, 0.25, na.rm = TRUE) - cs[c] * (quantile(temp19, 0.75, na.rm = TRUE) - quantile(temp19, 0.25, na.rm = TRUE))

  ls_cercas_resistentes_dois_anteriores = quantile(temp19, 0.75, na.rm = TRUE) + cs[c] * (quantile(temp19, 0.75, na.rm = TRUE) - quantile(temp19, 0.25, na.rm = TRUE))

  metodo_cercas_resistentes_dois_anteriores[which(temp3 < li_cercas_resistentes_dois_anteriores | temp3 > ls_cercas_resistentes_dois_anteriores)] = 1

  li_quartil_dois_anteriores = median(temp19) - cs[c] * max((median(temp19) - quantile(temp19, 0.25, na.rm = TRUE)), abs(a * median(temp19)))

  ls_quartil_dois_anteriores = median(temp19) + cs[c] * max((quantile(temp19, 0.75, na.rm = TRUE) - median(temp19)), abs(a * median(temp19)))

  metodo_quartil_dois_anteriores[which(temp3 < li_quartil_dois_anteriores | temp3 > ls_quartil_dois_anteriores)] = 1

  li_quartil_mod_dois_anteriores = median(temp19) - max(cs[c] * (median(temp19) - quantile(temp19, 0.25, na.rm = TRUE)), abs(a * median(temp19)))

```

```

ls_quartil_mod_dois_anteriores = median(temp19) + max(cs[c] * (
  quantile(temp19, 0.75, na.rm = TRUE) - median(temp19)), abs(a *
  median(temp19)))

metodo_quartil_mod_dois_anteriores[which(temp3 <
  li_quartil_mod_dois_anteriores | temp3 >
  ls_quartil_mod_dois_anteriores)] = 1

mediana = rep(median(temp19), length(temp19))

dif = abs(temp19 - mediana)

mad_dif = median(dif)

li_mad_dois_anteriores = median(temp19) - cs[c] * mad_dif

ls_mad_dois_anteriores = median(temp19) + cs[c] * mad_dif

metodo_mad_dois_anteriores[which(temp3 < li_mad_dois_anteriores |
  temp3 > ls_mad_dois_anteriores)] = 1

etapa1 = temp19[order(temp19)]

pos_1 = which(etapa1 == 1)

if (length(pos_1) == 0) {

  etapa2 = etapa1

  etapa3 = etapa2[c((arredonda(0.025 * length(etapa2)) + 1) : (
    length(etapa2) - arredonda(0.025 * length(etapa2))))]

  if ((length(etapa3[etapa3 < median(etapa3)]) == 0) | (length(
    etapa3[etapa3 > median(etapa3)]) == 0)) {

    li_algoritmo_tukey_dois_anteriores = "-"

    ls_algoritmo_tukey_dois_anteriores = "-"

    metodo_algoritmo_tukey_dois_anteriores = rep(9, qtdade_obs)

  }

```

```

else {

  xd = mean(etapa3)

  xl = mean(etapa3[etapa3 < median(etapa3)])

  xu = mean(etapa3[etapa3 > median(etapa3)])

  delta_xl = xd - xl

  delta_xu = xu - xd

  li_algoritmo_tukey_dois_anteriores = xd - cs[c] * delta_xl

  ls_algoritmo_tukey_dois_anteriores = xd + cs[c] * delta_xu

  metodo_algoritmo_tukey_dois_anteriores[which(temp3 <
    li_algoritmo_tukey_dois_anteriores | temp3 >
    ls_algoritmo_tukey_dois_anteriores)] = 1

}

}

else if ((length(pos_1) > 0) & (length(pos_1) < length(temp19))) {

  etapa2 = etapa1[-c(pos_1)]

  if (length(etapa2) == 1) {

    li_algoritmo_tukey_dois_anteriores = "-"

    ls_algoritmo_tukey_dois_anteriores = "-"

    metodo_algoritmo_tukey_dois_anteriores = rep(9, qtdade_obs)

  }

  else {

    etapa3 = etapa2[c((arredonda(0.025 * length(etapa2)) + 1) : (
      length(etapa2) - arredonda(0.025 * length(etapa2))))]

```

```

if ((length(etapa3[etapa3 < median(etapa3)]) == 0) | (length(
  etapa3[etapa3 > median(etapa3)]) == 0)) {

  li_algoritmo_tukey_dois_anteriores = "-"

  ls_algoritmo_tukey_dois_anteriores = "-"

  metodo_algoritmo_tukey_dois_anteriores = rep(9, qtdade_obs)

}

else {

  xd = mean(etapa3)

  xl = mean(etapa3[etapa3 < median(etapa3)])

  xu = mean(etapa3[etapa3 > median(etapa3)])

  delta_xl = xd - xl

  delta_xu = xu - xd

  li_algoritmo_tukey_dois_anteriores = xd - cs[c] * delta_xl

  ls_algoritmo_tukey_dois_anteriores = xd + cs[c] * delta_xu

  metodo_algoritmo_tukey_dois_anteriores[which(temp3 <
    li_algoritmo_tukey_dois_anteriores | temp3 >
    ls_algoritmo_tukey_dois_anteriores)] = 1

}

}

}

else {

  li_algoritmo_tukey_dois_anteriores = "-"

  ls_algoritmo_tukey_dois_anteriores = "-"

  metodo_algoritmo_tukey_dois_anteriores = rep(9, qtdade_obs)

```

```

    }

}

else {

    metodo_cercas_resistentes_dois_anteriores =
        metodo_cercas_resistentes

    metodo_quartil_dois_anteriores = metodo_quartil

    metodo_quartil_mod_dois_anteriores = metodo_quartil_mod

    metodo_mad_dois_anteriores = metodo_mad

    metodo_algoritmo_tukey_dois_anteriores = metodo_algoritmo_tukey

    li_cercas_resistentes_dois_anteriores = li_cercas_resistentes

    ls_cercas_resistentes_dois_anteriores = ls_cercas_resistentes

    li_quartil_dois_anteriores = li_quartil

    ls_quartil_dois_anteriores = ls_quartil

    li_quartil_mod_dois_anteriores = li_quartil_mod

    ls_quartil_mod_dois_anteriores = ls_quartil_mod

    li_mad_dois_anteriores = li_mad

    ls_mad_dois_anteriores = ls_mad

    li_algoritmo_tukey_dois_anteriores = li_algoritmo_tukey

    ls_algoritmo_tukey_dois_anteriores = ls_algoritmo_tukey

}

metodo_cercas_resistentes = c(metodo_cercas_resistentes ,
    metodo_cercas_resistentes)

metodo_quartil = c(metodo_quartil , metodo_quartil)

```

```

metodo_quartil_mod = c(metodo_quartil_mod , metodo_quartil_mod)

metodo_mad = c(metodo_mad , metodo_mad)

metodo_algoritmo_tukey = c(metodo_algoritmo_tukey ,
                             metodo_algoritmo_tukey)

metodo_cercas_resistentes_mes_anterior = c(
    metodo_cercas_resistentes_mes_anterior ,
    metodo_cercas_resistentes_mes_anterior)

metodo_quartil_mes_anterior = c(metodo_quartil_mes_anterior ,
                                  metodo_quartil_mes_anterior)

metodo_quartil_mod_mes_anterior = c(metodo_quartil_mod_mes_anterior ,
                                      metodo_quartil_mod_mes_anterior)

metodo_mad_mes_anterior = c(metodo_mad_mes_anterior ,
                              metodo_mad_mes_anterior)

metodo_algoritmo_tukey_mes_anterior = c(
    metodo_algoritmo_tukey_mes_anterior ,
    metodo_algoritmo_tukey_mes_anterior)

metodo_cercas_resistentes_dois_anteriores = c(
    metodo_cercas_resistentes_dois_anteriores ,
    metodo_cercas_resistentes_dois_anteriores)

metodo_quartil_dois_anteriores = c(metodo_quartil_dois_anteriores ,
                                     metodo_quartil_dois_anteriores)

metodo_quartil_mod_dois_anteriores = c(
    metodo_quartil_mod_dois_anteriores ,
    metodo_quartil_mod_dois_anteriores)

metodo_mad_dois_anteriores = c(metodo_mad_dois_anteriores ,
                                 metodo_mad_dois_anteriores)

metodo_algoritmo_tukey_dois_anteriores = c(
    metodo_algoritmo_tukey_dois_anteriores ,
    metodo_algoritmo_tukey_dois_anteriores)

```

```

metodo_cercas_resistentes_mesmo_mes_anos_anteriores = c(
  metodo_cercas_resistentes_mesmo_mes_anos_anteriores ,
  metodo_cercas_resistentes_mesmo_mes_anos_anteriores )

metodo_quartil_mesmo_mes_anos_anteriores = c(
  metodo_quartil_mesmo_mes_anos_anteriores ,
  metodo_quartil_mesmo_mes_anos_anteriores )

metodo_quartil_mod_mesmo_mes_anos_anteriores = c(
  metodo_quartil_mod_mesmo_mes_anos_anteriores ,
  metodo_quartil_mod_mesmo_mes_anos_anteriores )

metodo_mad_mesmo_mes_anos_anteriores = c(
  metodo_mad_mesmo_mes_anos_anteriores ,
  metodo_mad_mesmo_mes_anos_anteriores )

metodo_algoritmo_tukey_mesmo_mes_anos_anteriores = c(
  metodo_algoritmo_tukey_mesmo_mes_anos_anteriores ,
  metodo_algoritmo_tukey_mesmo_mes_anos_anteriores )

metodo_cercas_resistentes_anteriores = c(
  metodo_cercas_resistentes_anteriores ,
  metodo_cercas_resistentes_anteriores )

metodo_quartil_anteriores = c(metodo_quartil_anteriores ,
  metodo_quartil_anteriores )

metodo_quartil_mod_anteriores = c(metodo_quartil_mod_anteriores ,
  metodo_quartil_mod_anteriores )

metodo_mad_anteriores = c(metodo_mad_anteriores ,
  metodo_mad_anteriores )

metodo_algoritmo_tukey_anteriores = c(
  metodo_algoritmo_tukey_anteriores ,
  metodo_algoritmo_tukey_anteriores )

cod_ins = c(cod_ins , temp4$cod_ins)

cod_inf = c(cod_inf , temp4$cod_inf)

cod_ins_inf = c(cod_ins_inf , temp4$cod_ins_inf)

ano_dcol = c(ano_dcol , temp4$ano_dcol)

```

```

mes_dcol = c(mes_dcol, temp4$mes_dcol)

rel = c(rel, temp3)

outlier = c(outlier, temp4$outlier)

ano = c(ano, i)

mes = c(mes, j)

ins = c(ins, k)

vet_li_cercas_resistentes = c(vet_li_cercas_resistentes,
                               li_cercas_resistentes)

vet_ls_cercas_resistentes = c(vet_ls_cercas_resistentes,
                               ls_cercas_resistentes)

vet_li_quartil = c(vet_li_quartil, li_quartil)

vet_ls_quartil = c(vet_ls_quartil, ls_quartil)

vet_li_quartil_mod = c(vet_li_quartil_mod, li_quartil_mod)

vet_ls_quartil_mod = c(vet_ls_quartil_mod, ls_quartil_mod)

vet_li_mad = c(vet_li_mad, li_mad)

vet_ls_mad = c(vet_ls_mad, ls_mad)

vet_li_algoritmo_tukey = c(vet_li_algoritmo_tukey,
                           li_algoritmo_tukey)

vet_ls_algoritmo_tukey = c(vet_ls_algoritmo_tukey,
                           ls_algoritmo_tukey)

vet_li_cercas_resistentes_mes_anterior = c(
    vet_li_cercas_resistentes_mes_anterior,
    li_cercas_resistentes_mes_anterior)

vet_ls_cercas_resistentes_mes_anterior = c(
    vet_ls_cercas_resistentes_mes_anterior,
    ls_cercas_resistentes_mes_anterior)

```



```

vet_li_quartil_mes_anterior = c(vet_li_quartil_mes_anterior ,
    li_quartil_mes_anterior)

vet_ls_quartil_mes_anterior = c(vet_ls_quartil_mes_anterior ,
    ls_quartil_mes_anterior)

vet_li_quartil_mod_mes_anterior = c(vet_li_quartil_mod_mes_anterior ,
    li_quartil_mod_mes_anterior)

vet_ls_quartil_mod_mes_anterior = c(vet_ls_quartil_mod_mes_anterior ,
    ls_quartil_mod_mes_anterior)

vet_li_mad_mes_anterior = c(vet_li_mad_mes_anterior ,
    li_mad_mes_anterior)

vet_ls_mad_mes_anterior = c(vet_ls_mad_mes_anterior ,
    ls_mad_mes_anterior)

vet_li_algoritmo_tukey_mes_anterior = c(
    vet_li_algoritmo_tukey_mes_anterior ,
    li_algoritmo_tukey_mes_anterior)

vet_ls_algoritmo_tukey_mes_anterior = c(
    vet_ls_algoritmo_tukey_mes_anterior ,
    ls_algoritmo_tukey_mes_anterior)

vet_li_cercas_resistentes_dois_anteriores = c(
    vet_li_cercas_resistentes_dois_anteriores ,
    li_cercas_resistentes_dois_anteriores)

vet_ls_cercas_resistentes_dois_anteriores = c(
    vet_ls_cercas_resistentes_dois_anteriores ,
    ls_cercas_resistentes_dois_anteriores)

vet_li_quartil_dois_anteriores = c(vet_li_quartil_dois_anteriores ,
    li_quartil_dois_anteriores)

vet_ls_quartil_dois_anteriores = c(vet_ls_quartil_dois_anteriores ,
    ls_quartil_dois_anteriores)

vet_li_quartil_mod_dois_anteriores = c(
    vet_li_quartil_mod_dois_anteriores ,
    li_quartil_mod_dois_anteriores)

```

```

vet_ls_quartil_mod_dois_anteriores = c(
  vet_ls_quartil_mod_dois_anteriores ,
  ls_quartil_mod_dois_anteriores )

vet_li_mad_dois_anteriores = c(vet_li_mad_dois_anteriores ,
  li_mad_dois_anteriores )

vet_ls_mad_dois_anteriores = c(vet_ls_mad_dois_anteriores ,
  ls_mad_dois_anteriores )

vet_li_algoritmo_tukey_dois_anteriores = c(
  vet_li_algoritmo_tukey_dois_anteriores ,
  li_algoritmo_tukey_dois_anteriores )

vet_ls_algoritmo_tukey_dois_anteriores = c(
  vet_ls_algoritmo_tukey_dois_anteriores ,
  ls_algoritmo_tukey_dois_anteriores )

vet_li_cercas_resistentes_mesmo_mes_anos_anteriores = c(
  vet_li_cercas_resistentes_mesmo_mes_anos_anteriores ,
  li_cercas_resistentes_mesmo_mes_anos_anteriores )

vet_ls_cercas_resistentes_mesmo_mes_anos_anteriores = c(
  vet_ls_cercas_resistentes_mesmo_mes_anos_anteriores ,
  ls_cercas_resistentes_mesmo_mes_anos_anteriores )

vet_li_quartil_mesmo_mes_anos_anteriores = c(
  vet_li_quartil_mesmo_mes_anos_anteriores ,
  li_quartil_mesmo_mes_anos_anteriores )

vet_ls_quartil_mesmo_mes_anos_anteriores = c(
  vet_ls_quartil_mesmo_mes_anos_anteriores ,
  ls_quartil_mesmo_mes_anos_anteriores )

vet_li_quartil_mod_mesmo_mes_anos_anteriores = c(
  vet_li_quartil_mod_mesmo_mes_anos_anteriores ,
  li_quartil_mod_mesmo_mes_anos_anteriores )

vet_ls_quartil_mod_mesmo_mes_anos_anteriores = c(
  vet_ls_quartil_mod_mesmo_mes_anos_anteriores ,
  ls_quartil_mod_mesmo_mes_anos_anteriores )

```

