UNIVERSIDADE FEDERAL DE PERNAMBUCO

CENTRO ACADÊMICO DA VITÓRIA

ARIANNE KARLA FERREIRA TAVARES

OUTLIERS: UMA ABORDAGEM POR MEIO DA CAPACIDADE FISCAL E DE INVESTIMENTO EM SAÚDE

VITÓRIA DE SANTO ANTÃO

2022

UNIVERSIDADE FEDERAL DE PERNAMBUCO

CENTRO ACADÊMICO DA VITÓRIA

CURSO DE BACHARELADO EM SAÚDE COLETIVA

NÚCLEO DE SAÚDE COLETIVA

ARIANNE KARLA FERREIRA TAVARES

OUTLIERS: UMA ABORDAGEM POR MEIO DA CAPACIDADE FISCAL E DE INVESTIMENTO EM SAÚDE

TCC apresentado ao Curso de Saúde Coletiva da Universidade Federal de Pernambuco, Centro Acadêmico de Vitória, como requisito para a obtenção do título de Bacharelado em Saúde Coletiva.

Orientador: Dr. Carlos Renato dos Santos

VITÓRIA DE SANTO ANTÃO

2022

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Tavares, Arianne Karla Ferreira.
Outliers: uma abordagem por meio da capacidade fiscal e de investimento em saúde / Arianne Karla Ferreira Tavares. - Vitória de Santo Antão, 2022.
43 : il., tab.

Orientador(a): Carlos Renato dos Santos
Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro Acadêmico de Vitória, Saúde Coletiva, 2022.
Inclui referências, anexos.

1. Outliers. 2. Siops. 3. Investimento em Saúde. 4. Orçamento Fiscal. I. Santos, Carlos Renato dos. (Orientação). II. Título.

310 CDD (22.ed.)

ARIANNE KARLA FERREIRA TAVARES

OUTLIERS: UMA ABORDAGEM POR MEIO DA CAPACIDADE FISCAL E DE INVESTIMENTO EM SAÚDE

TCC apresentado ao Curso de Saúde Coletiva da Universidade Federal de Pernambuco, Centro Acadêmico de Vitória, como requisito para a obtenção do título de Bacharelado em Saúde Coletiva.

Aprovado em: 28/10/2022.

BANCA EXAMINADORA

_____
Profº. Dr. Carlos Renato dos Santos (Orientador)
Universidade Federal de Pernambuco

_____
Profº. Dra. Amanda Priscila de Santana Cabral Silva (Examinador Interno)
Universidade Federal de Pernambuco

_____
Profº. Dra. Simara Lopes Cruz Damásio (Examinador Externo)
Universidade Federal de Pernambuco

Dedico está monografia à minha família.

AGRADECIMENTOS

Agradeço primeiramente a Deus e a Nossa Senhora de Fátima por terem sido meu refúgio e minha fortaleza nos momentos mais difíceis.

A minha mãe Maria Paulina Ferreira que sempre esteve ao meu lado me dando forças e me motivando ao longo de toda a minha trajetória.

Agradeço de uma forma especial ao meu orientador, o professor Carlos Renato dos Santos, que contribuiu com seu enorme conhecimento no campo da estatística durante a graduação e no processo de construção do Trabalho de Conclusão de Curso, você é incrível.

As minhas irmãs Adriana Ferreira Tavares e Alexsandra Ferreira Tavares pela compreensão, paciência e espera todas as noites durante os 4 anos de graduação.

Ao meu namorado Artur Lima da Silva que esteve do meu lado me apoiando e incentivando na minha jornada.

Também agradeço às minhas amigas Suellen, Tathyana, Gabriela, Larissa, Luana, Juliana, Laís, Vívian e Rirlany, vocês fizeram a caminhada ser mais leve, gratidão!

A todos os meus professores do Curso de Saúde Coletiva da Universidade Federal de Pernambuco.

RESUMO

Outliers são observações que na maioria dos casos se mostram inconsistentes com o restante do conjunto de dados, contudo, sua identificação tem fundamental importância, visando identificá-los para explorar seus possíveis padrões de anormalidade. Nessa perspectiva, o objetivo desta pesquisa é avaliar a ocorrência de *outliers* no Sistema de Informações Sobre Orçamentos Públicos em Saúde (SIOPS) e seu impacto na qualidade da informação em indicadores referentes à capacidade fiscal e de investimentos em saúde no período de 2002 a 2021, totalizando 20 anos. Trata-se de um estudo descritivo, transversal, com abordagem quantitativa e inferencial. O local do estudo é o Brasil em suas municipalidades, ou seja, a partir da varredura exaustiva dos 5.569 municípios disponíveis no referido sistema. Para a análise dos dados foram utilizados o Programa R, por meio dos pacotes *outliers* para execução dos testes de Grubbs e Dixon, o *EnvStats* para execução do teste de Rosner, o *caret* para obtenção das medidas F1 e Kappa, além da planilha eletrônica Excel. Também foi necessária uma avaliação da concordância destes no sentido de validar os achados e compreender melhor os municípios com *outliers* detectados. Os resultados encontrados evidenciam que o SIOPS apresentou uma alta completitude no período estudado. Para o indicador 1.1, o teste de Grubbs encontrou 1523 (29,1%) municípios que tem algum *outlier,* por Rosner 1048 (26,9%) e por Dixon 1410 (20%). Para o indicador 1.2, o teste de Grubbs encontrou 823 (15,7%) municípios que tem algum *outlier,* por Rosner 788 (15%) e por Dixon 682 (13%). Para o indicador 2.1, o teste de Grubbs encontrou 437 (8,3%) municípios que tem algum *outlier*, por Rosner 323 (6,2%) e por Dixon 201 (3,8%). Para o indicador 2.2, o teste de Grubbs encontrou 749 (14,3%) municípios que tem algum *outlier,* por Rosner 748 (14,3%) e por Dixon 593 (11,3%). Para o indicador 3.1, o teste de Grubbs encontrou 1848 (35,3%) municípios que tem algum *outlier*, por Rosner 1731 (33,2%) e por Dixon 1240 (23,8%). Portanto, os métodos aplicados para detecção de *outliers* se mostraram eficientes na observação do comportamento dos municípios.

Palavras-chave: *outliers*; SIOPS; investimento em saúde; orçamento fiscal.

ABSTRACT

Outliers are observations that in most cases are inconsistent with the rest of the data set, however, their identification is of fundamental importance, in order to identify them to explore their possible patterns of abnormality. From this perspective, the objective of this research is to evaluate the occurrence of outliers in the Information System on Public Health Budgets (SIOPS) and their impact on the quality of information in indicators referring to fiscal capacity and health investments in the period from 2002 to 2021, totaling 20 years. This is a descriptive, cross-sectional study with a quantitative and inferential approach. The study site is Brazil in its municipalities, that is, from the exhaustive scan of the 5,569 municipalities available in the referred system. For data analysis, the R Program was used, through outliers packages to perform the Grubbs and Dixon tests, EnvStats to perform the Rosner test, the caret to obtain the F1 and Kappa measurements, in addition to the Excel spreadsheet. It was also necessary to assess their agreement in order to validate the findings and better understand the municipalities with detected outliers. The results found show that the SIOPS presented a high completeness in the studied period. For indicator 1.1, the Grubbs test found 1523 (29.1%) municipalities that have some outlier, by Rosner 1048 (26.9%) and by Dixon 1410 (20%). For indicator 1.2, the Grubbs test found 823 (15.7%) municipalities that have some outlier, by Rosner 788 (15%) and by Dixon 682 (13%). For indicator 2.1, the Grubbs test found 437 (8.3%) municipalities that have some outlier, by Rosner 323 (6.2%) and by Dixon 201 (3.8%). For indicator 2.2, the Grubbs test found 749 (14.3%) municipalities that have some outlier, by Rosner 748 (14.3%) and by Dixon 593 (11.3%). For indicator 3.1, the Grubbs test found 1848 (35.3%) municipalities that have some outlier, by Rosner 1731 (33.2%) and by Dixon 1240 (23.8%). Therefore, the methods applied to detect outliers proved to be efficient in observing the behavior of municipalities.

Keywords: *outliers;* SIOPS; investment in health; fiscal budget.