```

vet_li_mad_mesmo_mes_anos_anteriores = c(
  vet_li_mad_mesmo_mes_anos_anteriores ,
  li_mad_mesmo_mes_anos_anteriores )

vet_ls_mad_mesmo_mes_anos_anteriores = c(
  vet_ls_mad_mesmo_mes_anos_anteriores ,
  ls_mad_mesmo_mes_anos_anteriores )

vet_li_algoritmo_tukey_mesmo_mes_anos_anteriores = c(
  vet_li_algoritmo_tukey_mesmo_mes_anos_anteriores ,
  li_algoritmo_tukey_mesmo_mes_anos_anteriores )

vet_ls_algoritmo_tukey_mesmo_mes_anos_anteriores = c(
  vet_ls_algoritmo_tukey_mesmo_mes_anos_anteriores ,
  ls_algoritmo_tukey_mesmo_mes_anos_anteriores )

vet_li_cercas_resistentes_anteriores = c(
  vet_li_cercas_resistentes_anteriores ,
  li_cercas_resistentes_anteriores )

vet_ls_cercas_resistentes_anteriores = c(
  vet_ls_cercas_resistentes_anteriores ,
  ls_cercas_resistentes_anteriores )

vet_li_quartil_anteriores = c(vet_li_quartil_anteriores ,
  li_quartil_anteriores )

vet_ls_quartil_anteriores = c(vet_ls_quartil_anteriores ,
  ls_quartil_anteriores )

vet_li_quartil_mod_anteriores = c(vet_li_quartil_mod_anteriores ,
  li_quartil_mod_anteriores )

vet_ls_quartil_mod_anteriores = c(vet_ls_quartil_mod_anteriores ,
  ls_quartil_mod_anteriores )

vet_li_mad_anteriores = c(vet_li_mad_anteriores , li_mad_anteriores)

vet_ls_mad_anteriores = c(vet_ls_mad_anteriores , ls_mad_anteriores)

vet_li_algoritmo_tukey_anteriores = c(
  vet_li_algoritmo_tukey_anteriores , li_algoritmo_tukey_anteriores)

```

```

        vet_ls_algoritmo_tukey_anteriores = c(
            vet_ls_algoritmo_tukey_anteriores , ls_algoritmo_tukey_anteriores)

    }

}

resultado_c = data.frame(cod_ins ,

    cod_inf ,

    cod_ins_inf ,

    ano_dcol ,

    mes_dcol ,

    rel ,

    metodo_cercas_resistentes ,

    metodo_quartil ,

    metodo_quartil_mod ,

    metodo_mad ,

    metodo_algoritmo_tukey ,

    metodo_cercas_resistentes_mes_anterior ,

    metodo_quartil_mes_anterior ,

    metodo_quartil_mod_mes_anterior ,

    metodo_mad_mes_anterior ,

    metodo_algoritmo_tukey_mes_anterior ,

    metodo_cercas_resistentes_dois_anteriores ,

    metodo_quartil_dois_anteriores ,

    metodo_quartil_mod_dois_anteriores ,

```

```

metodo_mad_dois_anteriores ,

metodo_algoritmo_tukey_dois_anteriores ,

metodo_cercas_resistentes_mesmo_mes_anos_anteriores ,

metodo_quartil_mesmo_mes_anos_anteriores ,

metodo_quartil_mod_mesmo_mes_anos_anteriores ,

metodo_mad_mesmo_mes_anos_anteriores ,

metodo_algoritmo_tukey_mesmo_mes_anos_anteriores ,

metodo_cercas_resistentes_anteriores ,

metodo_quartil_anteriores ,

metodo_quartil_mod_anteriores ,

metodo_mad_anteriores ,

metodo_algoritmo_tukey_anteriores ,

outlier)

resultado_cs[[c]] = resultado_c

intervalos_c = data.frame(ano,

    mes,

    ins,

    vet_li_cercas_resistentes ,

    vet_ls_cercas_resistentes ,

    vet_li_quartil ,

    vet_ls_quartil ,

    vet_li_quartil_mod ,

```

vet\_ls\_quartil\_mod ,  
  
vet\_li\_mad ,  
  
vet\_ls\_mad ,  
  
vet\_li\_algoritmo\_tukey ,  
  
vet\_ls\_algoritmo\_tukey ,  
  
vet\_li\_cercas\_resistentes\_mes\_anterior ,  
  
vet\_ls\_cercas\_resistentes\_mes\_anterior ,  
  
vet\_li\_quartil\_mes\_anterior ,  
  
vet\_ls\_quartil\_mes\_anterior ,  
  
vet\_li\_quartil\_mod\_mes\_anterior ,  
  
vet\_ls\_quartil\_mod\_mes\_anterior ,  
  
vet\_li\_mad\_mes\_anterior ,  
  
vet\_ls\_mad\_mes\_anterior ,  
  
vet\_li\_algoritmo\_tukey\_mes\_anterior ,  
  
vet\_ls\_algoritmo\_tukey\_mes\_anterior ,  
  
vet\_li\_cercas\_resistentes\_dois\_anteriores ,  
  
vet\_ls\_cercas\_resistentes\_dois\_anteriores ,  
  
vet\_li\_quartil\_dois\_anteriores ,  
  
vet\_ls\_quartil\_dois\_anteriores ,  
  
vet\_li\_quartil\_mod\_dois\_anteriores ,  
  
vet\_ls\_quartil\_mod\_dois\_anteriores ,  
  
vet\_li\_mad\_dois\_anteriores ,

vet\_ls\_mad\_dois\_anteriores ,  
vet\_li\_algoritmo\_tukey\_dois\_anteriores ,  
vet\_ls\_algoritmo\_tukey\_dois\_anteriores ,  
vet\_li\_cercas\_resistentes\_mesmo\_mes\_anos\_anteriores ,  
vet\_ls\_cercas\_resistentes\_mesmo\_mes\_anos\_anteriores ,  
vet\_li\_quartil\_mesmo\_mes\_anos\_anteriores ,  
vet\_ls\_quartil\_mesmo\_mes\_anos\_anteriores ,  
vet\_li\_quartil\_mod\_mesmo\_mes\_anos\_anteriores ,  
vet\_ls\_quartil\_mod\_mesmo\_mes\_anos\_anteriores ,  
vet\_li\_mad\_mesmo\_mes\_anos\_anteriores ,  
vet\_ls\_mad\_mesmo\_mes\_anos\_anteriores ,  
vet\_li\_algoritmo\_tukey\_mesmo\_mes\_anos\_anteriores ,  
vet\_ls\_algoritmo\_tukey\_mesmo\_mes\_anos\_anteriores ,  
vet\_li\_cercas\_resistentes\_anteriores ,  
vet\_ls\_cercas\_resistentes\_anteriores ,  
vet\_li\_quartil\_anteriores ,  
vet\_ls\_quartil\_anteriores ,  
vet\_li\_quartil\_mod\_anteriores ,  
vet\_ls\_quartil\_mod\_anteriores ,  
vet\_li\_mad\_anteriores ,  
vet\_ls\_mad\_anteriores ,  
vet\_li\_algoritmo\_tukey\_anteriores ,

```

        vet_ls_algoritmo_tukey_anteriores)

    intervalos_cs[[c]] = intervalos_c

}

resultado = list(resultado_cs, intervalos_cs)

names(resultado[[1]]) = paste("c =", cs)

names(resultado[[2]]) = paste("c =", cs)

return(resultado)