LISTA DE ILUSTRAÇÕES

Figura 1 – Ilustração de um *outlier* global. 15

Figura 2 – Ilustração de um *outlier* contextual. 16

Figura 3 – Ilustração de um *outlier* coletivo. 16

Figura 4 – Ilustração do cálculo da precisão e do *recall.* 24

Figura 5 – Diagrama de Venn da quantidade de municípios com *outliers* detectados por diferentes testes para capacidade de arrecadação municipal. 28

Figura 6 – Diagrama de Venn da quantidade de municípios com *outliers* detectados por diferentes testes para a dependência municipal. 29

Figura 7 – Diagrama de Venn da quantidade de municípios com *outliers* detectados por diferentes testes para o gasto per capita com saúde. 31

Figura 8 – Diagrama de Venn da quantidade de municípios com *outliers* detectados por diferentes testes para o percentual da despesa total com saúde em relação ao gasto total com saúde. 32

Figura 9 - Diagrama de Venn da quantidade de municípios com *outliers* detectados por diferentes testes para o percentual de despesas com saúde financiadas com recursos transferidos por outras esferas de governo. 33

Gráfico 1 – MMC x PCMU. 35

Gráfico 2 – Percentual de concordância dois a dois entre os métodos nos indicadores avaliados. 36

Gráfico 3 – Percentual de ocorrência de *outliers* nas regiões do país Norte, Nordeste, Sudeste, Sul e Centro-oeste 37

Quadro 1 – Indicadores referentes aos valores de receitas, despesas e o relacionado a receitas e despesas. 20

Quadro 2 – Completitude dos indicadores estudados. 26

Quadro 3 – Nulidade dos indicadores estudados. 27

LISTA DE TABELAS

Tabela 1 – Classificação da concordância a partir de Kappa. 23

Tabela 2 – Matriz de confusão genérica. 24

Tabela 3 – Análise de concordância entre os diferentes testes para capacidade de arrecadação municipal. 28

Tabela 4 – Análise de concordância entre os diferentes testes para a dependência municipal. 30

Tabela 5 – Análise de concordância entre os diferentes testes para o gasto per capita com saúde. 31

Tabela 6 – Análise de concordância entre os diferentes testes para o percentual da despesa total com saúde em relação ao gasto total com saúde. 32

Tabela 7 – Análise de concordância entre os diferentes testes para o percentual de despesas com saúde financiadas com recursos transferidos por outras esferas de governo. 34

Tabela 8 – Análise conjunta dos cinco indicadores 36

LISTA DE SIGLAS

ABS	Atenção Básica em Saúde
APS	Atenção Primária à Saúde
ASPS	Ações e Serviços Públicos de Saúde
CAM	Capacidade de Arrecadação Municipal
CONEP	Comissão Nacional de Ética em Pesquisa
DEA	Data Envelopment Analysis
DEP	Despesa Total com Saúde em Relação ao Gasto Total com Saúde
DM	Mineração de Dados
DSO	Despesas com Saúde Financiadas com Recursos Transferidos por outras Esferas de Governo
FN	Falsos Negativos
FP	Falsos Positivos
GDM	Grau de Dependência Municipal
GPS	Gasto Per Capita com Saúde
MMC	Municípios com Mútua Concordância
PCMU	Percentual de Concordância Mútua
PFDC	Procuradoria Federal dos Direitos do Cidadão
SIOPS	Sistema de Informações sobre Orçamentos Públicos em Saúde
SNIS	Sistema Nacional de Informações sobre Saneamento
SUS	Sistema Único de Saúde
VN	Verdadeiros Negativos
VP	Verdadeiros Positivos

SUMÁRIO

1 INTRODUÇÃO 12

2 REVISÃO DE LITERATURA 15

2.1 Tipos de *Outliers* 15

2.2 *Outliers* em Saúde 17

3 OBJETIVOS 19

3.1 Objetivo Geral 19

3.2 Objetivos Específicos: 19

4 METODOLOGIA 20

4.1 Indicadores estudados 20

4.2 Preparação da base de dados 21

4.3 Recursos Computacionais e testes de detecção 22

4.4 Medidas de concordância 22

4.5 Aspectos Éticos 25

5 RESULTADOS E DISCUSSÃO 26

5.1 Análise da completitude das informações 26

5.2 Análise da nulidade das informações no conjunto completo 26

5.3 Análise descritiva e de concordância da quantidade de *outliers* detectados pelos diferentes testes 27

5.3.1 Análise conjunta dos indicadores - Municípios com mútua concordância 34

5.3.2 Análise conjunta dos indicadores - Análise regional 36

6 CONCLUSÃO 38

REFERÊNCIAS 40

1 INTRODUÇÃO

O termo “*Outliers*” tem sido descrito por Freitas (2019), como um objeto que se afasta significativamente do restante dos dados. Em consonância com sua visão, autores como Camacho et al., (2016) também se referem aos *outliers* como observações que na maioria dos casos se mostram inconsistentes com o restante do conjunto de dados, mas que apresentam uma importância ímpar no processo de identificação e busca de possíveis padrões de anormalidade.

De uma forma geral os *outliers* podem ser encontrados a partir de erros de medição, variação natural, medidas de dados e erros de coleta dados de diferentes classes (FREITAS, 2019; SOUZA; MAGALHAES; AQUINO; GOMES, 2018).

Dentre as vantagens dos estudos aplicados nessa área estão a identificação de fraudes, análise de desempenho, detecção de intrusão, fraude de cartão de crédito, sensores de eventos e diagnóstico médico. Todo esse leque de possibilidades e benefícios para a sociedade são pautadas nas informações sobre características anormais dos sistemas e entidades, que afetam o processo de geração de dados (AGGARWAL, 2017; BARBOSA, 2017).

Estritamente no âmbito da saúde também há uma preocupação com os *outliers,* considerando que é essencial entendê-los no campo da análise de dados, pois estes podem comprometer de forma negativa os resultados de uma análise, do mesmo modo podem ser exatamente o que se deseja encontrar (NEPOMUCENA; CIRILLO, 2012).

Por este motivo, estudar *outliers* em saúde é de extrema relevância em uma análise de dados, pois a presença deles impactam duas importantes dimensões da qualidade da informação: a completitude e a confiabilidade. Segundo Damázio (2022), a completitude é definida como o grau de valores não nulos. Já a confiabilidade é definida como o grau de concordância entre diferentes medidas.

Uma questão importante a ser considerada é a capacidade fiscal que pode ser definida como a habilidade seja do governo, de grupos ou instituições para gerar receita. Já a capacidade de investimento são as despesas correntes anuais, sendo uma variável importante na avaliação da autonomia orçamentária e perspectivas de desenvolvimento, considerando que investimento é qualquer gasto ou aplicação de recursos que produza um retorno futuro. Nesse sentido, é notório a magnitude do

investimento em saúde, tendo em vista que, a análise de indicadores de orçamento e gasto possibilita a discussão da disponibilidade e composição das receitas e despesas, tanto do ponto de vista da política fiscal, quanto do financiamento da política de saúde. Além disso, o financiamento significa alocar mais recursos para o funcionamento dos serviços com qualidade, para a ampliação da oferta e a melhoria da gestão (FELICIANO, 2019).

Em 1993, no Conselho Nacional de Saúde teve a ideia de criar um sistema que fornecesse informações com relação às despesas em saúde para todos os entes federativos. A Procuradoria Federal dos Direitos do Cidadão (PFDC) julgou relevante a proposta nos Inquéritos Civis Públicos de nº 001/ 94 e 002/ 94 relativos ao funcionamento e financiamento do Sistema Único de Saúde (SUS) (BRASIL, 2021).

Em 1999, o Ministério da Saúde criou o Sistema de Informações Sobre Orçamentos Públicos em Saúde (SIOPS), que reúne informações sobre o financiamento de gastos com saúde pública referentes às receitas e despesas dos municípios, dos estados e da União. Assim, as informações dispostas podem ser vistas de acordo com o nível de agregação esperado, com o custo e a capacidade de financiamento das ações e serviços públicos de saúde, nas três esferas de governo local, estadual e federal (TEIXEIRA; TEIXEIRA, 2003).

Constituindo um banco de dados único no âmbito das políticas públicas no Brasil, o SIOPS gera informações de forma regular e com qualidade crescente, mostrando que o sistema é uma relevante ferramenta para a realização de estudos por instituições de pesquisa, órgãos de controle e gestão, a fim de acompanhar e fiscalizar as ações e investimentos no âmbito do Sistema Único de Saúde (SUS) (TEIXEIRA, 2003).

Desse modo, os objetivos específicos do SIOPS são:

> consolidar informações de gasto público em Saúde no Brasil de forma a propiciar insumos para a melhoria da gestão, diagnósticos sobre o setor e formulação de políticas públicas; tornar mais ágil a consulta online sobre as receitas totais e despesas com Ações e Serviços Públicos de Saúde (ASPS); promover a transparência ao assegurar, tanto a Conselhos de Saúde como a cidadãos, o acesso público às informações, atendendo ao disposto na Lei nº 12.527/2011, e contribuir para a correta aplicação dos recursos públicos no setor; fornecer indicadores sobre gastos em saúde no país, proporcionando a toda a população o conhecimento sobre quanto cada unidade político-administrativa do país tem aplicado na área; permitir o monitoramento da alocação de recursos próprios em ações e serviços públicos de saúde, de acordo com as regras constitucionais de aplicação mínima em saúde; dimensionar a participação da União, Estados e Municípios no financiamento da saúde pública (BRASIL, 2021, p. 2).

Para tanto, o presente trabalho justificou-se pela necessidade de trabalhar a temática de *outliers* em saúde mediante ao cenário da capacidade fiscal e de investimentos realizados nos municípios brasileiros, através do banco de dados do Sistema de Informações Sobre Orçamentos Públicos em Saúde (SIOPS). O levantamento de informações subsidiara a elaboração de novos estudos, bem como foi cooperado para o melhoramento do conhecimento de profissionais, acadêmicos da área de saúde coletiva e população em geral.

Diante do exposto surge a seguinte pergunta norteadora: Como os *outliers* de receitas e despesas se apresentam nas informações em saúde?

2 REVISÃO DE LITERATURA

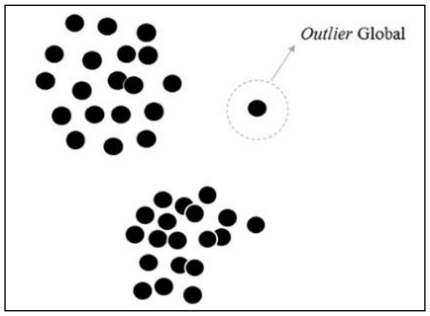
Nesta seção, serão abordados os tipos de *outliers* e *outliers* em saúde para compreensão deste trabalho.

2.1 Tipos de *Outliers*

Os *outliers* se classificam de três maneiras: globais (ou pontuais), contextuais (ou condicionais) e coletivos (FREITAS, 2019):

***Outlier* global:** é classificado quando um determinado objeto de estudo se desvia consideravelmente dos demais. Também conhecido como *outliers* pontuais, em decorrência da característica singular. Esse tipo de *outliers* é considerado o mais simples entre as três formas, como mostrado na Figura 1.

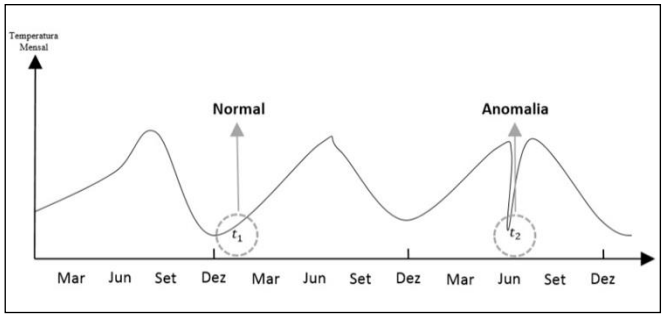
Figura 1 - Ilustração de um *outlier* global



Fonte: FREITAS, 2019.

- ***Outlier* contextual:** é denominado naqueles casos em que uma instância de dados é anômala apenas em um determinado contexto. Também chamado de *outliers* condicionais, tendo em vista que estão condicionados a um contexto específico. Assim, os atributos contextuais se definem como sendo aqueles que caracterizam o objeto em seu contexto e os atributos comportamentais definem as características do objeto e são úteis para avaliar se este objeto se trata de um *outlier* no contexto em que faz parte, como apresentado na figura 2.

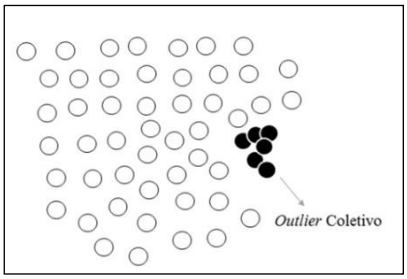
Figura 2 - Ilustração de um *outlier* contextual



Fonte: FREITAS, 2019.

· ***Outliers* coletivos:** são assim denominados por se tratarem de um subconjunto de dados considerado anormal assim que constituem um grupo. Logo, essa categoria é atribuída ao subconjunto de objetos que se desviam como um todo em relação ao conjunto de dados. A detecção desse tipo de *outlier* é muito importante para aplicações em que a anomalia está ligada ao comportamento de várias instâncias. Para tal, se faz necessário o conhecimento mais aprofundado do relacionamento entre os objetos, como medidas de distância ou similaridade. A figura 3 representa um exemplo desse tipo de *outlier.*

Figura 3 - Ilustração de um *outlier* coletivo



Fonte: FREITAS, 2019.

2.2 *Outliers* em Saúde

Os outliers requerem atenção especial, pois normalmente essas observações resultam de alguma violação das pressuposições necessárias para adequação ao modelo, produzindo consequentemente efeitos não confiáveis na eficiência dos estimadores (NEPOMUCENA; CIRILLO, 2012). O mesmo autor, em seu estudo, teve como objetivo de identificar outliers verificados nos dados amostrais de uma população de casos de HIV.

Outro estudo buscou mostrar o potencial do uso de análise de cluster para detecção de *outliers* utilizando técnicas estatísticas tradicionais em dados de acidentes de trânsito no Brasil, considerando que um dos aspectos que pode induzir ao enviesamento dos resultados é a não identificação ou tratamento de valores discrepantes, ou *outliers,* mostrando assim a importância da identificação e tratamento dos *outliers* (SILVA; FERREIRA; KOKKINOGENIS; ANDRADE, 2019).

Para tratar os *outliers* um estudo realizado para analisar a eficiência técnica dos serviços de água e esgoto no Brasil, utilizou a técnica *Jackstrap*, considerando que a base de dados do Sistema Nacional de Informações sobre Saneamento (SNIS), tem procedimento autodeclaratório que favorece alguns valores inconsistentes que pela sensibilidade do DEA aos valores extremos requer um cuidado especial com potenciais *outliers*. A técnica combina com o esquema *jackknife* e a técnica de reamostragem *bootstrap* de modo a automatizar a detecção de *outliers* com a vantagem de não possuir inspeções manuais o que viabiliza sua utilização em grandes amostras (CRUZ; MOTTA; MARINHO, 2019).

O estudo de Carvalho (2021) abordou os fatores políticos enquanto condicionantes na alocação de recursos federais no Sistema Único de Saúde (SUS) e seus efeitos nas desigualdades regionais. Para obter resultados satisfatório, viu-se a necessidade de aprofundar o estudo com objetivo de identificar os *outliers*, através do meio de análise comparada qualitativa, especificamente com o método fuzzy-setQCA, que possibilita a identificação de condições necessárias e/ou suficientes para explicação de um fenômeno em estudo. Considerando que o uso de casos atípicos é adequado para a realização de análise exploratória e construção de teorias, observando-se o que se chama de tipos ideais, que inclusive ajudam a refinar e testar modelos estatísticos futuros.

O estudo de Gontijo e Reis (2021) buscou avaliar a eficiência do gasto na Atenção Primária à Saúde (APS) em alguns municípios paulistas, por meio de um

modelo de Data Envelopment Analysis Supereficiência em detecção e remoção de *outliers* em vários pontos, além de identificar que a exclusão dos outliers têm os efeitos de elevar que a eficiência média dos municípios restantes, assim, as variáveis se tornam menos sensíveis a observações discrepantes (*outliers*).

Igualmente o estudo de Rodrigues, Gontijo e Gonçalves (2021) avaliou a eficiência dos gastos públicos em atenção básica em saúde (ABS) nos municípios do Rio de Janeiro, também por meio do Data Envelopment Analysis Supereficiência (DEA) para identificar possíveis *outliers,* contudo, sabendo que, o modelo de super eficiência necessita ser feito inspeção manual dos dados, utilizou a técnica jackstrap por não necessitar dessa inspeção, e assim notou-se que valores discrepantes à média estadual, gerando então a exclusão desses *outliers.*

Rani (2017) propôs em sua pesquisa uma abordagem para detectar anomalias composta por 3 estágios distintos: estágio de correlação, estágio de previsão e computação paralela. É considerado etapas separadas para tornar o modelo proposto escalável para lidar com grandes quantidades de dados de séries temporais. O sistema proposto é testado usando dados reais. Uma vez que os resultados experimentais verificam a eficácia pro-sistema proposto com alta taxa de detecção de *outliers* e baixa taxa de dados falsos de previsão de demanda de energia, ele pode ser usado para detectar anomalias em grandes conjuntos de dados no futuro.

Em contrapartida o estudo de Chai, Zhonghi e Haile (2015) traz a discussão acerca das técnicas de mineração de dados (DM), integradas para otimizar seu desempenho no campo da saúde e da ciência médica, considerando que o DM é uma ferramenta de análise de dados em larga escala moderna e avançada. A introdução do DM combinado para detecção de *outliers* baseada em técnicas; as técnicas combinadas de mineração em dados médicos, visto que técnicas combinadas de DM para cuidados de saúde podem incorporar com sucesso o processo de descoberta e detectar outliers com base em características de outlierness. Além de que, a detecção de valores discrepantes baseada em análise tradicional atravessa algumas deficiências no que se refere a definição dos resultados discrepantes apropriados e aceitáveis.

3 OBJETIVOS

3.1 Objetivo Geral

Avaliar a ocorrência de *outliers* no Sistema de Informações Sobre Orçamentos Públicos em Saúde (SIOPS) e seu impacto na qualidade da informação em indicadores referentes à capacidade fiscal e de investimentos em saúde no período de 2002 a 2021.

3.2 Objetivos Específicos:

- Analisar a completitude das informações;
- Descrever o quantitativo dos *outliers* pelos diferentes métodos de detecção;
- Avaliar a concordância mútua na detecção de *outliers;*
- Verificar o comportamento dos *outliers* nas cinco grandes regiões do país.