}

```



**APÊNDICE C - Tabelas de Resultados dos Métodos de Cercas Resistentes, do Quartil, do Quartil Modificado, do Desvio Mediano Absoluto e do Algoritmo de Tukey para Cada um dos Conjuntos de Dados Considerados**

**Tabela 15 Desempenho do Método de Cercas Resistentes (MR) no Município do Rio de Janeiro**

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	24914	89	0,74167	280
c = 0,1	20543	81	0,675	254
c = 0,2	17576	76	0,63333	231
c = 0,3	15078	70	0,58333	215
c = 0,4	13039	63	0,525	207
c = 0,5	11087	59	0,49167	188
c = 0,6	9587	52	0,43333	184
c = 0,7	8303	47	0,39167	177
c = 0,8	7310	46	0,38333	159
c = 0,9	6420	43	0,35833	149
c = 1	5634	41	0,34167	137
c = 1,1	4994	39	0,325	128
c = 1,2	4471	39	0,325	115
c = 1,3	4057	38	0,31667	107
c = 1,4	3727	37	0,30833	101
c = 1,5	3387	36	0,3	94
c = 1,6	3066	33	0,275	93
c = 1,7	2787	32	0,26667	87
c = 1,8	2575	31	0,25833	83
c = 1,9	2354	28	0,23333	84
c = 2	2166	28	0,23333	77
c = 2,1	2007	26	0,21667	77
c = 2,2	1845	25	0,20833	74
c = 2,3	1720	22	0,18333	78
c = 2,4	1603	22	0,18333	73
c = 2,5	1514	19	0,15833	80
c = 2,6	1406	18	0,15	78
c = 2,7	1326	17	0,14167	78
c = 2,8	1262	17	0,14167	74
c = 2,9	1190	17	0,14167	70
c = 3	1143	17	0,14167	67
c = 3,1	1070	16	0,13333	67
c = 3,2	1022	16	0,13333	64
c = 3,3	979	16	0,13333	61
c = 3,4	938	16	0,13333	59
c = 3,5	887	16	0,13333	55
c = 3,6	857	16	0,13333	54
c = 3,7	834	16	0,13333	52
c = 3,8	802	16	0,13333	50
c = 3,9	775	16	0,13333	48
c = 4	750	16	0,13333	47

Tabela 16 Desempenho do Método do Quartil (MR) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	45714	115	0,95833	398
c = 0,1	43137	112	0,93333	385
c = 0,2	40902	111	0,925	368
c = 0,3	38833	107	0,89167	363
c = 0,4	36909	105	0,875	352
c = 0,5	34987	102	0,85	343
c = 0,6	33153	102	0,85	325
c = 0,7	31315	101	0,84167	310
c = 0,8	29516	99	0,825	298
c = 0,9	27570	94	0,78333	293
c = 1	23309	88	0,73333	265
c = 1,1	20932	84	0,7	249
c = 1,2	19113	81	0,675	236
c = 1,3	17508	79	0,65833	222
c = 1,4	16072	76	0,63333	211
c = 1,5	14845	73	0,60833	203
c = 1,6	13699	70	0,58333	196
c = 1,7	12653	67	0,55833	189
c = 1,8	11689	64	0,53333	183
c = 1,9	10880	63	0,525	173
c = 2	9756	60	0,5	163
c = 2,1	8928	52	0,43333	172
c = 2,2	8290	51	0,425	163
c = 2,3	7660	49	0,40833	156
c = 2,4	7153	48	0,4	149
c = 2,5	6608	45	0,375	147
c = 2,6	6151	45	0,375	137
c = 2,7	5716	43	0,35833	133
c = 2,8	5328	43	0,35833	124
c = 2,9	4960	42	0,35	118
c = 3	4598	42	0,35	109
c = 3,1	4295	41	0,34167	105
c = 3,2	4036	39	0,325	103
c = 3,3	3768	39	0,325	97
c = 3,4	3522	38	0,31667	93
c = 3,5	3290	37	0,30833	89
c = 3,6	3063	36	0,3	85
c = 3,7	2871	31	0,25833	93
c = 3,8	2714	31	0,25833	88
c = 3,9	2538	31	0,25833	82
c = 4	2359	29	0,24167	81

Tabela 17 Desempenho do Método do Quartil Modificado (MR) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	31535	106	0,88333	298
c = 0,1	31535	106	0,88333	298
c = 0,2	31530	106	0,88333	297
c = 0,3	31466	105	0,875	300
c = 0,4	31241	103	0,85833	303
c = 0,5	30816	101	0,84167	305
c = 0,6	30167	101	0,84167	299
c = 0,7	29314	100	0,83333	293
c = 0,8	28284	98	0,81667	289
c = 0,9	27045	94	0,78333	288
c = 1	23309	88	0,73333	265
c = 1,1	21193	84	0,7	252
c = 1,2	19549	82	0,68333	238
c = 1,3	18100	80	0,66667	226
c = 1,4	16809	77	0,64167	218
c = 1,5	15696	74	0,61667	212
c = 1,6	14654	71	0,59167	206
c = 1,7	13674	68	0,56667	201
c = 1,8	12792	65	0,54167	197
c = 1,9	12046	64	0,53333	188
c = 2	10834	61	0,50833	178
c = 2,1	9968	53	0,44167	188
c = 2,2	9389	52	0,43333	181
c = 2,3	8825	50	0,41667	177
c = 2,4	8325	49	0,40833	170
c = 2,5	7802	46	0,38333	170
c = 2,6	7380	46	0,38333	160
c = 2,7	6978	44	0,36667	159
c = 2,8	6586	44	0,36667	150
c = 2,9	6235	43	0,35833	145
c = 3	5880	43	0,35833	137
c = 3,1	5562	42	0,35	132
c = 3,2	5289	40	0,33333	132
c = 3,3	5034	40	0,33333	126
c = 3,4	4784	39	0,325	123
c = 3,5	4560	38	0,31667	120
c = 3,6	4353	37	0,30833	118
c = 3,7	4158	33	0,275	126
c = 3,8	3992	33	0,275	121
c = 3,9	3822	32	0,26667	119
c = 4	3620	31	0,25833	117

Tabela 18 Desempenho do Método do Desvio Mediano Absoluto (MR) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	45714	115	0,95833	398
c = 0,1	43285	113	0,94167	383
c = 0,2	41006	111	0,925	369
c = 0,3	38972	109	0,90833	358
c = 0,4	37042	106	0,88333	349
c = 0,5	35099	101	0,84167	348
c = 0,6	33307	99	0,825	336
c = 0,7	31569	97	0,80833	325
c = 0,8	29784	92	0,76667	324
c = 0,9	28140	92	0,76667	306
c = 1	23570	84	0,7	281
c = 1,1	21511	82	0,68333	262
c = 1,2	19951	77	0,64167	259
c = 1,3	18496	76	0,63333	243
c = 1,4	17253	72	0,6	240
c = 1,5	16006	69	0,575	232
c = 1,6	14837	67	0,55833	221
c = 1,7	13790	62	0,51667	222
c = 1,8	12871	60	0,5	215
c = 1,9	12034	60	0,5	201
c = 2	11178	59	0,49167	189
c = 2,1	10464	57	0,475	184
c = 2,2	9792	56	0,46667	175
c = 2,3	9163	51	0,425	180
c = 2,4	8627	50	0,41667	173
c = 2,5	8105	50	0,41667	162
c = 2,6	7660	48	0,4	160
c = 2,7	7221	45	0,375	160
c = 2,8	6797	42	0,35	162
c = 2,9	6450	42	0,35	154
c = 3	6125	42	0,35	146
c = 3,1	5814	42	0,35	138
c = 3,2	5539	42	0,35	132
c = 3,3	5236	40	0,33333	131
c = 3,4	4993	40	0,33333	125
c = 3,5	4738	37	0,30833	128
c = 3,6	4533	35	0,29167	130
c = 3,7	4329	35	0,29167	124
c = 3,8	4146	35	0,29167	118
c = 3,9	3978	34	0,28333	117
c = 4	3834	34	0,28333	113

Tabela 19 Desempenho do Algoritmo de Tukey (MR) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	49718	120	1	414
c = 0,1	46689	116	0,96667	402
c = 0,2	43650	113	0,94167	386
c = 0,3	40553	111	0,925	365
c = 0,4	37646	106	0,88333	355
c = 0,5	34762	101	0,84167	344
c = 0,6	31995	98	0,81667	326
c = 0,7	29194	97	0,80833	301
c = 0,8	26392	93	0,775	284
c = 0,9	23843	90	0,75	265
c = 1	20270	81	0,675	250
c = 1,1	18094	77	0,64167	235
c = 1,2	16022	72	0,6	223
c = 1,3	14219	70	0,58333	203
c = 1,4	12611	65	0,54167	194
c = 1,5	11185	60	0,5	186
c = 1,6	9813	57	0,475	172
c = 1,7	8598	55	0,45833	156
c = 1,8	7556	50	0,41667	151
c = 1,9	6620	46	0,38333	144
c = 2	5770	44	0,36667	131
c = 2,1	5025	39	0,325	129
c = 2,2	4440	36	0,3	123
c = 2,3	3896	35	0,29167	111
c = 2,4	3427	31	0,25833	111
c = 2,5	3044	29	0,24167	105
c = 2,6	2695	28	0,23333	96
c = 2,7	2412	26	0,21667	93
c = 2,8	2127	26	0,21667	82
c = 2,9	1889	26	0,21667	73
c = 3	1666	23	0,19167	72
c = 3,1	1499	21	0,175	71
c = 3,2	1338	21	0,175	64
c = 3,3	1199	21	0,175	57
c = 3,4	1081	20	0,16667	54
c = 3,5	966	20	0,16667	48
c = 3,6	870	18	0,15	48
c = 3,7	793	17	0,14167	47
c = 3,8	713	17	0,14167	42
c = 3,9	658	14	0,11667	47
c = 4	613	13	0,10833	47

Tabela 20 Desempenho do Método de Cercas Resistentes (MA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	34162	103	0,85833	332
c = 0,1	31797	100	0,83333	318
c = 0,2	29600	96	0,8	308
c = 0,3	27460	94	0,78333	292
c = 0,4	25432	94	0,78333	271
c = 0,5	23532	92	0,76667	256
c = 0,6	21772	89	0,74167	245
c = 0,7	20099	83	0,69167	242
c = 0,8	18630	81	0,675	230
c = 0,9	17273	76	0,63333	227
c = 1	16019	74	0,61667	216
c = 1,1	14879	74	0,61667	201
c = 1,2	13818	72	0,6	192
c = 1,3	12831	70	0,58333	183
c = 1,4	11920	68	0,56667	175
c = 1,5	11102	64	0,53333	173
c = 1,6	10271	61	0,50833	168
c = 1,7	9542	60	0,5	159
c = 1,8	8927	60	0,5	149
c = 1,9	8328	57	0,475	146
c = 2	7807	55	0,45833	142
c = 2,1	7307	51	0,425	143
c = 2,2	6818	49	0,40833	139
c = 2,3	6405	46	0,38333	139
c = 2,4	6028	45	0,375	134
c = 2,5	5655	45	0,375	126
c = 2,6	5360	43	0,35833	125
c = 2,7	5025	42	0,35	120
c = 2,8	4731	42	0,35	113
c = 2,9	4470	42	0,35	106
c = 3	4249	40	0,33333	106
c = 3,1	4009	39	0,325	103
c = 3,2	3802	39	0,325	97
c = 3,3	3618	39	0,325	93
c = 3,4	3422	38	0,31667	90
c = 3,5	3244	37	0,30833	88
c = 3,6	3103	35	0,29167	89
c = 3,7	2962	34	0,28333	87
c = 3,8	2819	32	0,26667	88
c = 3,9	2708	32	0,26667	85
c = 4	2600	32	0,26667	81

Tabela 21 Desempenho do Método do Quartil (MA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	47703	119	0,99167	401
c = 0,1	44779	113	0,94167	396
c = 0,2	38480	113	0,94167	341
c = 0,3	34254	110	0,91667	311
c = 0,4	30598	100	0,83333	306
c = 0,5	27108	99	0,825	274
c = 0,6	24035	97	0,80833	248
c = 0,7	21194	92	0,76667	230
c = 0,8	18637	90	0,75	207
c = 0,9	16415	86	0,71667	191
c = 1	14507	83	0,69167	175
c = 1,1	12898	81	0,675	159
c = 1,2	11453	76	0,63333	151
c = 1,3	10084	74	0,61667	136
c = 1,4	8909	72	0,6	124
c = 1,5	7838	71	0,59167	110
c = 1,6	6989	68	0,56667	103
c = 1,7	6209	66	0,55	94
c = 1,8	5550	63	0,525	88
c = 1,9	4915	58	0,48333	85
c = 2	4314	55	0,45833	78
c = 2,1	3837	53	0,44167	72
c = 2,2	3463	49	0,40833	71
c = 2,3	3119	47	0,39167	66
c = 2,4	2820	44	0,36667	64
c = 2,5	2562	43	0,35833	60
c = 2,6	2318	42	0,35	55
c = 2,7	2112	41	0,34167	52
c = 2,8	1925	39	0,325	49
c = 2,9	1788	37	0,30833	48
c = 3	1677	36	0,3	47
c = 3,1	1548	35	0,29167	44
c = 3,2	1441	34	0,28333	42
c = 3,3	1341	32	0,26667	42
c = 3,4	1234	31	0,25833	40
c = 3,5	1131	28	0,23333	40
c = 3,6	1044	28	0,23333	37
c = 3,7	980	28	0,23333	35
c = 3,8	928	26	0,21667	36
c = 3,9	868	26	0,21667	33
c = 4	801	26	0,21667	31

Tabela 22 Desempenho do Método do Quartil Modificado (MA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	38497	113	0,94167	341
c = 0,1	24064	97	0,80833	248
c = 0,2	14566	84	0,7	173
c = 0,3	14566	84	0,7	173
c = 0,4	14566	84	0,7	173
c = 0,5	14565	84	0,7	173
c = 0,6	14561	84	0,7	173
c = 0,7	14552	84	0,7	173
c = 0,8	14545	84	0,7	173
c = 0,9	14537	83	0,69167	175
c = 1	14507	83	0,69167	175
c = 1,1	14481	83	0,69167	174
c = 1,2	14438	83	0,69167	174
c = 1,3	14377	82	0,68333	175
c = 1,4	14305	82	0,68333	174
c = 1,5	14211	82	0,68333	173
c = 1,6	14103	82	0,68333	172
c = 1,7	13978	82	0,68333	170
c = 1,8	13833	82	0,68333	169
c = 1,9	13676	81	0,675	169
c = 2	13465	79	0,65833	170
c = 2,1	13267	78	0,65	170
c = 2,2	13042	78	0,65	167
c = 2,3	12805	77	0,64167	166
c = 2,4	12554	77	0,64167	163
c = 2,5	12305	76	0,63333	162
c = 2,6	12063	75	0,625	161
c = 2,7	11814	75	0,625	158
c = 2,8	11568	74	0,61667	156
c = 2,9	11267	74	0,61667	152
c = 3	10984	74	0,61667	148
c = 3,1	10715	74	0,61667	145
c = 3,2	10434	73	0,60833	143
c = 3,3	10165	71	0,59167	143
c = 3,4	9913	70	0,58333	142
c = 3,5	9641	68	0,56667	142
c = 3,6	9346	67	0,55833	139
c = 3,7	9099	65	0,54167	140
c = 3,8	8864	65	0,54167	136
c = 3,9	8623	63	0,525	137
c = 4	8368	62	0,51667	135



Tabela 23 Desempenho do Método do Desvio Mediano Absoluto (MA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	47703	119	0,99167	401
c = 0,1	46329	117	0,975	396
c = 0,2	44882	114	0,95	394
c = 0,3	43530	113	0,94167	385
c = 0,4	42158	113	0,94167	373
c = 0,5	40737	111	0,925	367
c = 0,6	39346	110	0,91667	358
c = 0,7	38014	108	0,9	352
c = 0,8	36753	106	0,88333	347
c = 0,9	35447	104	0,86667	341
c = 1	33694	104	0,86667	324