4 METODOLOGIA

Trata-se de um estudo descritivo, transversal, com abordagem quantitativa e inferencial para avaliar a ocorrência de *outliers* no SIOPS e seu impacto na qualidade da informação indicadores do Sistema de Informações sobre Orçamentos Públicos em Saúde (SIOPS), referentes à capacidade fiscal e de investimento em saúde, no período de 2002 a 2021, totalizando 20 anos. O local do estudo é o Brasil em suas municipalidades, ou seja, a partir da varredura exaustiva dos 5.569 municípios disponíveis no referido sistema.

Com fins de tornar clara a metodologia deste estudo, serão apresentados a seguir, os indicadores estudados, a preparação da base de dados, recursos computacionais, testes de detecção de *outliers*, medidas de concordância e aspectos éticos.

4.1 Indicadores estudados

Do total de 13 indicadores possíveis de serem avaliados no SIOPS, foram utilizados cinco. No quadro 1 a seguir, estão apresentados os indicadores referentes aos valores de receitas (1.1 e 1.2), despesas (2.1 e 2.2) e o relacionado a receitas e despesas (3.1).

Quadro 1 - Indicadores referentes aos valores de receitas, despesas e o relacionado a receitas e despesas

Indicador	**O que mede**
1.1	Capacidade de arrecadação do município: Representa a participação do total de Impostos diretamente arrecadados pelo município (IPTU, ITBI, ISS e IRRF) em relação à receita total do município.
1.2	Grau de dependência do município em relação às transferências de outras esferas de governo.

2.1	Gasto per capita em saúde: O gasto médio com Saúde, sob responsabilidade do município, por habitante: Representa a despesa total com Saúde (exceto inativos), inclusive aquela financiada por outras esferas de governo, por habitante.
2.2	Participação percentual da despesa com pessoal na despesa total com Saúde.
3.1	Percentual de despesas com saúde financiadas com recursos transferidos por outras esferas de governo.

Fonte: A autora, 2022.

4.2 Preparação da base de dados

Os dados foram inicialmente baixados livremente por meio do site do SIOPS, a nível municipal nas linhas, cujo link pode ser conferido a seguir: http://siops-asp.datasus.gov.br/cgi/siops/serhist/MUNICIPIO/indicadores.HTM .

Uma vez obtidos, foram particionados por indicador, criando-se assim cinco diferentes bancos de dados, ou seja, um para cada indicador. Em cada um destes conjuntos foi feita uma varredura em busca de municípios com valores ausentes, os quais foram prontamente retirados da análise, o que não consiste necessariamente de conjuntos íntegros, uma vez que valores nulos ou mesmo inconsistentes podem fazer parte de uma série temporal completa, sendo possível, em alguns casos, a detecção destas dimensões por meio dos testes de detecção aqui aplicados e posteriormente detalhados.

A integridade de dados refere-se a manter e garantir a precisão e consistências dos dados no decorrer de todo o ciclo da informação (DAMÁZIO, 2022), não sendo este, objetivo desta análise e sim atuar sobre um conjunto que apresente necessariamente completitude, mas sem garantia de outras dimensões da qualidade da informação.

4.3 Recursos Computacionais e testes de detecção

Os recursos computacionais utilizados para a análise dos dados foram o Programa R, por meio dos pacotes *outliers* para execução dos testes de Grubbs e Dixon, o *EnvStats* para execução do teste de Rosner, o *caret* para obtenção das medidas F1 e Kappa, além da planilha eletrônica Excel.

O teste de Grubbs (1969), se baseia em uma distribuição normal no qual a hipótese nula testada é que todas as discrepâncias da amostra provêm do mesmo grupo, ou seja, não são candidatas à *outliers*. Portanto, o teste de Grubbs é útil para avaliar se existem *outliers,* permitindo encontrar, estatisticamente, valores que devem ser descartados ou desconsiderados em uma análise. Além disso, esse método também é utilizado para confirmar o ajuste de todos os pontos considerados em uma curva analítica (CRISTO, 2021).

O teste de Dixon é assim chamado por ter sido desenvolvido por Dixon (1950), ele consiste em um teste que apresenta dois lados, em que a hipótese nula acredita que não há discrepância sugestiva a *outlier* no grupo das discrepâncias. Desse modo, o teste de Dixon é utilizado em testes de proporção menor em relação ao espaço amostral. É importante destacar que o teste apresenta algumas limitações para $n \leq 30$, sendo posteriormente estendido para $n \leq 40$ (OLIVEIRA *et al.,* 2014).

O teste de Rosner (1983) ou teste do desvio extremo da distribuição t de Student para *outliers,* tem o objetivo de detectar um ou mais valores discrepantes em um grupo de dados uni variado que segue próximo a uma distribuição normal. Esse teste requer apenas a existência de um limite superior para o número suspeito de *outliers* a ser especificado. Assim, o teste generalizado ESD especialmente realiza testes separados, ou seja, um teste para um *outlier,* outro teste para dois *outliers*, e assim por diante (OLIVEIRA *et al.,* 2014).

4.4 Medidas de concordância

Uma vez que serão adotados três testes em simultâneo no conjunto completo para detecção de municípios possivelmente influenciados por *outliers*, faz-se necessário uma avaliação da concordância destes no sentido de validar os achados e compreender melhor os municípios com *outliers* detectados, que podem ser

classificados em verdadeiros positivos (VP) e verdadeiros negativos (VN), além dos municípios com *outliers* não detectados, que podem ser classificados como falsos positivos (FP) e falsos negativos (FN) através da obtenção de matrizes de confusão com o formato apresentado na Tabela 2 .

A partir destas matrizes, pode-se obter medidas de concordância entre os testes dois a dois, como as medidas F1 e de Kappa. Esta primeira é uma medida harmônica entre a precisão e a revocação (*recall*), ver Figura 4, e é dada por:

$$F1 = 2 * \frac{Precisão * Recall}{Precisão + Recall}$$

Valores de F1 abaixo de 0,5 são considerados de baixa concordância entre os métodos, principalmente na detecção de VP (ALLWRIGHT, 2022). Já a medida Kappa (LANDIS; KOCH, 1977), trata-se de um método estatístico para avaliar o nível de concordância ou reprodutibilidade entre dois conjuntos de dados. Kappa é dado por uma razão que envolve a diferença da probabilidade de aceitação relativa (p_o), ou seja, o número de concordâncias entre os dois especialistas dividido pelo número total de avaliações e a probabilidade de concordância randômica (p_e), ou seja, a probabilidade conjunta de cada teste detectar ou não outliers no conjunto total de municípios e é dada conforme a seguir:

$$k = \frac{p_o - p_e}{1 - p_e}$$

O mesmo autor classifica a medida de magnitude de k de acordo com a Tabela 1, da seguinte forma:

Tabela 1 - Classificação da concordância a partir de Kappa

Kappa	Concordância
<0	Não existe
0 - 0,2	Mínima
0,21 - 0,4	Razoável
0,41 - 0,6	Moderada
0,61 - 0,8	Substancial
0,8 - 1	Perfeita

Fonte: (LANDIS; KOCH, 1977)

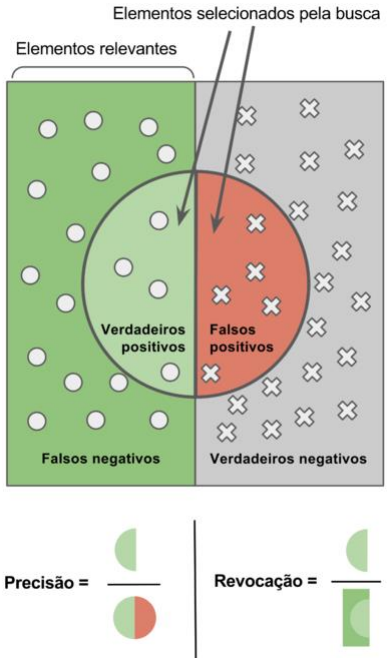
Ambas medidas foram utilizadas em outro recente trabalho com o mesmo propósito de avaliar a concordância de métodos quanto a detecção de *outliers* em séries temporais (ALIMOHAMMADI; NANCY CHEN, 2022).

Tabela 2 - Matriz de confusão genérica

Teste A	Teste B	
	Outlier não detectado	Outlier detectado
Outlier não detectado	*VN*	*FN*
Outlier detectado	*FP*	*VP*

Fonte: A autora, 2022.

Figura 4 - Ilustração do cálculo da precisão e do *recall*



https://en.wikipedia.org/wiki/F-score

4.5 Aspectos Éticos

O referido trabalho não consiste em pesquisa clínica, epidemiológica ou experimental com seres humanos e/ou animais, sendo voltado a análise de informações públicas e agregadas a nível domiciliar do sistema de informação SIOPS no período de 2002 a 2021. Conforme Resolução 510/2016 do CONEP em seu Artigo I, Parágrafo Único, itens II e V, não serão registradas nem avaliadas pelo sistema CEP/CONEP:

II- pesquisa que utilize informações de acesso público, nos termos da Lei nº 12.527, de 18 de novembro de 2011;

V- pesquisa com bancos de dados, cujas informações são agregadas, sem possibilidades de identificação individual;

5 RESULTADOS E DISCUSSÃO

Esta seção traz os principais resultados organizados em duas subseções:

- Análise da completitude das informações, uma vez que é necessário obter o conjunto de municípios com preenchimento em todos os anos;
- Análise descritiva da quantidade e concordância de *outliers* detectados pelos três testes via diagramas de Venn e medidas Kappa e F1.