c = 1,1	32459	101	0,84167	321
c = 1,2	31286	100	0,83333	313
c = 1,3	30206	98	0,81667	308
c = 1,4	29105	96	0,8	303
c = 1,5	28024	93	0,775	301
c = 1,6	27014	93	0,775	290
c = 1,7	26021	92	0,76667	283
c = 1,8	24991	91	0,75833	275
c = 1,9	24065	91	0,75833	264
c = 2	23172	87	0,725	266
c = 2,1	22211	85	0,70833	261
c = 2,2	21354	83	0,69167	257
c = 2,3	20567	82	0,68333	251
c = 2,4	19775	81	0,675	244
c = 2,5	19043	80	0,66667	238
c = 2,6	18325	79	0,65833	232
c = 2,7	17676	77	0,64167	230
c = 2,8	17058	74	0,61667	231
c = 2,9	16455	72	0,6	229
c = 3	15868	72	0,6	220
c = 3,1	15293	71	0,59167	215
c = 3,2	14736	71	0,59167	208
c = 3,3	14209	70	0,58333	203
c = 3,4	13741	69	0,575	199
c = 3,5	13259	67	0,55833	198
c = 3,6	12809	67	0,55833	191
c = 3,7	12311	64	0,53333	192
c = 3,8	11879	64	0,53333	186
c = 3,9	11462	62	0,51667	185
c = 4	11052	60	0,5	184

Tabela 24 Desempenho do Algoritmo de Tukey (MA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	49696	120	1	414
c = 0,1	47522	118	0,98333	403
c = 0,2	45343	115	0,95833	394
c = 0,3	43306	114	0,95	380
c = 0,4	41290	110	0,91667	375
c = 0,5	39271	109	0,90833	360
c = 0,6	37311	106	0,88333	352
c = 0,7	35433	105	0,875	337
c = 0,8	33612	102	0,85	330
c = 0,9	31823	102	0,85	312
c = 1	30139	101	0,84167	298
c = 1,1	28560	101	0,84167	283
c = 1,2	27086	96	0,8	282
c = 1,3	25725	94	0,78333	274
c = 1,4	24329	94	0,78333	259
c = 1,5	23009	93	0,775	247
c = 1,6	21762	91	0,75833	239
c = 1,7	20560	87	0,725	236
c = 1,8	19437	86	0,71667	226
c = 1,9	18426	84	0,7	219
c = 2	17477	82	0,68333	213
c = 2,1	16541	80	0,66667	207
c = 2,2	15662	78	0,65	201
c = 2,3	14786	75	0,625	197
c = 2,4	14026	70	0,58333	200
c = 2,5	13307	69	0,575	193
c = 2,6	12638	68	0,56667	186
c = 2,7	11988	66	0,55	182
c = 2,8	11370	66	0,55	172
c = 2,9	10788	65	0,54167	166
c = 3	10227	63	0,525	162
c = 3,1	9709	62	0,51667	157
c = 3,2	9208	60	0,5	153
c = 3,3	8750	57	0,475	154
c = 3,4	8311	56	0,46667	148
c = 3,5	7885	51	0,425	155
c = 3,6	7461	51	0,425	146
c = 3,7	7077	50	0,41667	142
c = 3,8	6716	49	0,40833	137
c = 3,9	6374	47	0,39167	136
c = 4	6060	47	0,39167	129

Tabela 25 Desempenho do Método de Cercas Resistentes (2MA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	30676	103	0,85833	298
c = 0,1	27900	100	0,83333	279
c = 0,2	25316	97	0,80833	261
c = 0,3	22953	91	0,75833	252
c = 0,4	20881	86	0,71667	243
c = 0,5	19014	83	0,69167	229
c = 0,6	17328	79	0,65833	219
c = 0,7	15750	76	0,63333	207
c = 0,8	14337	72	0,6	199
c = 0,9	13113	69	0,575	190
c = 1	11934	65	0,54167	184
c = 1,1	10910	62	0,51667	176
c = 1,2	9982	60	0,5	166
c = 1,3	9113	59	0,49167	154
c = 1,4	8415	59	0,49167	143
c = 1,5	7748	57	0,475	136
c = 1,6	7168	53	0,44167	135
c = 1,7	6620	53	0,44167	125
c = 1,8	6184	50	0,41667	124
c = 1,9	5739	47	0,39167	122
c = 2	5337	45	0,375	119
c = 2,1	4988	44	0,36667	113
c = 2,2	4665	44	0,36667	106
c = 2,3	4365	41	0,34167	106
c = 2,4	4085	38	0,31667	108
c = 2,5	3835	36	0,3	107
c = 2,6	3588	35	0,29167	103
c = 2,7	3415	35	0,29167	98
c = 2,8	3251	34	0,28333	96
c = 2,9	3070	32	0,26667	96
c = 3	2914	30	0,25	97
c = 3,1	2768	29	0,24167	95
c = 3,2	2620	27	0,225	97
c = 3,3	2499	27	0,225	93
c = 3,4	2375	27	0,225	88
c = 3,5	2260	27	0,225	84
c = 3,6	2168	27	0,225	80
c = 3,7	2065	27	0,225	76
c = 3,8	1990	27	0,225	74
c = 3,9	1927	27	0,225	71
c = 4	1866	25	0,20833	75

Tabela 26 Desempenho do Método do Quartil (2MA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	47620	120	1	397
c = 0,1	44524	115	0,95833	387
c = 0,2	38121	111	0,925	343
c = 0,3	33742	105	0,875	321
c = 0,4	29829	102	0,85	292
c = 0,5	26234	100	0,83333	262
c = 0,6	23001	100	0,83333	230
c = 0,7	20185	94	0,78333	215
c = 0,8	17702	91	0,75833	195
c = 0,9	15471	87	0,725	178
c = 1	13602	84	0,7	162
c = 1,1	11881	78	0,65	152
c = 1,2	10387	75	0,625	138
c = 1,3	9147	72	0,6	127
c = 1,4	7942	69	0,575	115
c = 1,5	6924	69	0,575	100
c = 1,6	6072	67	0,55833	91
c = 1,7	5328	63	0,525	85
c = 1,8	4652	60	0,5	78
c = 1,9	4115	58	0,48333	71
c = 2	3632	55	0,45833	66
c = 2,1	3233	53	0,44167	61
c = 2,2	2885	50	0,41667	58
c = 2,3	2578	47	0,39167	55
c = 2,4	2320	44	0,36667	53
c = 2,5	2082	43	0,35833	48
c = 2,6	1884	40	0,33333	47
c = 2,7	1674	38	0,31667	44
c = 2,8	1529	34	0,28333	45
c = 2,9	1414	34	0,28333	42
c = 3	1307	34	0,28333	38
c = 3,1	1202	33	0,275	36
c = 3,2	1105	32	0,26667	35
c = 3,3	1005	30	0,25	34
c = 3,4	915	26	0,21667	35
c = 3,5	842	24	0,2	35
c = 3,6	801	24	0,2	33
c = 3,7	737	24	0,2	31
c = 3,8	684	24	0,2	29
c = 3,9	635	23	0,19167	28
c = 4	577	22	0,18333	26

Tabela 27 Desempenho do Método do Quartil Modificado (2MA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	38293	111	0,925	345
c = 0,1	23445	100	0,83333	234
c = 0,2	14177	85	0,70833	167
c = 0,3	14174	85	0,70833	167
c = 0,4	14162	85	0,70833	167
c = 0,5	14129	85	0,70833	166
c = 0,6	14072	85	0,70833	166
c = 0,7	13968	85	0,70833	164
c = 0,8	13864	85	0,70833	163
c = 0,9	13732	84	0,7	163
c = 1	13602	84	0,7	162
c = 1,1	13432	82	0,68333	164
c = 1,2	13272	82	0,68333	162
c = 1,3	13090	81	0,675	162
c = 1,4	12854	78	0,65	165
c = 1,5	12613	77	0,64167	164
c = 1,6	12348	76	0,63333	162
c = 1,7	12072	75	0,625	161
c = 1,8	11800	74	0,61667	159
c = 1,9	11493	73	0,60833	157
c = 2	11189	71	0,59167	158
c = 2,1	10887	69	0,575	158
c = 2,2	10570	68	0,56667	155
c = 2,3	10283	68	0,56667	151
c = 2,4	9997	66	0,55	151
c = 2,5	9715	65	0,54167	149
c = 2,6	9429	63	0,525	150
c = 2,7	9104	62	0,51667	147
c = 2,8	8815	61	0,50833	145
c = 2,9	8533	60	0,5	142
c = 3	8270	59	0,49167	140
c = 3,1	7962	58	0,48333	137
c = 3,2	7697	58	0,48333	133
c = 3,3	7452	57	0,475	131
c = 3,4	7197	56	0,46667	129
c = 3,5	6952	56	0,46667	124
c = 3,6	6725	54	0,45	125
c = 3,7	6501	53	0,44167	123
c = 3,8	6316	53	0,44167	119
c = 3,9	6110	50	0,41667	122
c = 4	5896	50	0,41667	118

Tabela 28 Desempenho do Método do Desvio Mediano Absoluto (2MA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	47620	120	1	397
c = 0,1	45837	118	0,98333	388
c = 0,2	44007	114	0,95	386
c = 0,3	42244	113	0,94167	374
c = 0,4	40485	113	0,94167	358
c = 0,5	38618	112	0,93333	345
c = 0,6	36990	110	0,91667	336
c = 0,7	35339	106	0,88333	333
c = 0,8	33805	104	0,86667	325
c = 0,9	32211	104	0,86667	310
c = 1	30370	104	0,86667	292
c = 1,1	28855	103	0,85833	280
c = 1,2	27485	102	0,85	269
c = 1,3	26177	99	0,825	264
c = 1,4	24919	95	0,79167	262
c = 1,5	23800	94	0,78333	253
c = 1,6	22672	91	0,75833	249
c = 1,7	21612	87	0,725	248
c = 1,8	20596	85	0,70833	242
c = 1,9	19671	81	0,675	243
c = 2	18752	80	0,66667	234
c = 2,1	17890	80	0,66667	224
c = 2,2	17126	80	0,66667	214
c = 2,3	16358	75	0,625	218
c = 2,4	15561	74	0,61667	210
c = 2,5	14828	73	0,60833	203
c = 2,6	14193	71	0,59167	200
c = 2,7	13534	70	0,58333	193
c = 2,8	12895	68	0,56667	190
c = 2,9	12310	66	0,55	187
c = 3	11777	64	0,53333	184
c = 3,1	11296	61	0,50833	185
c = 3,2	10799	61	0,50833	177
c = 3,3	10359	60	0,5	173
c = 3,4	9904	59	0,49167	168
c = 3,5	9452	58	0,48333	163
c = 3,6	9097	56	0,46667	162
c = 3,7	8710	54	0,45	161
c = 3,8	8323	53	0,44167	157
c = 3,9	8001	53	0,44167	151
c = 4	7691	53	0,44167	145

Tabela 29 Desempenho do Algoritmo de Tukey (2MA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	49733	120	1	414
c = 0,1	47182	116	0,96667	407
c = 0,2	44512	114	0,95	390
c = 0,3	41945	111	0,925	378
c = 0,4	39494	110	0,91667	359
c = 0,5	37114	107	0,89167	347
c = 0,6	34763	106	0,88333	328
c = 0,7	32576	104	0,86667	313
c = 0,8	30462	104	0,86667	293
c = 0,9	28537	103	0,85833	277
c = 1	26527	97	0,80833	273
c = 1,1	24793	95	0,79167	261
c = 1,2	23179	91	0,75833	255
c = 1,3	21614	91	0,75833	238
c = 1,4	20178	86	0,71667	235
c = 1,5	18845	83	0,69167	227
c = 1,6	17582	81	0,675	217
c = 1,7	16428	78	0,65	211
c = 1,8	15293	77	0,64167	199
c = 1,9	14297	73	0,60833	196
c = 2	13368	70	0,58333	191
c = 2,1	12501	66	0,55	189
c = 2,2	11710	65	0,54167	180
c = 2,3	10912	63	0,525	173
c = 2,4	10243	60	0,5	171
c = 2,5	9611	57	0,475	169
c = 2,6	8962	53	0,44167	169
c = 2,7	8383	49	0,40833	171
c = 2,8	7842	47	0,39167	167
c = 2,9	7350	46	0,38333	160
c = 3	6829	45	0,375	152
c = 3,1	6392	44	0,36667	145
c = 3,2	5989	42	0,35	143
c = 3,3	5591	40	0,33333	140
c = 3,4	5231	38	0,31667	138
c = 3,5	4959	37	0,30833	134
c = 3,6	4627	37	0,30833	125
c = 3,7	4380	36	0,3	122
c = 3,8	4106	34	0,28333	121
c = 3,9	3860	31	0,25833	125
c = 4	3672	31	0,25833	118

Tabela 30 Desempenho do Método de Cercas Resistentes (MMAA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	27149	94	0,78333	289
c = 0,1	23858	93	0,775	257
c = 0,2	20889	88	0,73333	237
c = 0,3	18476	81	0,675	228
c = 0,4	16265	76	0,63333	214
c = 0,5	14397	72	0,6	200
c = 0,6	12712	62	0,51667	205
c = 0,7	11208	56	0,46667	200
c = 0,8	9975	54	0,45	185
c = 0,9	8897	48	0,4	185
c = 1	7927	46	0,38333	172
c = 1,1	7146	42	0,35	170
c = 1,2	6464	41	0,34167	158
c = 1,3	5851	38	0,31667	154
c = 1,4	5317	37	0,30833	144
c = 1,5	4868	36	0,3	135
c = 1,6	4448	36	0,3	124
c = 1,7	4054	34	0,28333	119
c = 1,8	3720	32	0,26667	116
c = 1,9	3449	30	0,25	115
c = 2	3181	29	0,24167	110
c = 2,1	2960	28	0,23333	106
c = 2,2	2744	27	0,225	102
c = 2,3	2570	25	0,20833	103
c = 2,4	2392	24	0,2	100
c = 2,5	2239	22	0,18333	102
c = 2,6	2086	21	0,175	99
c = 2,7	1955	19	0,15833	103
c = 2,8	1848	19	0,15833	97
c = 2,9	1753	19	0,15833	92
c = 3	1665	18	0,15	93
c = 3,1	1580	15	0,125	105
c = 3,2	1497	15	0,125	100
c = 3,3	1433	15	0,125	96
c = 3,4	1365	15	0,125	91
c = 3,5	1294	15	0,125	86
c = 3,6	1239	15	0,125	83
c = 3,7	1185	15	0,125	79
c = 3,8	1150	14	0,11667	82
c = 3,9	1107	14	0,11667	79
c = 4	1069	14	0,11667	76



Tabela 31 Desempenho do Método do Quartil (MMAA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	47271	119	0,99167	397
c = 0,1	45048	117	0,975	385
c = 0,2	42799	116	0,96667	369
c = 0,3	40550	110	0,91667	369
c = 0,4	38444	108	0,9	356
c = 0,5	36356	107	0,89167	340
c = 0,6	34445	106	0,88333	325
c = 0,7	32452	103	0,85833	315
c = 0,8	30649	102	0,85	300
c = 0,9	28829	101	0,84167	285
c = 1	26371	94	0,78333	281
c = 1,1	24681	93	0,775	265
c = 1,2	23072	93	0,775	248
c = 1,3	21586	91	0,75833	237
c = 1,4	20207	90	0,75	225
c = 1,5	18954	82	0,68333	231
c = 1,6	17732	78	0,65	227
c = 1,7	16603	74	0,61667	224
c = 1,8	15512	73	0,60833	212
c = 1,9	14515	70	0,58333	207
c = 2	13554	68	0,56667	199
c = 2,1	12738	60	0,5	212
c = 2,2	11950	60	0,5	199
c = 2,3	11221	56	0,46667	200
c = 2,4	10556	54	0,45	195
c = 2,5	9899	51	0,425	194
c = 2,6	9296	51	0,425	182
c = 2,7	8728	48	0,4	182
c = 2,8	8225	46	0,38333	179
c = 2,9	7752	45	0,375	172
c = 3	7294	44	0,36667	166
c = 3,1	6894	43	0,35833	160
c = 3,2	6533	42	0,35	156
c = 3,3	6183	42	0,35	147
c = 3,4	5811	40	0,33333	145
c = 3,5	5486	40	0,33333	137
c = 3,6	5194	39	0,325	133
c = 3,7	4871	31	0,25833	157
c = 3,8	4599	31	0,25833	148
c = 3,9	4352	30	0,25	145
c = 4	4096	28	0,23333	146

Tabela 32 Desempenho do Método do Quartil Modificado (MMAA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	36835	109	0,90833	338
c = 0,1	36828	109	0,90833	338
c = 0,2	36715	109	0,90833	337
c = 0,3	36279	108	0,9	336
c = 0,4	35547	106	0,88333	335
c = 0,5	34486	106	0,88333	325
c = 0,6	33176	105	0,875	316
c = 0,7	31673	103	0,85833	308
c = 0,8	30164	102	0,85	296
c = 0,9	28610	101	0,84167	283
c = 1	26371	94	0,78333	281
c = 1,1	24802	93	0,775	267
c = 1,2	23298	93	0,775	251
c = 1,3	21905	91	0,75833	241
c = 1,4	20586	90	0,75	229
c = 1,5	19422	82	0,68333	237
c = 1,6	18277	78	0,65	234
c = 1,7	17243	74	0,61667	233
c = 1,8	16205	73	0,60833	222
c = 1,9	15257	70	0,58333	218
c = 2	14320	68	0,56667	211
c = 2,1	13527	60	0,5	225
c = 2,2	12779	60	0,5	213
c = 2,3	12124	56	0,46667	217
c = 2,4	11489	54	0,45	213
c = 2,5	10878	51	0,425	213
c = 2,6	10314	51	0,425	202
c = 2,7	9796	48	0,4	204
c = 2,8	9310	46	0,38333	202
c = 2,9	8863	45	0,375	197
c = 3	8424	44	0,36667	191
c = 3,1	8034	43	0,35833	187
c = 3,2	7682	42	0,35	183
c = 3,3	7356	42	0,35	175
c = 3,4	7017	40	0,33333	175
c = 3,5	6701	40	0,33333	168
c = 3,6	6425	39	0,325	165
c = 3,7	6126	32	0,26667	191
c = 3,8	5854	32	0,26667	183
c = 3,9	5620	32	0,26667	176
c = 4	5389	30	0,25	180

Tabela 33 Desempenho do Método do Desvio Mediano Absoltuo (MMAA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	47271	119	0,99167	397
c = 0,1	45106	118	0,98333	382
c = 0,2	42894	115	0,95833	373
c = 0,3	40758	112	0,93333	364
c = 0,4	38638	109	0,90833	354
c = 0,5	36728	106	0,88333	346
c = 0,6	34784	106	0,88333	328
c = 0,7	32827	103	0,85833	319
c = 0,8	31031	101	0,84167	307
c = 0,9	29220	100	0,83333	292
c = 1	26896	97	0,80833	277
c = 1,1	25191	96	0,8	262
c = 1,2	23768	93	0,775	256
c = 1,3	22325	90	0,75	248
c = 1,4	20981	87	0,725	241
c = 1,5	19744	86	0,71667	230
c = 1,6	18582	83	0,69167	224
c = 1,7	17441	80	0,66667	218
c = 1,8	16421	77	0,64167	213
c = 1,9	15486	76	0,63333	204
c = 2	14556	73	0,60833	199
c = 2,1	13681	69	0,575	198
c = 2,2	12823	68	0,56667	189
c = 2,3	12089	66	0,55	183
c = 2,4	11375	62	0,51667	183
c = 2,5	10755	61	0,50833	176
c = 2,6	10193	58	0,48333	176
c = 2,7	9602	55	0,45833	175
c = 2,8	9060	53	0,44167	171
c = 2,9	8623	52	0,43333	166
c = 3	8178	51	0,425	160
c = 3,1	7768	48	0,4	162
c = 3,2	7396	48	0,4	154
c = 3,3	7045	47	0,39167	150
c = 3,4	6685	42	0,35	159
c = 3,5	6357	40	0,33333	159
c = 3,6	6051	40	0,33333	151
c = 3,7	5779	39	0,325	148
c = 3,8	5544	38	0,31667	146
c = 3,9	5297	37	0,30833	143
c = 4	5094	37	0,30833	138

Tabela 34 Desempenho do Algoritmo de Tukey (MMAA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	49911	120	1	416
c = 0,1	46943	119	0,99167	394
c = 0,2	44040	116	0,96667	380
c = 0,3	41019	113	0,94167	363
c = 0,4	38274	107	0,89167	358
c = 0,5	35425	104	0,86667	341
c = 0,6	32762	102	0,85	321
c = 0,7	30244	100	0,83333	302
c = 0,8	27757	95	0,79167	292
c = 0,9	25505	95	0,79167	268
c = 1	23248	91	0,75833	255
c = 1,1	21295	90	0,75	237
c = 1,2	19526	86	0,71667	227
c = 1,3	17855	83	0,69167	215
c = 1,4	16304	81	0,675	201
c = 1,5	14917	75	0,625	199
c = 1,6	13622	70	0,58333	195
c = 1,7	12425	66	0,55	188
c = 1,8	11406	62	0,51667	184
c = 1,9	10416	57	0,475	183
c = 2	9493	55	0,45833	173
c = 2,1	8683	51	0,425	170
c = 2,2	7967	48	0,4	166
c = 2,3	7332	44	0,36667	167
c = 2,4	6725	41	0,34167	164
c = 2,5	6186	38	0,31667	163
c = 2,6	5718	37	0,30833	155
c = 2,7	5262	34	0,28333	155
c = 2,8	4874	34	0,28333	143
c = 2,9	4554	32	0,26667	142
c = 3	4218	30	0,25	141
c = 3,1	3933	28	0,23333	140
c = 3,2	3626	27	0,225	134
c = 3,3	3386	27	0,225	125
c = 3,4	3162	26	0,21667	122
c = 3,5	2927	24	0,2	122
c = 3,6	2712	23	0,19167	118
c = 3,7	2565	22	0,18333	117
c = 3,8	2392	21	0,175	114
c = 3,9	2231	20	0,16667	112
c = 4	2100	19	0,15833	111

Tabela 35 Desempenho do Método de Cercas Resistentes (TA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	24910	104	0,86667	240
c = 0,1	21660	100	0,83333	217
c = 0,2	18692	96	0,8	195
c = 0,3	16176	91	0,75833	178
c = 0,4	13994	84	0,7	167
c = 0,5	12000	79	0,65833	152
c = 0,6	10336	75	0,625	138
c = 0,7	9004	71	0,59167	127
c = 0,8	7774	66	0,55	118
c = 0,9	6808	64	0,53333	106
c = 1	5961	58	0,48333	103
c = 1,1	5215	56	0,46667	93
c = 1,2	4626	56	0,46667	83
c = 1,3	4139	54	0,45	77
c = 1,4	3682	52	0,43333	71
c = 1,5	3305	48	0,4	69
c = 1,6	3018	45	0,375	67
c = 1,7	2784	43	0,35833	65
c = 1,8	2548	40	0,33333	64
c = 1,9	2347	40	0,33333	59
c = 2	2174	38	0,31667	57
c = 2,1	2011	38	0,31667	53
c = 2,2	1860	36	0,3	52
c = 2,3	1722	34	0,28333	51
c = 2,4	1612	32	0,26667	50
c = 2,5	1518	31	0,25833	49
c = 2,6	1421	29	0,24167	49
c = 2,7	1335	28	0,23333	48
c = 2,8	1266	27	0,225	47
c = 2,9	1189	26	0,21667	46
c = 3	1119	24	0,2	47
c = 3,1	1064	24	0,2	44
c = 3,2	1014	24	0,2	42
c = 3,3	967	23	0,19167	42
c = 3,4	909	21	0,175	43
c = 3,5	870	21	0,175	41
c = 3,6	831	21	0,175	40
c = 3,7	784	20	0,16667	39
c = 3,8	758	20	0,16667	38
c = 3,9	729	20	0,16667	36
c = 4	706	20	0,16667	35

Tabela 36 Desempenho do Método do Quartil (TA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	45816	119	0,99167	385
c = 0,1	43432	116	0,96667	374
c = 0,2	40979	115	0,95833	356
c = 0,3	38658	112	0,93333	345
c = 0,4	36370	112	0,93333	325
c = 0,5	34202	111	0,925	308
c = 0,6	32087	111	0,925	289
c = 0,7	30073	109	0,90833	276
c = 0,8	28108	108	0,9	260
c = 0,9	26302	108	0,9	244
c = 1	24477	104	0,86667	235
c = 1,1	22868	102	0,85	224
c = 1,2	21239	99	0,825	215
c = 1,3	19708	97	0,80833	203
c = 1,4	18300	96	0,8	191
c = 1,5	16949	94	0,78333	180
c = 1,6	15773	87	0,725	181
c = 1,7	14674	86	0,71667	171
c = 1,8	13629	83	0,69167	164
c = 1,9	12599	81	0,675	156
c = 2	11629	76	0,63333	153
c = 2,1	10693	74	0,61667	145
c = 2,2	9934	73	0,60833	136
c = 2,3	9156	73	0,60833	125
c = 2,4	8440	72	0,6	117
c = 2,5	7833	70	0,58333	112
c = 2,6	7262	69	0,575	105
c = 2,7	6728	67	0,55833	100
c = 2,8	6251	64	0,53333	98
c = 2,9	5802	62	0,51667	94
c = 3	5368	61	0,50833	88
c = 3,1	5017	59	0,49167	85
c = 3,2	4668	58	0,48333	80
c = 3,3	4373	57	0,475	77
c = 3,4	4019	57	0,475	71
c = 3,5	3720	57	0,475	65
c = 3,6	3458	53	0,44167	65
c = 3,7	3248	50	0,41667	65
c = 3,8	3072	49	0,40833	63
c = 3,9	2871	45	0,375	64
c = 4	2654	43	0,35833	62

Tabela 37 Desempenho do Método do Quartil Modificado (TA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	36375	112	0,93333	325
c = 0,1	36375	112	0,93333	325
c = 0,2	36331	112	0,93333	324
c = 0,3	35896	112	0,93333	321
c = 0,4	34785	112	0,93333	311
c = 0,5	33264	111	0,925	300
c = 0,6	31490	111	0,925	284
c = 0,7	29714	109	0,90833	273
c = 0,8	27888	108	0,9	258
c = 0,9	26193	108	0,9	243
c = 1	24477	104	0,86667	235
c = 1,1	22954	102	0,85	225
c = 1,2	21391	100	0,83333	214
c = 1,3	19910	98	0,81667	203
c = 1,4	18560	97	0,80833	191
c = 1,5	17263	95	0,79167	182
c = 1,6	16123	88	0,73333	183
c = 1,7	15084	87	0,725	173
c = 1,8	14075	85	0,70833	166
c = 1,9	13086	83	0,69167	158
c = 2	12156	78	0,65	156
c = 2,1	11262	76	0,63333	148
c = 2,2	10526	75	0,625	140
c = 2,3	9806	75	0,625	131
c = 2,4	9115	74	0,61667	123
c = 2,5	8521	72	0,6	118
c = 2,6	7965	71	0,59167	112
c = 2,7	7437	69	0,575	108
c = 2,8	6965	66	0,55	106
c = 2,9	6528	64	0,53333	102
c = 3	6106	63	0,525	97
c = 3,1	5761	61	0,50833	94
c = 3,2	5435	60	0,5	91
c = 3,3	5133	59	0,49167	87
c = 3,4	4816	59	0,49167	82
c = 3,5	4531	58	0,48333	78
c = 3,6	4272	54	0,45	79
c = 3,7	4057	51	0,425	80
c = 3,8	3887	50	0,41667	78
c = 3,9	3677	46	0,38333	80
c = 4	3485	44	0,36667	79

Tabela 38 Desempenho do Método do Desvio Mediano Absoluto (TA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	45816	119	0,99167	385
c = 0,1	43481	117	0,975	372
c = 0,2	41091	114	0,95	360
c = 0,3	38871	112	0,93333	347
c = 0,4	36657	111	0,925	330
c = 0,5	34402	110	0,91667	313
c = 0,6	32353	109	0,90833	297
c = 0,7	30401	106	0,88333	287
c = 0,8	28452	104	0,86667	274
c = 0,9	26581	102	0,85	261
c = 1	24836	101	0,84167	246
c = 1,1	23208	100	0,83333	232
c = 1,2	21615	97	0,80833	223
c = 1,3	20135	95	0,79167	212
c = 1,4	18738	92	0,76667	204
c = 1,5	17439	90	0,75	194
c = 1,6	16212	87	0,725	186
c = 1,7	15076	86	0,71667	175
c = 1,8	14037	83	0,69167	169
c = 1,9	13055	81	0,675	161
c = 2	12072	80	0,66667	151
c = 2,1	11220	77	0,64167	146
c = 2,2	10507	75	0,625	140
c = 2,3	9785	74	0,61667	132
c = 2,4	9133	71	0,59167	129
c = 2,5	8516	70	0,58333	122
c = 2,6	7953	69	0,575	115
c = 2,7	7407	67	0,55833	111
c = 2,8	6963	64	0,53333	109
c = 2,9	6549	63	0,525	104
c = 3	6122	61	0,50833	100
c = 3,1	5726	60	0,5	95
c = 3,2	5384	59	0,49167	91
c = 3,3	5054	57	0,475	89
c = 3,4	4749	54	0,45	88
c = 3,5	4487	54	0,45	83
c = 3,6	4244	54	0,45	79
c = 3,7	4005	52	0,43333	77
c = 3,8	3795	52	0,43333	73
c = 3,9	3609	52	0,43333	69
c = 4	3439	47	0,39167	73



Tabela 39 Desempenho do Algoritmo de Tukey (TA) no Município do Rio de Janeiro

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	50072	120	1	417
c = 0,1	46830	120	1	390
c = 0,2	41550	118	0,98333	352
c = 0,3	37558	116	0,96667	324
c = 0,4	34641	113	0,94167	307
c = 0,5	31885	112	0,93333	285
c = 0,6	29259	107	0,89167	273
c = 0,7	26808	106	0,88333	253
c = 0,8	24490	103	0,85833	238
c = 0,9	22295	99	0,825	225
c = 1	20280	95	0,79167	213
c = 1,1	18372	94	0,78333	195
c = 1,2	16553	91	0,75833	182
c = 1,3	14941	88	0,73333	170
c = 1,4	13559	86	0,71667	158
c = 1,5	12195	83	0,69167	147
c = 1,6	10993	82	0,68333	134
c = 1,7	9798	76	0,63333	129
c = 1,8	8697	74	0,61667	118
c = 1,9	7829	71	0,59167	110
c = 2	7018	69	0,575	102
c = 2,1	6298	67	0,55833	94
c = 2,2	5602	62	0,51667	90
c = 2,3	5039	60	0,5	84
c = 2,4	4490	54	0,45	83
c = 2,5	4025	53	0,44167	76
c = 2,6	3627	52	0,43333	70
c = 2,7	3259	52	0,43333	63
c = 2,8	2962	51	0,425	58
c = 2,9	2666	47	0,39167	57
c = 3	2438	47	0,39167	52
c = 3,1	2206	45	0,375	49
c = 3,2	2019	43	0,35833	47
c = 3,3	1830	37	0,30833	49
c = 3,4	1686	36	0,3	47
c = 3,5	1532	33	0,275	46
c = 3,6	1416	33	0,275	43
c = 3,7	1309	33	0,275	40
c = 3,8	1203	33	0,275	36
c = 3,9	1136	32	0,26667	36
c = 4	1058	29	0,24167	36

Tabela 40 Desempenho do Método de Cercas Resistentes (MR) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	27365	152	0,77551	180
c = 0,1	23353	131	0,66837	178
c = 0,2	20424	120	0,61224	170
c = 0,3	17972	114	0,58163	158
c = 0,4	16016	109	0,55612	147
c = 0,5	13543	97	0,4949	140
c = 0,6	11830	81	0,41327	146
c = 0,7	10570	73	0,37245	145
c = 0,8	9530	68	0,34694	140
c = 0,9	8626	62	0,31633	139
c = 1	7691	57	0,29082	135
c = 1,1	6969	56	0,28571	124
c = 1,2	6394	55	0,28061	116