5.1 Análise da completitude das informações

Para cada indicador, o universo de estudo consiste do conjunto de municípios com informações completas nos 20 anos da série. Os percentuais de cada indicador seguem no quadro 2 abaixo, o que mostrou uma alta completitude para o SIOPS neste período.

Quadro 2 - Completitude dos indicadores estudados

Indicador	***Municípios com dados completos***	***%***
1.1	5230	93,9
1.2	5226	93,8
2.1	5226	93,8
2.2	5225	93,8
3.1	5209	93,5

Fonte: A autora, 2022.

5.2 Análise da nulidade das informações no conjunto completo

Conjuntos de dados completos podem apresentar tanto valores inconsistentes quanto à nulidade das informações. Embora não seja objetivo deste trabalho, verificar o tamanho da nulidade pode ajudar a entender melhor o comportamento da detecção

de *outliers.* Verificou-se que para o caso do SIOPS, nos conjuntos completos, esta dimensão não gerou tanto impacto no período estudado, em geral, correspondendo a valores menores que 1% do universo completo (Quadro 3).

Quadro 3 - Nulidade dos indicadores estudados

Indicador	***Municípios com nulidade / universo***	***%***
1.1	34 / 5230	< 1%
1.2	0 / 5226	< 1%
2.1	1 / 5226	< 1%
2.2	72 / 5225	1,37%
3.1	27 / 5209	< 1%

Fonte: A autora, 2022.

5.3 Análise descritiva e de concordância da quantidade de *outliers* detectados pelos diferentes testes

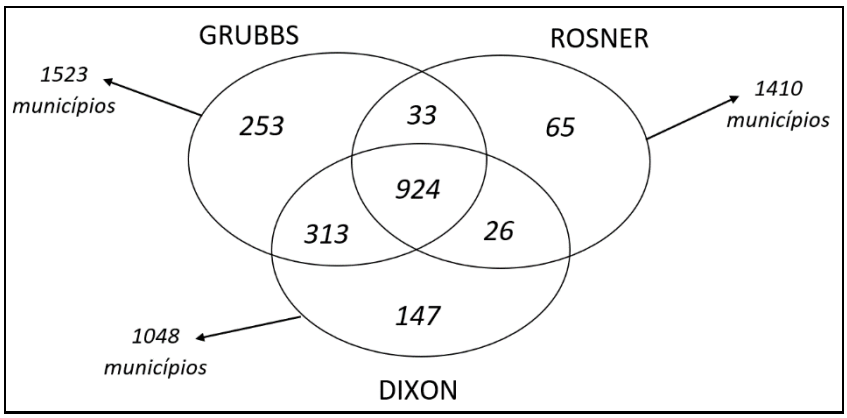
Para a capacidade de arrecadação municipal (CAM), indicador 1.1, 1761 municípios, 33,6% do total, foram identificados por pelo menos um método com *outliers,* da seguinte forma: por Grubbs, 1523 (29,1%), por Rosner, 1410 (20%) e por Dixon, 1048 (26,9%), o que mostra que no mínimo, pelo menos 1/5 das informações podem estar sendo influenciadas por possíveis valores discrepantes, uma vez que, um *outlier* não significa necessariamente uma informação errada. Uma observação mais detalhada do comportamento destes municípios entre os três métodos aplicados pode ser melhor visualizada na Figura 5, a seguir.

Para a CAM, os seguintes resultados podem ser assim resumidos:

- Os três métodos concordam mutuamente em 924 municípios, 17,6% do total e 52,4% daqueles detectados por pelo menos um método;
- Grubbs e Rosner detectaram conjuntamente 1696 casos, com uma concordância mútua em 1237 casos (56%);

- Já Grubbs e Dixon detectaram conjuntamente 1614, casos com uma concordância mútua em 957 casos (76%);
- Rosner e Dixon detectaram conjuntamente 1508 casos, com uma concordância mútua em 950 casos (63%);
- Tanto F1 quanto Kappa apontaram para uma concordância substancial nas três combinações dois a dois dos métodos.

Figura 5 - Diagrama de Venn da quantidade de municípios com *outliers* detectados por diferentes testes para capacidade de arrecadação municipal



Fonte: A autora, 2022.

Tabela 3 - Análise de concordância entre os diferentes testes para a capacidade de arrecadação municipal

Testes	F1	Kappa	Matriz de confusão	
GRUBBS, ROSNER	0,84	0,78	**3534**	173
			286	**1237**
GRUBBS, DIXON	0,74	0,66	**3616**	91
			566	**957**
ROSNER, DIXON	0,77	0,7	**3722**	98
			460	**950**

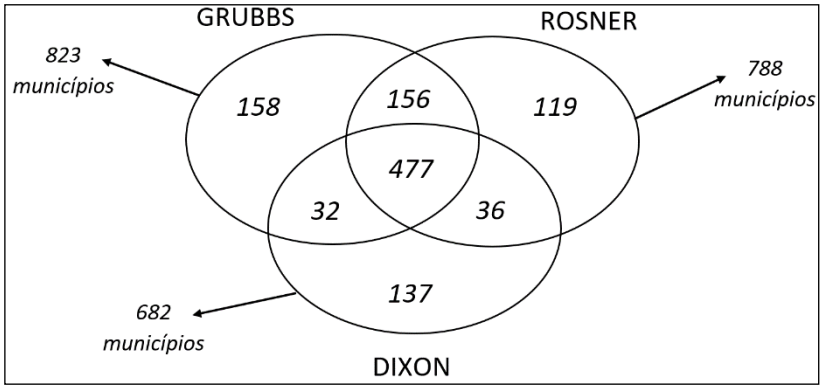
Fonte: A autora, 2022.

Para o grau de dependência municipal (GDM), indicador 1.2, 1115 municípios, 21,3% do total, foram identificados por pelo menos um método com *outliers,* da seguinte forma: por Grubbs, 823 (15,7%), por Rosner, 788 (15%) e por Dixon, 682 (13%), o que mostra que no mínimo, pelo menos 1/8 das informações podem estar sendo influenciadas por possíveis valores discrepantes. Uma observação mais detalhada do comportamento destes municípios entre os três métodos aplicados pode ser melhor visualizada na Figura 6, a seguir.

Para o GDM, os seguintes resultados podem ser assim resumidos:

- Os três métodos concordam mutuamente em 477 municípios, 9,1% do total e 42,8% daqueles detectados por pelo menos um método;
- Grubbs e Rosner detectaram conjuntamente 978 casos, com uma concordância mútua em 633 casos (64,7%);
- já Grubbs e Dixon detectaram conjuntamente 996, casos com uma concordância mútua em 509 casos (51,1%);
- Rosner e Dixon detectaram conjuntamente 957 casos, com uma concordância mútua em 513 casos (53,6%);
- Tanto F1 quanto Kappa apontaram para uma concordância substancial nas três combinações dois a dois dos métodos.

Figura 6 - Diagrama de Venn da quantidade de municípios com *outliers* detectados por diferentes testes para o grau de dependência municipal



Fonte: A autora, 2022.

Tabela 4 - Análise de concordância entre os diferentes testes para o grau dependência municipal

Testes	F1	Kappa	Matriz de confusão	
GRUBBS, ROSNER	0,78	0,74	**4248**	155
			190	**633**
GRUBBS, DIXON	0,67	0,62	**4230**	173
			314	**509**
ROSNER, DIXON	0,69	0,64	**4269**	169
			275	**513**

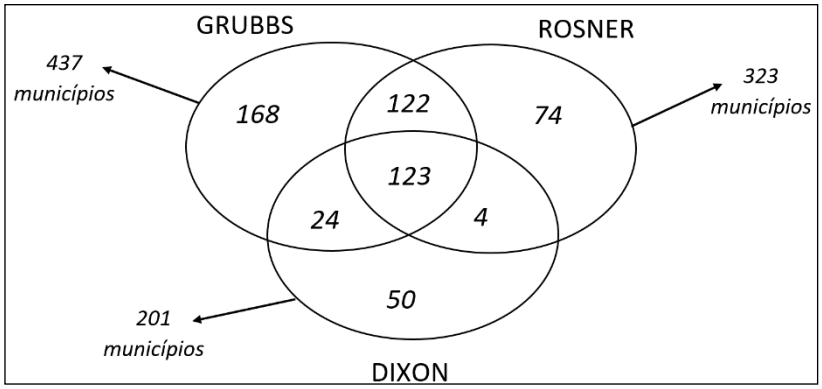
Fonte: A autora, 2022.