c = 1,3	5864	52	0,26531	113
c = 1,4	5422	49	0,25	111
c = 1,5	5015	48	0,2449	104
c = 1,6	4662	48	0,2449	97
c = 1,7	4366	46	0,23469	95
c = 1,8	4059	44	0,22449	92
c = 1,9	3789	43	0,21939	88
c = 2	3552	42	0,21429	85
c = 2,1	3331	42	0,21429	79
c = 2,2	3136	41	0,20918	76
c = 2,3	2976	39	0,19898	76
c = 2,4	2819	37	0,18878	76
c = 2,5	2663	34	0,17347	78
c = 2,6	2549	33	0,16837	77
c = 2,7	2450	33	0,16837	74
c = 2,8	2354	31	0,15816	76
c = 2,9	2242	30	0,15306	75
c = 3	2136	29	0,14796	74
c = 3,1	2034	29	0,14796	70
c = 3,2	1963	29	0,14796	68
c = 3,3	1891	29	0,14796	65
c = 3,4	1821	29	0,14796	63
c = 3,5	1757	28	0,14286	63
c = 3,6	1696	26	0,13265	65
c = 3,7	1644	25	0,12755	66
c = 3,8	1593	23	0,11735	69
c = 3,9	1555	22	0,11224	71
c = 4	1510	22	0,11224	69

Tabela 41 Desempenho do Método do Quartil (MR) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	46858	193	0,98469	243
c = 0,1	44647	192	0,97959	233
c = 0,2	42596	185	0,94388	230
c = 0,3	40711	183	0,93367	222
c = 0,4	38914	179	0,91327	217
c = 0,5	37052	175	0,89286	212
c = 0,6	35235	169	0,86224	208
c = 0,7	33458	166	0,84694	202
c = 0,8	31668	164	0,83673	193
c = 0,9	29709	156	0,79592	190
c = 1	24929	145	0,7398	172
c = 1,1	22740	138	0,70408	165
c = 1,2	20881	126	0,64286	166
c = 1,3	19244	119	0,60714	162
c = 1,4	17798	117	0,59694	152
c = 1,5	16461	113	0,57653	146
c = 1,6	15319	108	0,55102	142
c = 1,7	14278	105	0,53571	136
c = 1,8	13320	100	0,5102	133
c = 1,9	12432	93	0,47449	134
c = 2	10873	83	0,42347	131
c = 2,1	9951	77	0,39286	129
c = 2,2	9314	75	0,38265	124
c = 2,3	8692	72	0,36735	121
c = 2,4	8122	69	0,35204	118
c = 2,5	7613	66	0,33673	115
c = 2,6	7160	65	0,33163	110
c = 2,7	6696	62	0,31633	108
c = 2,8	6328	59	0,30102	107
c = 2,9	5914	58	0,29592	102
c = 3	5566	55	0,28061	101
c = 3,1	5231	54	0,27551	97
c = 3,2	4960	54	0,27551	92
c = 3,3	4689	52	0,26531	90
c = 3,4	4380	52	0,26531	84
c = 3,5	4118	52	0,26531	79
c = 3,6	3906	52	0,26531	75
c = 3,7	3670	50	0,2551	73
c = 3,8	3477	49	0,25	71
c = 3,9	3290	48	0,2449	69
c = 4	3065	46	0,23469	67

Tabela 42 Desempenho do Método do Quartil Modificado (MR) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	33057	167	0,85204	198
c = 0,1	33057	167	0,85204	198
c = 0,2	33048	167	0,85204	198
c = 0,3	32993	167	0,85204	198
c = 0,4	32787	167	0,85204	196
c = 0,5	32410	166	0,84694	195
c = 0,6	31830	162	0,82653	196
c = 0,7	31105	159	0,81122	196
c = 0,8	30231	158	0,80612	191
c = 0,9	29049	155	0,79082	187
c = 1	24929	145	0,7398	172
c = 1,1	23100	138	0,70408	167
c = 1,2	21562	126	0,64286	171
c = 1,3	20105	120	0,61224	168
c = 1,4	18876	118	0,60204	160
c = 1,5	17771	114	0,58163	156
c = 1,6	16819	109	0,55612	154
c = 1,7	15940	106	0,54082	150
c = 1,8	15084	101	0,51531	149
c = 1,9	14361	95	0,48469	151
c = 2	12503	84	0,42857	149
c = 2,1	11526	78	0,39796	148
c = 2,2	10953	76	0,38776	144
c = 2,3	10456	73	0,37245	143
c = 2,4	9965	69	0,35204	144
c = 2,5	9515	66	0,33673	144
c = 2,6	9106	65	0,33163	140
c = 2,7	8716	62	0,31633	141
c = 2,8	8353	59	0,30102	142
c = 2,9	8008	58	0,29592	138
c = 3	7708	56	0,28571	138
c = 3,1	7404	55	0,28061	135
c = 3,2	7150	55	0,28061	130
c = 3,3	6879	53	0,27041	130
c = 3,4	6638	53	0,27041	125
c = 3,5	6395	53	0,27041	121
c = 3,6	6185	53	0,27041	117
c = 3,7	5946	51	0,2602	117
c = 3,8	5761	50	0,2551	115
c = 3,9	5588	49	0,25	114
c = 4	5319	47	0,2398	113

Tabela 43 Desempenho do Método do Desvio Mediano Absoluto (MR) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	46858	193	0,98469	243
c = 0,1	44877	192	0,97959	234
c = 0,2	42930	186	0,94898	231
c = 0,3	41123	181	0,92347	227
c = 0,4	39355	177	0,90306	222
c = 0,5	37714	174	0,88776	217
c = 0,6	36131	169	0,86224	214
c = 0,7	34616	164	0,83673	211
c = 0,8	33010	160	0,81633	206
c = 0,9	31465	153	0,78061	206
c = 1	25116	134	0,68367	187
c = 1,1	23222	125	0,63776	186
c = 1,2	21731	122	0,62245	178
c = 1,3	20397	117	0,59694	174
c = 1,4	19203	111	0,56633	173
c = 1,5	18138	107	0,54592	170
c = 1,6	17119	103	0,52551	166
c = 1,7	16163	99	0,5051	163
c = 1,8	15255	93	0,47449	164
c = 1,9	14420	91	0,46429	158
c = 2	13530	88	0,44898	154
c = 2,1	12848	85	0,43367	151
c = 2,2	12201	81	0,41327	151
c = 2,3	11577	78	0,39796	148
c = 2,4	11044	75	0,38265	147
c = 2,5	10519	74	0,37755	142
c = 2,6	10052	73	0,37245	138
c = 2,7	9607	69	0,35204	139
c = 2,8	9202	69	0,35204	133
c = 2,9	8795	69	0,35204	127
c = 3	8446	67	0,34184	126
c = 3,1	8123	65	0,33163	125
c = 3,2	7841	64	0,32653	123
c = 3,3	7521	63	0,32143	119
c = 3,4	7241	63	0,32143	115
c = 3,5	6975	61	0,31122	114
c = 3,6	6756	59	0,30102	115
c = 3,7	6567	58	0,29592	113
c = 3,8	6378	53	0,27041	120
c = 3,9	6173	53	0,27041	116
c = 4	5990	52	0,26531	115

Tabela 44 Desempenho do Algoritmo de Tukey (MR) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	54339	192	0,97959	283
c = 0,1	50585	187	0,95408	271
c = 0,2	47004	181	0,92347	260
c = 0,3	43426	173	0,88265	251
c = 0,4	39966	168	0,85714	238
c = 0,5	36621	163	0,83163	225
c = 0,6	33311	154	0,78571	216
c = 0,7	30215	147	0,75	206
c = 0,8	27211	144	0,73469	189
c = 0,9	24455	139	0,70918	176
c = 1	19653	108	0,55102	182
c = 1,1	17355	101	0,51531	172
c = 1,2	15243	92	0,46939	166
c = 1,3	13388	86	0,43878	156
c = 1,4	11672	80	0,40816	146
c = 1,5	10102	76	0,38776	133
c = 1,6	8781	74	0,37755	119
c = 1,7	7681	67	0,34184	115
c = 1,8	6618	60	0,30612	110
c = 1,9	5717	56	0,28571	102
c = 2	4910	51	0,2602	96
c = 2,1	4229	49	0,25	86
c = 2,2	3648	46	0,23469	79
c = 2,3	3203	42	0,21429	76
c = 2,4	2768	37	0,18878	75
c = 2,5	2388	37	0,18878	65
c = 2,6	2104	36	0,18367	58
c = 2,7	1831	32	0,16327	57
c = 2,8	1604	31	0,15816	52
c = 2,9	1426	30	0,15306	48
c = 3	1273	28	0,14286	45
c = 3,1	1139	28	0,14286	41
c = 3,2	1009	26	0,13265	39
c = 3,3	892	26	0,13265	34
c = 3,4	808	23	0,11735	35
c = 3,5	716	21	0,10714	34
c = 3,6	648	17	0,08673	38
c = 3,7	586	15	0,07653	39
c = 3,8	544	15	0,07653	36
c = 3,9	502	13	0,06633	39
c = 4	456	13	0,06633	35

Tabela 45 Desempenho do Método de Cercas Resistentes (MA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	33208	167	0,85204	199
c = 0,1	30613	165	0,84184	186
c = 0,2	28249	157	0,80102	180
c = 0,3	26113	153	0,78061	171
c = 0,4	24138	149	0,7602	162
c = 0,5	22012	139	0,70918	158
c = 0,6	20352	138	0,70408	147
c = 0,7	18800	135	0,68878	139
c = 0,8	17378	129	0,65816	135
c = 0,9	16139	126	0,64286	128
c = 1	14976	123	0,62755	122
c = 1,1	13978	121	0,61735	116
c = 1,2	13017	120	0,61224	108
c = 1,3	12149	119	0,60714	102
c = 1,4	11351	114	0,58163	100
c = 1,5	10628	112	0,57143	95
c = 1,6	9969	108	0,55102	92
c = 1,7	9393	107	0,54592	88
c = 1,8	8825	104	0,53061	85
c = 1,9	8282	101	0,51531	82
c = 2	7793	100	0,5102	78
c = 2,1	7359	94	0,47959	78
c = 2,2	6978	89	0,45408	78
c = 2,3	6601	85	0,43367	78
c = 2,4	6285	79	0,40306	80
c = 2,5	5959	78	0,39796	76
c = 2,6	5684	77	0,39286	74
c = 2,7	5402	72	0,36735	75
c = 2,8	5169	71	0,36224	73
c = 2,9	4908	71	0,36224	69
c = 3	4689	70	0,35714	67
c = 3,1	4495	70	0,35714	64
c = 3,2	4304	67	0,34184	64
c = 3,3	4130	67	0,34184	62
c = 3,4	3985	67	0,34184	59
c = 3,5	3842	67	0,34184	57
c = 3,6	3724	66	0,33673	56
c = 3,7	3620	65	0,33163	56
c = 3,8	3504	64	0,32653	55
c = 3,9	3385	63	0,32143	54
c = 4	3284	62	0,31633	53

Tabela 46 Desempenho do Método do Quartil (MA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	49494	195	0,9949	254
c = 0,1	47703	190	0,96939	251
c = 0,2	45982	187	0,95408	246
c = 0,3	44144	185	0,94388	239
c = 0,4	42439	182	0,92857	233
c = 0,5	40625	179	0,91327	227
c = 0,6	38887	176	0,89796	221
c = 0,7	37204	173	0,88265	215
c = 0,8	35545	171	0,87245	208
c = 0,9	34065	168	0,85714	203
c = 1	31314	167	0,85204	188
c = 1,1	29840	166	0,84694	180
c = 1,2	28398	165	0,84184	172
c = 1,3	27103	161	0,82143	168
c = 1,4	25893	157	0,80102	165
c = 1,5	24671	152	0,77551	162
c = 1,6	23543	149	0,7602	158
c = 1,7	22437	149	0,7602	151
c = 1,8	21484	148	0,7551	145
c = 1,9	20452	147	0,75	139
c = 2	19394	144	0,73469	135
c = 2,1	18433	142	0,72449	130
c = 2,2	17571	138	0,70408	127
c = 2,3	16765	136	0,69388	123
c = 2,4	16009	132	0,67347	121
c = 2,5	15287	127	0,64796	120
c = 2,6	14598	127	0,64796	115
c = 2,7	13922	127	0,64796	110
c = 2,8	13355	124	0,63265	108
c = 2,9	12723	124	0,63265	103
c = 3	12175	121	0,61735	101
c = 3,1	11673	118	0,60204	99
c = 3,2	11216	117	0,59694	96
c = 3,3	10755	117	0,59694	92
c = 3,4	10185	115	0,58673	89
c = 3,5	9761	112	0,57143	87
c = 3,6	9360	110	0,56122	85
c = 3,7	8965	107	0,54592	84
c = 3,8	8580	105	0,53571	82
c = 3,9	8245	100	0,5102	82
c = 4	7871	98	0,5	80



Tabela 47 Desempenho do Método do Quartil Modificado (MA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	38424	184	0,93878	209
c = 0,1	38424	184	0,93878	209
c = 0,2	38403	184	0,93878	209
c = 0,3	38247	183	0,93367	209
c = 0,4	37849	180	0,91837	210
c = 0,5	37249	179	0,91327	208
c = 0,6	36420	176	0,89796	207
c = 0,7	35542	173	0,88265	205
c = 0,8	34548	171	0,87245	202
c = 0,9	33554	168	0,85714	200
c = 1	31314	167	0,85204	188
c = 1,1	30175	166	0,84694	182
c = 1,2	29051	166	0,84694	175
c = 1,3	27992	163	0,83163	172
c = 1,4	27043	158	0,80612	171
c = 1,5	26089	153	0,78061	171
c = 1,6	25175	150	0,76531	168
c = 1,7	24326	150	0,76531	162
c = 1,8	23534	149	0,7602	158
c = 1,9	22668	148	0,7551	153
c = 2	21641	144	0,73469	150
c = 2,1	20837	143	0,72959	146
c = 2,2	20088	140	0,71429	143
c = 2,3	19454	138	0,70408	141
c = 2,4	18806	136	0,69388	138
c = 2,5	18196	134	0,68367	136
c = 2,6	17609	133	0,67857	132
c = 2,7	17077	132	0,67347	129
c = 2,8	16556	128	0,65306	129
c = 2,9	16031	128	0,65306	125
c = 3	15539	125	0,63776	124
c = 3,1	15107	124	0,63265	122
c = 3,2	14702	123	0,62755	120
c = 3,3	14297	123	0,62755	116
c = 3,4	13898	120	0,61224	116
c = 3,5	13511	119	0,60714	114
c = 3,6	13162	117	0,59694	112
c = 3,7	12822	114	0,58163	112
c = 3,8	12460	111	0,56633	112
c = 3,9	12161	108	0,55102	113
c = 4	11845	104	0,53061	114

Tabela 48 Desempenho do Método do Desvio Mediano Absoluto (MA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	49494	195	0,9949	254
c = 0,1	47862	189	0,96429	253
c = 0,2	46284	186	0,94898	249
c = 0,3	44691	185	0,94388	242
c = 0,4	43056	182	0,92857	237
c = 0,5	41522	178	0,90816	233
c = 0,6	39931	176	0,89796	227
c = 0,7	38462	172	0,87755	224
c = 0,8	36950	171	0,87245	216
c = 0,9	35476	169	0,86224	210
c = 1	32349	162	0,82653	200
c = 1,1	31052	161	0,82143	193
c = 1,2	29755	159	0,81122	187
c = 1,3	28622	155	0,79082	185
c = 1,4	27480	155	0,79082	177
c = 1,5	26352	151	0,77041	175
c = 1,6	25334	150	0,76531	169
c = 1,7	24303	146	0,7449	166
c = 1,8	23396	145	0,7398	161
c = 1,9	22458	144	0,73469	156
c = 2	21596	141	0,71939	153
c = 2,1	20788	140	0,71429	148
c = 2,2	19951	138	0,70408	145
c = 2,3	19179	138	0,70408	139
c = 2,4	18416	134	0,68367	137
c = 2,5	17706	133	0,67857	133
c = 2,6	17056	128	0,65306	133
c = 2,7	16458	127	0,64796	130
c = 2,8	15884	127	0,64796	125
c = 2,9	15328	126	0,64286	122
c = 3	14794	123	0,62755	120
c = 3,1	14289	120	0,61224	119
c = 3,2	13837	119	0,60714	116
c = 3,3	13376	115	0,58673	116
c = 3,4	12908	114	0,58163	113
c = 3,5	12494	113	0,57653	111
c = 3,6	12106	111	0,56633	109
c = 3,7	11758	109	0,55612	108
c = 3,8	11400	107	0,54592	107
c = 3,9	11062	105	0,53571	105
c = 4	10723	103	0,52551	104

Tabela 49 Desempenho do Algoritmo de Tukey (MA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	54357	190	0,96939	286
c = 0,1	51207	188	0,95918	272
c = 0,2	48045	184	0,93878	261
c = 0,3	45099	183	0,93367	246
c = 0,4	42042	178	0,90816	236
c = 0,5	39242	174	0,88776	226
c = 0,6	36425	170	0,86735	214
c = 0,7	33829	166	0,84694	204
c = 0,8	31313	161	0,82143	194
c = 0,9	29030	157	0,80102	185
c = 1	26776	146	0,7449	183
c = 1,1	24841	140	0,71429	177
c = 1,2	22992	135	0,68878	170
c = 1,3	21326	133	0,67857	160
c = 1,4	19772	128	0,65306	154
c = 1,5	18282	125	0,63776	146
c = 1,6	16976	123	0,62755	138
c = 1,7	15777	119	0,60714	133
c = 1,8	14645	113	0,57653	130
c = 1,9	13649	109	0,55612	125
c = 2	12628	104	0,53061	121
c = 2,1	11734	99	0,5051	119
c = 2,2	10915	98	0,5	111
c = 2,3	10236	98	0,5	104
c = 2,4	9574	95	0,48469	101
c = 2,5	8946	94	0,47959	95
c = 2,6	8396	91	0,46429	92
c = 2,7	7874	90	0,45918	87
c = 2,8	7374	85	0,43367	87
c = 2,9	6931	83	0,42347	84
c = 3	6518	81	0,41327	80
c = 3,1	6140	80	0,40816	77
c = 3,2	5752	77	0,39286	75
c = 3,3	5393	73	0,37245	74
c = 3,4	5094	72	0,36735	71
c = 3,5	4804	68	0,34694	71
c = 3,6	4512	66	0,33673	68
c = 3,7	4260	62	0,31633	69
c = 3,8	4029	60	0,30612	67
c = 3,9	3818	59	0,30102	65
c = 4	3613	58	0,29592	62

Tabela 50 Desempenho do Método de Cercas Resistentes (2MA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	31907	168	0,85714	190
c = 0,1	29220	159	0,81122	184
c = 0,2	26845	155	0,79082	173
c = 0,3	24637	149	0,7602	165
c = 0,4	22642	142	0,72449	159
c = 0,5	20496	136	0,69388	151
c = 0,6	18828	132	0,67347	143
c = 0,7	17342	128	0,65306	135
c = 0,8	15946	125	0,63776	128
c = 0,9	14724	119	0,60714	124
c = 1	13614	111	0,56633	123
c = 1,1	12575	107	0,54592	118
c = 1,2	11690	101	0,51531	116
c = 1,3	10893	95	0,48469	115
c = 1,4	10166	93	0,47449	109
c = 1,5	9518	91	0,46429	105
c = 1,6	8902	91	0,46429	98
c = 1,7	8368	87	0,44388	96
c = 1,8	7873	81	0,41327	97
c = 1,9	7452	80	0,40816	93
c = 2	7052	79	0,40306	89
c = 2,1	6662	77	0,39286	87
c = 2,2	6326	73	0,37245	87
c = 2,3	5992	72	0,36735	83
c = 2,4	5709	70	0,35714	82
c = 2,5	5437	66	0,33673	82
c = 2,6	5184	64	0,32653	81
c = 2,7	4981	63	0,32143	79
c = 2,8	4750	62	0,31633	77
c = 2,9	4549	60	0,30612	76
c = 3	4370	60	0,30612	73
c = 3,1	4209	59	0,30102	71
c = 3,2	4030	58	0,29592	69
c = 3,3	3890	57	0,29082	68
c = 3,4	3749	56	0,28571	67
c = 3,5	3624	53	0,27041	68
c = 3,6	3490	53	0,27041	66
c = 3,7	3399	52	0,26531	65
c = 3,8	3301	49	0,25	67
c = 3,9	3198	46	0,23469	70
c = 4	3109	46	0,23469	68

Tabela 51 Desempenho do Método do Quartil (2MA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	49352	196	1	252
c = 0,1	47467	194	0,9898	245
c = 0,2	45583	189	0,96429	241
c = 0,3	43526	187	0,95408	233
c = 0,4	41666	184	0,93878	226
c = 0,5	39752	180	0,91837	221
c = 0,6	37828	180	0,91837	210
c = 0,7	36060	178	0,90816	203
c = 0,8	34474	174	0,88776	198
c = 0,9	32738	170	0,86735	193
c = 1	30161	166	0,84694	182
c = 1,1	28692	158	0,80612	182
c = 1,2	27236	156	0,79592	175
c = 1,3	25971	154	0,78571	169
c = 1,4	24698	150	0,76531	165
c = 1,5	23529	144	0,73469	163
c = 1,6	22384	143	0,72959	157
c = 1,7	21254	141	0,71939	151
c = 1,8	20250	140	0,71429	145
c = 1,9	19221	138	0,70408	139
c = 2	18130	133	0,67857	136
c = 2,1	17232	128	0,65306	135
c = 2,2	16399	126	0,64286	130
c = 2,3	15583	122	0,62245	128
c = 2,4	14863	122	0,62245	122
c = 2,5	14203	119	0,60714	119
c = 2,6	13512	115	0,58673	117
c = 2,7	12864	112	0,57143	115
c = 2,8	12259	110	0,56122	111
c = 2,9	11602	109	0,55612	106
c = 3	11103	108	0,55102	103
c = 3,1	10587	105	0,53571	101
c = 3,2	10122	105	0,53571	96
c = 3,3	9686	105	0,53571	92
c = 3,4	9183	104	0,53061	88
c = 3,5	8794	103	0,52551	85
c = 3,6	8385	102	0,52041	82
c = 3,7	8017	101	0,51531	79
c = 3,8	7673	97	0,4949	79
c = 3,9	7332	96	0,4898	76
c = 4	7006	95	0,48469	74

Tabela 52 Desempenho do Método do Quartil Modificado (2MA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	38067	184	0,93878	207
c = 0,1	38059	184	0,93878	207
c = 0,2	37982	184	0,93878	206
c = 0,3	37768	182	0,92857	208
c = 0,4	37277	180	0,91837	207
c = 0,5	36470	178	0,90816	205
c = 0,6	35456	178	0,90816	199
c = 0,7	34426	176	0,89796	196
c = 0,8	33426	174	0,88776	192
c = 0,9	32238	170	0,86735	190
c = 1	30161	166	0,84694	182
c = 1,1	29004	159	0,81122	182
c = 1,2	27832	157	0,80102	177
c = 1,3	26777	156	0,79592	172
c = 1,4	25757	152	0,77551	169
c = 1,5	24835	146	0,7449	170
c = 1,6	23894	145	0,7398	165
c = 1,7	23004	143	0,72959	161
c = 1,8	22160	142	0,72449	156
c = 1,9	21310	141	0,71939	151
c = 2	20285	135	0,68878	150
c = 2,1	19496	130	0,66327	150
c = 2,2	18748	129	0,65816	145
c = 2,3	18087	127	0,64796	142
c = 2,4	17480	127	0,64796	138
c = 2,5	16940	125	0,63776	136
c = 2,6	16367	121	0,61735	135
c = 2,7	15862	118	0,60204	134
c = 2,8	15301	116	0,59184	132
c = 2,9	14781	115	0,58673	129
c = 3	14322	114	0,58163	126
c = 3,1	13885	111	0,56633	125
c = 3,2	13478	111	0,56633	121
c = 3,3	13071	111	0,56633	118
c = 3,4	12725	110	0,56122	116
c = 3,5	12372	110	0,56122	112
c = 3,6	11999	108	0,55102	111
c = 3,7	11674	107	0,54592	109
c = 3,8	11353	103	0,52551	110
c = 3,9	11072	102	0,52041	109
c = 4	10787	101	0,51531	107

Tabela 53 Desempenho do Método do Desvio Mediano Absoluto (2MA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	49352	196	1	252
c = 0,1	47639	194	0,9898	246
c = 0,2	45724	190	0,96939	241
c = 0,3	43941	187	0,95408	235
c = 0,4	42160	185	0,94388	228
c = 0,5	40362	184	0,93878	219
c = 0,6	38648	182	0,92857	212
c = 0,7	37007	182	0,92857	203
c = 0,8	35405	177	0,90306	200
c = 0,9	33894	174	0,88776	195
c = 1	31060	164	0,83673	189
c = 1,1	29632	155	0,79082	191
c = 1,2	28295	154	0,78571	184
c = 1,3	27005	150	0,76531	180
c = 1,4	25877	147	0,75	176
c = 1,5	24752	145	0,7398	171
c = 1,6	23694	136	0,69388	174
c = 1,7	22706	133	0,67857	171
c = 1,8	21709	130	0,66327	167
c = 1,9	20818	130	0,66327	160
c = 2	19875	130	0,66327	153
c = 2,1	19014	129	0,65816	147
c = 2,2	18253	126	0,64286	145
c = 2,3	17505	124	0,63265	141
c = 2,4	16810	122	0,62245	138
c = 2,5	16151	118	0,60204	137
c = 2,6	15505	117	0,59694	133
c = 2,7	14900	114	0,58163	131
c = 2,8	14327	113	0,57653	127
c = 2,9	13760	112	0,57143	123
c = 3	13242	105	0,53571	126
c = 3,1	12782	104	0,53061	123
c = 3,2	12335	104	0,53061	119
c = 3,3	11901	103	0,52551	116
c = 3,4	11499	102	0,52041	113
c = 3,5	11152	101	0,51531	110
c = 3,6	10778	101	0,51531	107
c = 3,7	10450	96	0,4898	109
c = 3,8	10108	94	0,47959	108
c = 3,9	9810	92	0,46939	107
c = 4	9509	91	0,46429	104

Tabela 54 Desempenho do Algoritmo de Tukey (2MA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	54487	191	0,97449	285
c = 0,1	50978	186	0,94898	274
c = 0,2	47386	182	0,92857	260
c = 0,3	44028	176	0,89796	250
c = 0,4	40692	170	0,86735	239
c = 0,5	37541	165	0,84184	228
c = 0,6	34646	164	0,83673	211
c = 0,7	31947	160	0,81633	200
c = 0,8	29410	154	0,78571	191
c = 0,9	27177	152	0,77551	179
c = 1	24802	136	0,69388	182
c = 1,1	22749	131	0,66837	174
c = 1,2	20927	129	0,65816	162
c = 1,3	19198	127	0,64796	151
c = 1,4	17651	121	0,61735	146
c = 1,5	16256	116	0,59184	140
c = 1,6	14873	109	0,55612	136
c = 1,7	13672	104	0,53061	131
c = 1,8	12586	103	0,52551	122
c = 1,9	11653	97	0,4949	120
c = 2	10809	91	0,46429	119
c = 2,1	10058	90	0,45918	112
c = 2,2	9258	87	0,44388	106
c = 2,3	8560	81	0,41327	106
c = 2,4	7948	77	0,39286	103
c = 2,5	7386	76	0,38776	97
c = 2,6	6864	74	0,37755	93
c = 2,7	6405	73	0,37245	88
c = 2,8	5979	71	0,36224	84
c = 2,9	5569	70	0,35714	80
c = 3	5212	69	0,35204	76
c = 3,1	4892	69	0,35204	71
c = 3,2	4593	68	0,34694	68
c = 3,3	4278	66	0,33673	65
c = 3,4	4000	63	0,32143	63
c = 3,5	3762	62	0,31633	61
c = 3,6	3547	61	0,31122	58
c = 3,7	3327	58	0,29592	57
c = 3,8	3155	57	0,29082	55
c = 3,9	2979	56	0,28571	53
c = 4	2812	55	0,28061	51



Tabela 55 Desempenho do Método de Cercas Resistentes (MMAA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	27319	158	0,80612	173
c = 0,1	24179	145	0,7398	167
c = 0,2	21451	138	0,70408	155
c = 0,3	18879	130	0,66327	145
c = 0,4	16669	122	0,62245	137
c = 0,5	14610	113	0,57653	129
c = 0,6	12825	99	0,5051	130
c = 0,7	11409	94	0,47959	121
c = 0,8	10197	90	0,45918	113
c = 0,9	9168	84	0,42857	109
c = 1	8179	80	0,40816	102
c = 1,1	7345	79	0,40306	93
c = 1,2	6657	76	0,38776	88
c = 1,3	6054	74	0,37755	82
c = 1,4	5524	71	0,36224	78
c = 1,5	5044	68	0,34694	74
c = 1,6	4640	65	0,33163	71
c = 1,7	4290	63	0,32143	68
c = 1,8	3958	61	0,31122	65
c = 1,9	3652	59	0,30102	62
c = 2	3411	57	0,29082	60
c = 2,1	3164	56	0,28571	57
c = 2,2	2961	55	0,28061	54
c = 2,3	2754	53	0,27041	52
c = 2,4	2581	51	0,2602	51
c = 2,5	2410	49	0,25	49
c = 2,6	2273	48	0,2449	47
c = 2,7	2152	48	0,2449	45
c = 2,8	2033	47	0,2398	43
c = 2,9	1919	46	0,23469	42
c = 3	1816	45	0,22959	40
c = 3,1	1729	45	0,22959	38
c = 3,2	1657	44	0,22449	38
c = 3,3	1569	44	0,22449	36
c = 3,4	1498	43	0,21939	35
c = 3,5	1440	42	0,21429	34
c = 3,6	1375	39	0,19898	35
c = 3,7	1321	38	0,19388	35
c = 3,8	1263	38	0,19388	33
c = 3,9	1209	36	0,18367	34
c = 4	1162	36	0,18367	32

Tabela 56 Desempenho do Método do Quartil (MMAA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	49187	193	0,98469	255
c = 0,1	46801	192	0,97959	244
c = 0,2	44549	187	0,95408	238
c = 0,3	42099	186	0,94898	226
c = 0,4	39805	184	0,93878	216
c = 0,5	37502	180	0,91837	208
c = 0,6	35237	172	0,87755	205
c = 0,7	33107	170	0,86735	195
c = 0,8	31112	168	0,85714	185
c = 0,9	29313	161	0,82143	182
c = 1	26234	151	0,77041	174
c = 1,1	24547	146	0,7449	168
c = 1,2	23093	140	0,71429	165
c = 1,3	21665	135	0,68878	160
c = 1,4	20337	133	0,67857	153
c = 1,5	19007	128	0,65306	148
c = 1,6	17802	124	0,63265	144
c = 1,7	16668	120	0,61224	139
c = 1,8	15627	116	0,59184	135
c = 1,9	14646	110	0,56122	133
c = 2	13513	106	0,54082	127
c = 2,1	12562	100	0,5102	126
c = 2,2	11797	98	0,5	120
c = 2,3	11085	95	0,48469	117
c = 2,4	10416	94	0,47959	111
c = 2,5	9794	91	0,46429	108
c = 2,6	9224	90	0,45918	102
c = 2,7	8674	87	0,44388	100
c = 2,8	8198	85	0,43367	96