Para o gasto per capita com saúde (GPS), indicador 2.1, 565 municípios, 10% do total, foram identificados por pelo menos um método com *outliers,* da seguinte forma: por Grubbs, 437 (8,3%), por Rosner, 323 (6,2%) e por Dixon, 201 (3,8%), o que mostra que no mínimo, pelo menos 1/25 das informações podem estar sendo influenciadas por possíveis valores discrepantes. Uma observação mais detalhada do comportamento destes municípios entre os três métodos aplicados pode ser melhor visualizada na Figura 7, a seguir.

Para o GPS, os seguintes resultados podem ser assim resumidos:

- Os três métodos concordam mutuamente em 123 municípios, 2,35% do total e 21,8% daqueles detectados por pelo menos um método;
- Grubbs e Rosner detectaram conjuntamente 515 casos, com uma concordância mútua em 245 casos (47,5%);
- já Grubbs e Dixon detectaram conjuntamente 491, casos com uma concordância mútua em 147 casos (30%);
- Rosner e Dixon detectaram conjuntamente 397 casos, com uma concordância mútua em 127 casos (32%);
- Para Grubbs e Rosner, tanto F1 quanto Kappa apontaram para uma concordância substancial, porém, para Grubbs e Dixon e para Dixon e Rosner, F1 foi menor que 0,5 e consequentemente indicou uma baixa correspondência, além disso, Kappa apontou para uma concordância apenas moderada.

Figura 7 - Diagrama de Venn da quantidade de municípios com *outliers* detectados por diferentes testes para o gasto per capita com saúde



Fonte: A autora, 2022.

Tabela 5 - Análise de concordância entre os diferentes testes para o gasto per capita com saúde

Testes	F1	Kappa	Matriz de confusão	
GRUBBS, ROSNER	0,64	0,61	**4711**	78
			192	**245**
GRUBBS, DIXON	0,46	0,43	**4735**	54
			290	**147**
ROSNER, DIXON	0,48	0,46	**4829**	74
			196	**127**

Fonte: A autora, 2022.

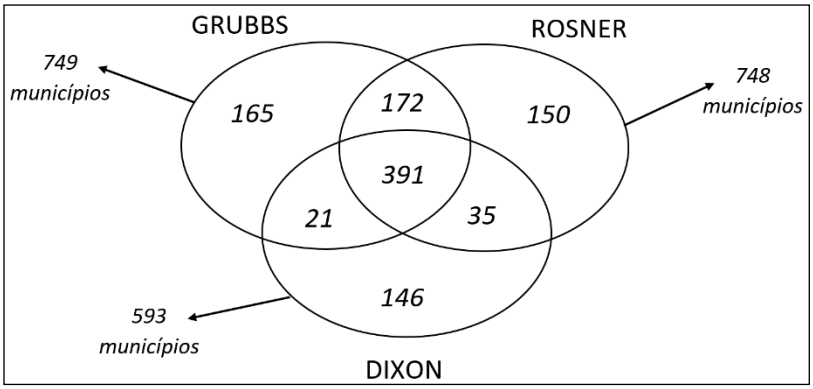
Para o percentual da despesa total com saúde em relação ao gasto total com saúde (DEP), indicador 2.2, 1080 municípios, 20,6% do total, foram identificados por pelo menos um método com *outliers,* da seguinte forma: por Grubbs, 749 (14,3%), por Rosner, 748 (14,3%) e por Dixon, 593 (11,3%), o que mostra que no mínimo, pelo menos 1/9 das informações podem estar sendo influenciadas por possíveis valores discrepantes. Uma observação mais detalhada do comportamento destes municípios entre os três métodos aplicados pode ser mais bem visualizada na Figura 8, a seguir.

Para o DEP, os seguintes resultados podem ser assim resumidos

- Os três métodos concordam mutuamente em 391 municípios, 7,5% do total e 36,2% daqueles detectados por pelo menos um método;

- Grubbs e Rosner detectaram conjuntamente 934 casos, com uma concordância mútua em 563 casos (60,2%);
- já Grubbs e Dixon detectaram conjuntamente 930, casos com uma concordância mútua em 412 casos (44,3%);
- Rosner e Dixon detectaram conjuntamente 915 casos, com uma concordância mútua em 426 casos (46,5%);
- Para Grubbs e Rosner, tanto F1 quanto Kappa apontaram para uma boa concordância entre os métodos. Para as demais combinações F1 apontou uma concordância fraca e Kappa para uma concordância apenas moderada.

Figura 8 - Diagrama de Venn da quantidade de municípios com *outliers* detectados por diferentes testes para o percentual da despesa total com saúde em relação ao gasto total com saúde



Fonte: A autora, 2022.

Tabela 6 - Análise de concordância entre os diferentes testes para o percentual da despesa total com saúde em relação ao gasto total com saúde

Testes	F1	Kappa	Matriz de confusão	
GRUBBS, ROSNER	0,75	0,71	**4291**	185
			186	**563**
GRUBBS, DIXON	0,61	0,56	**4295**	181
			337	**412**
ROSNER, DIXON	0,63	0,58	**4310**	167
			322	**426**

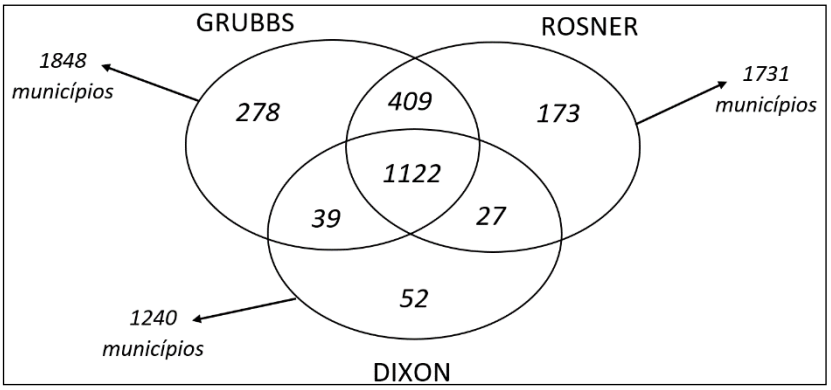
Fonte: A autora, 2022.

Para o percentual de despesas com saúde financiadas com recursos transferidos por outras esferas de governo (DSO), indicador 3.1, 2100 municípios, 40,3% do total, foram identificados por pelo menos um método com *outliers,* da seguinte forma: por Grubbs, 1848 (35,3%), por Rosner, 1731 (33,2%) e por Dixon, 1240 (23,8%), o que mostra que no mínimo, pelo menos 1/5 das informações podem estar sendo influenciadas por possíveis valores discrepantes. Uma observação mais detalhada do comportamento destes municípios entre os três métodos aplicados pode ser melhor visualizada na Figura 9, a seguir.

Para o DSO, os seguintes resultados podem ser assim resumidos

- Os três métodos concordam mutuamente em 1122 municípios, 21,5% do total e 53,4% daqueles detectados por pelo menos um método;
- Grubbs e Rosner detectaram conjuntamente 2048 casos, com uma concordância mútua em 1531 casos (74,7%);
- já Grubbs e Dixon detectaram conjuntamente 1927, casos com uma concordância mútua em 1161 casos (60,2%);
- Rosner e Dixon detectaram conjuntamente 1822 casos, com uma concordância mútua em 1149 casos (63%);
- Tanto F1 quanto Kappa apontaram para uma concordância substancial nas três combinações dois a dois dos métodos.

Figura 9 - Diagrama de Venn da quantidade de municípios com *outliers* detectados por diferentes testes para o percentual de despesas com saúde financiadas com recursos transferidos por outras esferas de governo



Fonte: A autora, 2022.

Tabela 7 - Análise de concordância entre os diferentes testes para o percentual de despesas com saúde financiadas com recursos transferidos por outras esferas de governo

Testes	F1	Kappa	Matriz de confusão	
GRUBBS, ROSNER	0,85	0,78	**3161**	200
			317	**1531**
GRUBBS, DIXON	0,75	0,65	**3282**	79
			687	**1161**
ROSNER, DIXON	0,77	0,68	**3387**	91
			582	**1149**

Fonte: A autora, 2022.

5.3.1 Análise conjunta dos indicadores - Municípios com mútua concordância

Analisando conjuntamente o desempenho de busca por municípios com potenciais *outlier* é possível também entender o perfil de cada indicador no tocante a este quesito. A concordância mútua em relação aos municípios, ou os municípios com mútua concordância (MMC), podem ser utilizados como indicador relevante para assegurar que aqueles municípios, de fato, apresentam esta característica, uma vez que passaram pela observação de três “juízes”, que neste caso, os métodos de detecção.

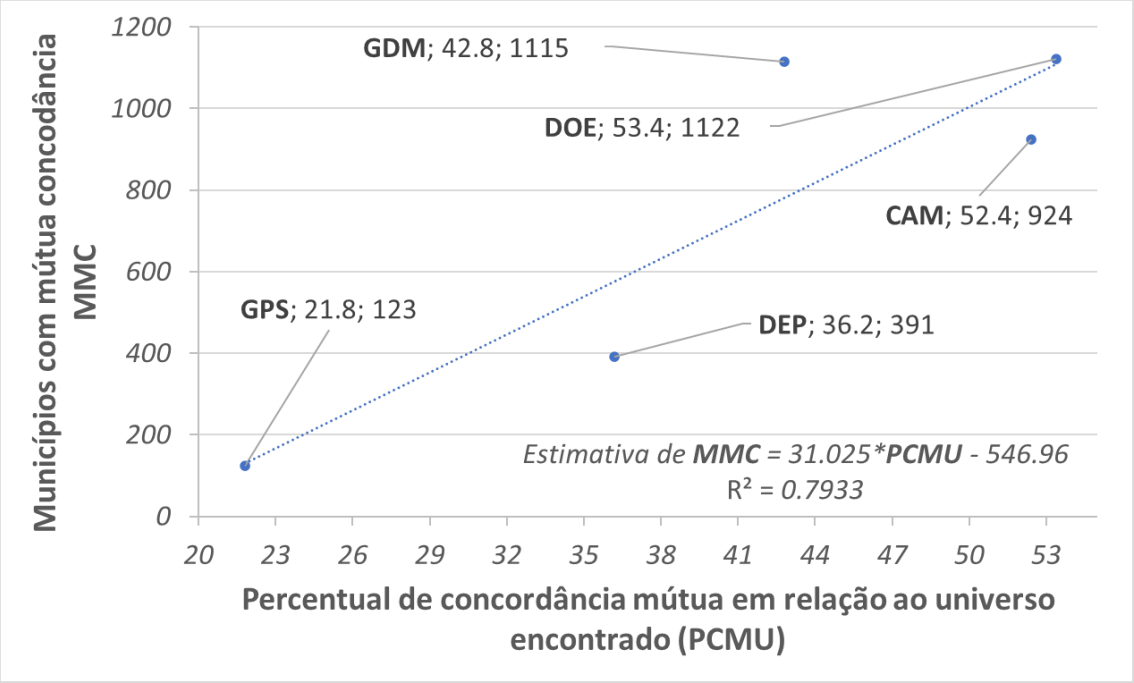
Quando MMC é confrontado com seu percentual em relação a todos os municípios apontados com a presença de *outliers* pelos três métodos, ou seja, com o percentual de concordância mútua em relação ao universo de municípios com *outliers* encontrado (PCMU), verifica-se que quanto maior MMC, em geral será maior PCMU. Isso, entre outras coisas, aponta para o fato de que indicadores como GPS, que no universo dos 5230, obteve MMC de 123 e PCMU de apenas 21,8%, apresentam não somente uma menor quantidade de municípios com *outliers,* mas também uma maior dificuldade de concordância entre os métodos. Esta questão também foi fortemente apontada pelas medidas F1 e de Kappa dois a dois para este indicador que apontou para concordâncias fracas ou moderadas em duas das três combinações (Gráficos 1 e 2).

Por outro lado, indicadores que obtiveram um MMC de aproximadamente 1000, tendem a concordar em mais de 40% de suas indicações, como foi o caso de GDM,

DOE e CAM. Estes mesmos indicadores avaliados dois a dois apresentaram tanto F1 quanto Kappa satisfatórios. No caso de CAM e DOE o PCMU se deu para mais de 50% do universo. Já o DEP, foi o mais intermediário dos indicadores, uma vez que se observou um MMC de 391 e um PCMU de 36,2%. Esta questão também foi reforçada anteriormente pelas medidas F1 e Kappa dois a dois entre os métodos. MMC e PCMU também apresentaram forte correlação positiva (0,89) e alta explicabilidade, aproximadamente 79% (Gráficos 1 e 2).

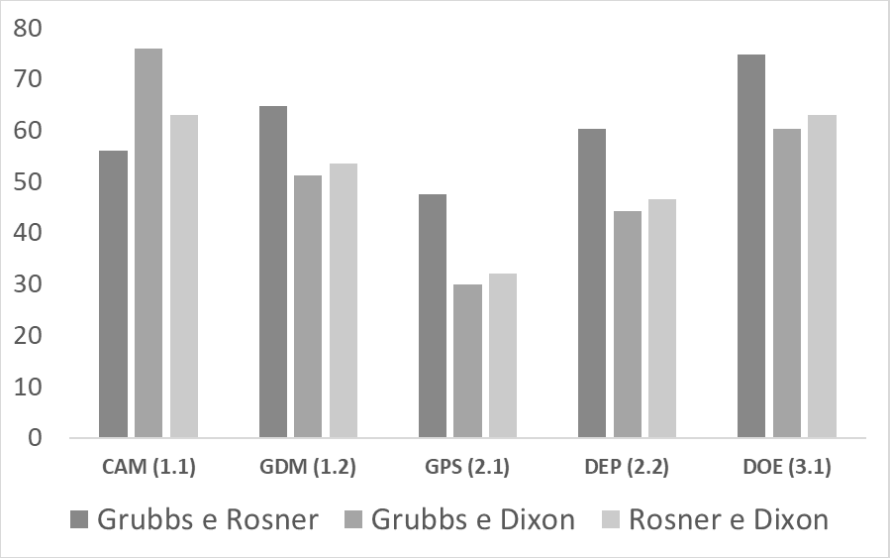
Em relação a nulidade, observou-se que a mesma não chega ter grande participação no SIOPS, porém, mesmo assim, uma parte das poucas situações de nulidade presentes foram identificadas como *outliers* por pelo menos um método, a saber, para CAM (em 18 das 34 situações, 53%), para GDM (não existem valores nulos), para GPS (em 1 de 1 situação, 100%), para DEP (em 47 das 72 situações, 65,2%) e para DOE (em 20 das 27 situações, 74%).

Gráfico 1 - MMC x PCMU



Fonte: A autora, 2022.

Gráfico 2 - Percentual de concordância dois a dois entre os métodos nos indicadores avaliados



Fonte: A autora, 2022.

5.3.2 Análise conjunta dos indicadores - Análise regional

Avaliando a ocorrência de *outliers* nas cinco grandes regiões do país, verificou-se associação com todos os indicadores. Porém, foi fraca para CAM, DEP e DOE e forte para GDM e GPS.

Tabela 8 - Análise conjunta dos cinco indicadores

Indicador	Qui-quadrado	p-valor	V de Cramer	Associação
CAM (1.1)	14.149	<0.001	0.052	Fraca
GDM (1.2)	150.2	<0.001	0.169	Forte
GPS (2.1)	140.63	<0.001	0.164	Forte
DEP (2.2)	19.921	<0.001	0.061	Fraca
DOE (3.1)	29.208	<0.001	0.074	Fraca

Fonte: A autora, 2022.

Para a capacidade de arrecadação municipal (CAM), indicador 1.1, o percentual de ocorrência de *outliers* na região Norte foi de 36,3%, região Nordeste 35,8%, região Sudeste 34,3%, região Sul 31% e na região Centro-oeste 28,4%.

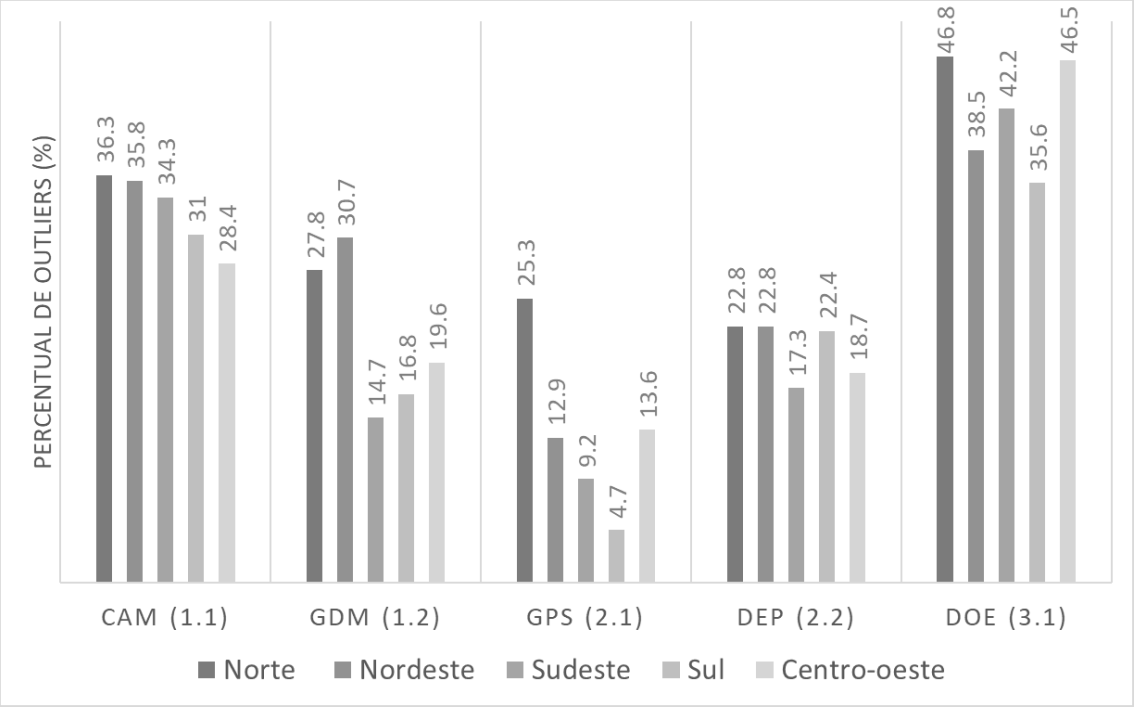
Para o grau de dependência municipal (GDM), indicador 1.2, o percentual de ocorrência de *outliers* na região Norte foi de 27,8%, região Nordeste 30,7%, região Sudeste 14,7%, região Sul 16,8% e na região Centro-oeste 19,6%.