c = 2,9	7760	83	0,42347	93
c = 3	7320	79	0,40306	93
c = 3,1	6924	75	0,38265	92
c = 3,2	6573	74	0,37755	89
c = 3,3	6220	74	0,37755	84
c = 3,4	5828	74	0,37755	79
c = 3,5	5516	72	0,36735	77
c = 3,6	5229	71	0,36224	74
c = 3,7	4951	70	0,35714	71
c = 3,8	4683	69	0,35204	68
c = 3,9	4467	67	0,34184	67
c = 4	4203	67	0,34184	63

Tabela 57 Desempenho do Método do Quartil Modificado (MMAA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	37047	173	0,88265	214
c = 0,1	37046	173	0,88265	214
c = 0,2	37000	173	0,88265	214
c = 0,3	36675	173	0,88265	212
c = 0,4	36019	173	0,88265	208
c = 0,5	34922	171	0,87245	204
c = 0,6	33497	165	0,84184	203
c = 0,7	32001	163	0,83163	196
c = 0,8	30476	162	0,82653	188
c = 0,9	29005	160	0,81633	181
c = 1	26234	151	0,77041	174
c = 1,1	24721	146	0,7449	169
c = 1,2	23445	140	0,71429	167
c = 1,3	22157	136	0,69388	163
c = 1,4	20997	134	0,68367	157
c = 1,5	19816	129	0,65816	154
c = 1,6	18735	125	0,63776	150
c = 1,7	17725	121	0,61735	146
c = 1,8	16767	117	0,59694	143
c = 1,9	15893	112	0,57143	142
c = 2	14746	107	0,54592	138
c = 2,1	13810	101	0,51531	137
c = 2,2	13100	99	0,5051	132
c = 2,3	12478	96	0,4898	130
c = 2,4	11884	94	0,47959	126
c = 2,5	11303	91	0,46429	124
c = 2,6	10775	90	0,45918	120
c = 2,7	10305	87	0,44388	118
c = 2,8	9853	85	0,43367	116
c = 2,9	9479	83	0,42347	114
c = 3	9068	80	0,40816	113
c = 3,1	8706	77	0,39286	113
c = 3,2	8335	76	0,38776	110
c = 3,3	8016	76	0,38776	105
c = 3,4	7704	76	0,38776	101
c = 3,5	7410	74	0,37755	100
c = 3,6	7139	73	0,37245	98
c = 3,7	6880	72	0,36735	96
c = 3,8	6619	71	0,36224	93
c = 3,9	6411	70	0,35714	92
c = 4	6155	70	0,35714	88

Tabela 58 Desempenho do Método do Desvio Mediano Absoltuo (MMAA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	49187	193	0,98469	255
c = 0,1	46843	191	0,97449	245
c = 0,2	44572	187	0,95408	238
c = 0,3	42231	185	0,94388	228
c = 0,4	39923	183	0,93367	218
c = 0,5	37611	179	0,91327	210
c = 0,6	35435	175	0,89286	202
c = 0,7	33314	169	0,86224	197
c = 0,8	31374	165	0,84184	190
c = 0,9	29526	158	0,80612	187
c = 1	26456	143	0,72959	185
c = 1,1	24831	141	0,71939	176
c = 1,2	23288	139	0,70918	168
c = 1,3	21827	133	0,67857	164
c = 1,4	20612	129	0,65816	160
c = 1,5	19365	125	0,63776	155
c = 1,6	18175	122	0,62245	149
c = 1,7	17052	114	0,58163	150
c = 1,8	15983	108	0,55102	148
c = 1,9	15072	106	0,54082	142
c = 2	14166	103	0,52551	138
c = 2,1	13318	101	0,51531	132
c = 2,2	12615	100	0,5102	126
c = 2,3	11894	98	0,5	121
c = 2,4	11240	96	0,4898	117
c = 2,5	10628	95	0,48469	112
c = 2,6	10031	92	0,46939	109
c = 2,7	9456	89	0,45408	106
c = 2,8	9009	88	0,44898	102
c = 2,9	8561	88	0,44898	97
c = 3	8152	87	0,44388	94
c = 3,1	7766	87	0,44388	89
c = 3,2	7412	85	0,43367	87
c = 3,3	7068	84	0,42857	84
c = 3,4	6767	82	0,41837	83
c = 3,5	6458	78	0,39796	83
c = 3,6	6179	76	0,38776	81
c = 3,7	5925	75	0,38265	79
c = 3,8	5671	72	0,36735	79
c = 3,9	5447	70	0,35714	78
c = 4	5229	68	0,34694	77

Tabela 59 Desempenho do Algoritmo de Tukey (MMAA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	55178	195	0,9949	283
c = 0,1	50823	192	0,97959	265
c = 0,2	46709	186	0,94898	251
c = 0,3	42722	179	0,91327	239
c = 0,4	38936	174	0,88776	224
c = 0,5	35342	169	0,86224	209
c = 0,6	31952	162	0,82653	197
c = 0,7	28939	157	0,80102	184
c = 0,8	26284	155	0,79082	170
c = 0,9	23798	153	0,78061	156
c = 1	20608	128	0,65306	161
c = 1,1	18489	122	0,62245	152
c = 1,2	16663	117	0,59694	142
c = 1,3	14932	110	0,56122	136
c = 1,4	13426	104	0,53061	129
c = 1,5	12063	103	0,52551	117
c = 1,6	10852	98	0,5	111
c = 1,7	9781	92	0,46939	106
c = 1,8	8835	87	0,44388	102
c = 1,9	7961	83	0,42347	96
c = 2	7183	82	0,41837	88
c = 2,1	6468	77	0,39286	84
c = 2,2	5847	76	0,38776	77
c = 2,3	5348	73	0,37245	73
c = 2,4	4842	66	0,33673	73
c = 2,5	4375	66	0,33673	66
c = 2,6	3978	62	0,31633	64
c = 2,7	3637	57	0,29082	64
c = 2,8	3321	55	0,28061	60
c = 2,9	3067	52	0,26531	59
c = 3	2806	51	0,2602	55
c = 3,1	2566	51	0,2602	50
c = 3,2	2361	47	0,2398	50
c = 3,3	2209	47	0,2398	47
c = 3,4	2040	46	0,23469	44
c = 3,5	1878	44	0,22449	43
c = 3,6	1736	43	0,21939	40
c = 3,7	1629	42	0,21429	39
c = 3,8	1512	39	0,19898	39
c = 3,9	1420	37	0,18878	38
c = 4	1336	35	0,17857	38

Tabela 60 Desempenho do Método de Cercas Resistentes (TA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	25397	161	0,82143	158
c = 0,1	22153	154	0,78571	144
c = 0,2	19366	147	0,75	132
c = 0,3	16941	142	0,72449	119
c = 0,4	14692	135	0,68878	109
c = 0,5	12707	124	0,63265	102
c = 0,6	11077	122	0,62245	91
c = 0,7	9681	113	0,57653	86
c = 0,8	8444	109	0,55612	77
c = 0,9	7387	106	0,54082	70
c = 1	6501	102	0,52041	64
c = 1,1	5708	96	0,4898	59
c = 1,2	5102	93	0,47449	55
c = 1,3	4566	86	0,43878	53
c = 1,4	4070	83	0,42347	49
c = 1,5	3673	81	0,41327	45
c = 1,6	3298	76	0,38776	43
c = 1,7	2979	73	0,37245	41
c = 1,8	2692	69	0,35204	39
c = 1,9	2429	66	0,33673	37
c = 2	2230	63	0,32143	35
c = 2,1	2043	61	0,31122	33
c = 2,2	1899	60	0,30612	32
c = 2,3	1751	58	0,29592	30
c = 2,4	1629	56	0,28571	29
c = 2,5	1502	54	0,27551	28
c = 2,6	1411	53	0,27041	27
c = 2,7	1312	53	0,27041	25
c = 2,8	1211	50	0,2551	24
c = 2,9	1142	50	0,2551	23
c = 3	1075	49	0,25	22
c = 3,1	996	47	0,2398	21
c = 3,2	947	46	0,23469	21
c = 3,3	889	46	0,23469	19
c = 3,4	841	45	0,22959	19
c = 3,5	802	45	0,22959	18
c = 3,6	761	44	0,22449	17
c = 3,7	727	44	0,22449	17
c = 3,8	684	44	0,22449	16
c = 3,9	645	44	0,22449	15
c = 4	601	43	0,21939	14

Tabela 61 Desempenho do Método do Quartil (TA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	44600	195	0,9949	229
c = 0,1	42644	193	0,98469	221
c = 0,2	40767	190	0,96939	215
c = 0,3	38721	188	0,95918	206
c = 0,4	36660	186	0,94898	197
c = 0,5	34583	181	0,92347	191
c = 0,6	32506	178	0,90816	183
c = 0,7	30540	173	0,88265	177
c = 0,8	28638	168	0,85714	170
c = 0,9	26822	166	0,84694	162
c = 1	25036	161	0,82143	156
c = 1,1	23357	158	0,80612	148
c = 1,2	21823	154	0,78571	142
c = 1,3	20380	152	0,77551	134
c = 1,4	18970	149	0,7602	127
c = 1,5	17737	147	0,75	121
c = 1,6	16578	144	0,73469	115
c = 1,7	15441	140	0,71429	110
c = 1,8	14457	140	0,71429	103
c = 1,9	13469	136	0,69388	99
c = 2	12523	132	0,67347	95
c = 2,1	11620	124	0,63265	94
c = 2,2	10834	119	0,60714	91
c = 2,3	10067	118	0,60204	85
c = 2,4	9394	113	0,57653	83
c = 2,5	8780	108	0,55102	81
c = 2,6	8230	107	0,54592	77
c = 2,7	7659	107	0,54592	72
c = 2,8	7161	103	0,52551	70
c = 2,9	6666	101	0,51531	66
c = 3	6195	96	0,4898	65
c = 3,1	5786	92	0,46939	63
c = 3,2	5434	90	0,45918	60
c = 3,3	5094	86	0,43878	59
c = 3,4	4728	85	0,43367	56
c = 3,5	4431	85	0,43367	52
c = 3,6	4170	83	0,42347	50
c = 3,7	3942	82	0,41837	48
c = 3,8	3700	80	0,40816	46
c = 3,9	3491	77	0,39286	45
c = 4	3259	77	0,39286	42

Tabela 62 Desempenho do Método do Quartil Modificado (TA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	35554	186	0,94898	191
c = 0,1	35554	186	0,94898	191
c = 0,2	35537	186	0,94898	191
c = 0,3	35306	185	0,94388	191
c = 0,4	34577	183	0,93367	189
c = 0,5	33384	181	0,92347	184
c = 0,6	31783	178	0,90816	179
c = 0,7	30090	173	0,88265	174
c = 0,8	28390	168	0,85714	169
c = 0,9	26700	166	0,84694	161
c = 1	25036	161	0,82143	156
c = 1,1	23452	158	0,80612	148
c = 1,2	21998	155	0,79082	142
c = 1,3	20623	153	0,78061	135
c = 1,4	19278	149	0,7602	129
c = 1,5	18090	147	0,75	123
c = 1,6	16992	144	0,73469	118
c = 1,7	15889	140	0,71429	113
c = 1,8	14928	140	0,71429	107
c = 1,9	13973	136	0,69388	103
c = 2	13049	132	0,67347	99
c = 2,1	12177	125	0,63776	97
c = 2,2	11414	120	0,61224	95
c = 2,3	10710	118	0,60204	91
c = 2,4	10045	115	0,58673	87
c = 2,5	9440	112	0,57143	84
c = 2,6	8897	111	0,56633	80
c = 2,7	8366	110	0,56122	76
c = 2,8	7839	105	0,53571	75
c = 2,9	7388	103	0,52551	72
c = 3	6914	98	0,5	71
c = 3,1	6499	95	0,48469	68
c = 3,2	6142	93	0,47449	66
c = 3,3	5804	89	0,45408	65
c = 3,4	5492	87	0,44388	63
c = 3,5	5179	87	0,44388	60
c = 3,6	4922	85	0,43367	58
c = 3,7	4703	84	0,42857	56
c = 3,8	4460	82	0,41837	54
c = 3,9	4253	79	0,40306	54
c = 4	4031	79	0,40306	51



Tabela 63 Desempenho do Método do Desvio Mediano Absoluto (TA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	44600	195	0,9949	229
c = 0,1	42629	193	0,98469	221
c = 0,2	40811	190	0,96939	215
c = 0,3	38845	188	0,95918	207
c = 0,4	36743	188	0,95918	195
c = 0,5	34682	180	0,91837	193
c = 0,6	32610	177	0,90306	184
c = 0,7	30624	169	0,86224	181
c = 0,8	28784	166	0,84694	173
c = 0,9	26939	164	0,83673	164
c = 1	25084	160	0,81633	157
c = 1,1	23461	156	0,79592	150
c = 1,2	21867	151	0,77041	145
c = 1,3	20428	148	0,7551	138
c = 1,4	19069	146	0,7449	131
c = 1,5	17748	141	0,71939	126
c = 1,6	16528	137	0,69898	121
c = 1,7	15343	136	0,69388	113
c = 1,8	14357	132	0,67347	109
c = 1,9	13326	128	0,65306	104
c = 2	12369	125	0,63776	99
c = 2,1	11561	123	0,62755	94
c = 2,2	10770	119	0,60714	91
c = 2,3	10051	117	0,59694	86
c = 2,4	9430	114	0,58163	83
c = 2,5	8823	112	0,57143	79
c = 2,6	8222	109	0,55612	75
c = 2,7	7740	108	0,55102	72
c = 2,8	7230	107	0,54592	68
c = 2,9	6722	106	0,54082	63
c = 3	6324	101	0,51531	63
c = 3,1	5981	98	0,5	61
c = 3,2	5645	95	0,48469	59
c = 3,3	5318	93	0,47449	57
c = 3,4	5042	91	0,46429	55
c = 3,5	4764	90	0,45918	53
c = 3,6	4488	88	0,44898	51
c = 3,7	4264	86	0,43878	50
c = 3,8	4049	83	0,42347	49
c = 3,9	3853	81	0,41327	48
c = 4	3667	80	0,40816	46

Tabela 64 Desempenho do Algoritmo de Tukey (TA) no Município do São Paulo

c	Outliers Id.	Outliers Id. Corretamente	Prop. de Acerto	Falsos Positivos
c = 0	55520	195	0,9949	285
c = 0,1	52183	193	0,98469	270
c = 0,2	43113	188	0,95918	229
c = 0,3	36520	185	0,94388	197
c = 0,4	33409	177	0,90306	189
c = 0,5	30522	176	0,89796	173
c = 0,6	27716	171	0,87245	162
c = 0,7	25116	166	0,84694	151
c = 0,8	22710	159	0,81122	143
c = 0,9	20515	151	0,77041	136
c = 1	18464	141	0,71939	131
c = 1,1	16649	138	0,70408	121
c = 1,2	14895	134	0,68367	111
c = 1,3	13346	130	0,66327	103
c = 1,4	11922	124	0,63265	96
c = 1,5	10649	123	0,62755	87
c = 1,6	9517	118	0,60204	81
c = 1,7	8481	115	0,58673	74
c = 1,8	7561	110	0,56122	69
c = 1,9	6750	106	0,54082	64
c = 2	6011	101	0,51531	60
c = 2,1	5348	98	0,5	55
c = 2,2	4750	95	0,48469	50
c = 2,3	4247	91	0,46429	47
c = 2,4	3790	87	0,44388	44
c = 2,5	3404	86	0,43878	40
c = 2,6	3075	80	0,40816	38
c = 2,7	2773	78	0,39796	36
c = 2,8	2498	74	0,37755	34
c = 2,9	2259	73	0,37245	31
c = 3	2009	69	0,35204	29
c = 3,1	1819	67	0,34184	27
c = 3,2	1650	64	0,32653	26
c = 3,3	1503	63	0,32143	24
c = 3,4	1356	60	0,30612	23
c = 3,5	1244	58	0,29592	21
c = 3,6	1126	58	0,29592	19
c = 3,7	1035	56	0,28571	18
c = 3,8	954	52	0,26531	18
c = 3,9	877	51	0,2602	17
c = 4	818	49	0,25	17