Para o gasto per capita com saúde (GPS), indicador 2.1, o percentual de ocorrência de *outliers* na região Norte foi de 25,3%, região Nordeste 12,9%, região Sudeste 9,2%, região Sul 4,7% e na região Centro-oeste 13,6%.

Para a despesa total com saúde em relação ao gasto total com saúde (DEP), indicador 2.2, o percentual de ocorrência de *outliers* na região Norte foi de 22,8%, região Nordeste 22,8%, região Sudeste 17,3%, região Sul 22,4% e na região Centro-oeste 18,7%.

Para as despesas com saúde financiadas com recursos transferidos por outras esferas de governo (DOE), indicador 3.1, o percentual de ocorrência de *outliers* na região Norte foi de 46,8%, região Nordeste 38,5%, região Sudeste 42,2%, região Sul 35,6% e na região Centro-oeste 46,5% (Gráfico 3).

Gráfico 3 - Percentual de ocorrência de *outliers* nas regiões do país Norte, Nordeste, Sudeste, Sul e Centro-oeste



Fonte: A autora, 2022.

6 CONCLUSÃO

Neste estudo identificou-se que o SIOPS apresentou uma alta completitude no período de 2002 a 2021. Dos cincos indicadores estudados o indicador 1.1, referente à capacidade de arrecadação municipal (CAM) apresentou completitude excelente. Em relação a nulidade observou-se que ela existe, mas é baixa, porém serviu para verificar se essa nulidade em algum momento foi detectada como um *outlier* e por fim foi constatado. Contudo, o tamanho da nulidade, pode ajudar a compreender o comportamento dos possíveis *outliers.* Para o SIOPS nos conjuntos completos, esta dimensão não gerou tanto impacto neste período. Entre os três testes estatísticos para detectar *outliers,* o teste de Grubbs foi o que mais detectou, seguido por Rosner e por fim Dixon.

O indicador 1.1, capacidade de arrecadação municipal, o teste de Grubbs encontrou 1523 (29,1%) municípios que tem algum *outlier*, por Rosner 1048 (26,9%) e por Dixon 1410 (20%). Vale ressaltar que um *outlier* não implica exatamente em uma informação incorreta. O indicador 1.2, grau de dependência municipal, o teste de Grubbs encontrou 823 (15,7%), municípios que possuem algum *outlier*, por Rosner 788 (15%) e por Dixon 682 (13%). É evidente que essas informações sofrem influências por possíveis valores discrepantes.

O indicador 2.1, gasto per capita com saúde, o teste de Grubbs encontrou 437 (8,3%), municípios que dispõem algum *outlier,* por Rosner, 323 (6,2%) e por Dixon, 201 (3,8%). Essas informações são afetadas por possíveis valores discrepantes. O indicador 2.2, percentual da despesa total com saúde em relação ao gasto total com saúde, o teste de Grubbs encontrou 749 (14,3%), por Rosner, 748 (14,3%) e por Dixon, 593 (11,3%). Essas informações podem estar sendo influenciadas por possíveis valores discrepantes.

Por fim, o indicador 3.1, percentual de despesas com saúde financiadas com recursos transferidos por outras esferas de governo, o teste de Grubbs encontrou 1848 (35,3%), por Rosner, 1731 (33,2%) e por Dixon, 1240 (23,8%). Podendo estar sofrendo influência por possíveis valores discrepantes.

No que diz respeito ao desempenho de busca por municípios com potenciais *outliers* é possível entender o perfil de cada indicador. Nesta análise, verifica-se que quanto maior for o MMC, em geral será maior o PCMU.

Com relação as cincos grandes regiões do país, observou-se que o menor percentual de *outliers* para o indicador 1.1, capacidade de arrecadação municipal (CAM) é na região Centro-oeste e o maior percentual na região Norte. O indicador 1.2, grau de dependência municipal (GDM) o menor percentual de *outliers* é na região Sudeste e o maior percentual na região Nordeste.

O indicador 2.1, gasto per capita com saúde (GPS) o menor percentual de *outliers* é na região Sul e o maior na região Norte. O indicador 2.2, despesa total com saúde em relação ao gasto total com saúde (DPE) o menor percentual de *outliers* é na região Sudeste e os maiores são nas regiões Norte e Nordeste.

O indicador 3.1, despesas com saúde financiadas com recursos transferidos por outras esferas de governo (DOE) o menor percentual de *outliers* é na região Sul e o maior na região Norte.

Os métodos aplicados para detecção de *outliers* se mostraram eficientes na observação do comportamento dos municípios. Outro ponto determinante que vale destacar é a capacidade fiscal, considerando sua magnitude do investimento em saúde, visto que, a análise de indicadores de orçamento e gasto favorece a discussão da disponibilidade e composição das receitas e despesas, tanto do ponto de vista da política fiscal, quanto do financiamento da política de saúde.

Tendo em vista o alcance destes resultados, é clara a importância do estudo sobre o assunto, pois a estes podem servir como subsídio para uma nova discussão. Conclui-se que, este trabalho acerca dos *outliers* em saúde é de extrema relevância visto que, a presença de valores discrepantes também tem influência direta nas dimensões da qualidade da informação: a completitude e a confiabilidade.

REFERÊNCIAS

AGGARWAL, Charu C. An Introduction to Outlier Analysis. *In*: AGGARWAL, Charu C. (Org.). **Outlier Analysis**. Cham: Springer International Publishing, 2017, p. 1–34. Disponível em: https://doi.org/10.1007/978-3-319-47578-3_1. Acesso em: 5 out. 2022.

AKOGLU, Haldun. User's guide to correlation coefficients. **Turkish journal of emergency medicine**, Mumbai, v. 18, n. 3, p. 91–93, 2018.

ALIMOHAMMADI, Hamzeh; NANCY CHEN, Shengnan. Performance evaluation of outlier detection techniques in production timeseries: A systematic review and meta-analysis. **Expert Systems with Applications**, United States, v. 191, p. 116371, 2022.

ALLWRIGHT, Stephen. **How to interpret F1 score (simply explained)**. [*S. l.: s. n.*], 2022. Disponível em: https://stephenallwright.com/interpret-f1-score/. Acesso em: 4 out. 2022.

BARBOSA, Josino José. **Identificação de outliers multivariados:** uma aplicação em dados de saúde. 2017. 51 f. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Viçosa, Viçosa, 2017.

BRASIL. Ministério da Saúde. **Sistema de Informações sobre Orçamentos Públicos em Saúde - SIOPS.** Brasília-DF: Ministério da Saúde, 18 maio 2021. Disponível em: https://www.gov.br/saude/pt-br/acesso-a-informacao/siops. Acesso em: 09 maio 2022.

BRASIL. Ministério da Saúde. Conselho Nacional de Saúde. Resolução nº 510, de 7 de abril de 2016. Trata sobre as diretrizes e normas regulamentadoras de pesquisa em ciências humanas e sociais. **Diário Oficial da União**, Brasília, DF, n. 98, seção 1, p. 44-46, 24 maio 2016. Disponível em: https://conselho.saude.gov.br/resolucoes/2016/Reso510.pdf. Acesso em 10 maio 2022.

CAMACHO, José *et al*. PCA-based multivariate statistical network monitoring for anomaly detection. **Computers & Security**, United States, v. 59, p. 118–137, 2016.

CARVALHO, Gregory dos Passos. Desigualdades regionais e o papel dos recursos federais no SUS: fatores políticos condicionam a alocação de recursos? **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 26, supl. 2, p. 3409–3421, 2021.

CRISTO, Sabrina Lourdes Pereira de. **Análise da detecção e da influência de outliers na avaliação da acurácia posicional de produtos cartográficos**. 2021. 82 f. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Viçosa, Viçosa. 2021. Disponível em: https://locus.ufv.br//handle/123456789/28638. Acesso em: 5 out. 2022.

CRUZ, Felipe Ponciano da da; MOTTA, Ronaldo Seroa da; MARINHO, Alexandre. Eficiência Técnica dos Serviços de água e esgoto no Brasil por tipo de atuação e gestão. **Econômica**, Niterói, v. 21, n. 1, 2019. Disponível em: https://periodicos.uff.br/revistaeconomica/article/view/35118. Acesso em: 6 out. 2022.

DAMÁZIO, Simara Lopes Cruz. **Financiamento da saúde no Brasil:** uma análise da qualidade dos dados do Finanças do Brasil (Finbra) e Sistema de Informações sobre Orçamentos Públicos em Saúde (Siops) no período de 2002 a 2020. 2022. 187 p. Tese (Doutorado) - Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, 2022.

FELICIANO, Marciana. **Federalismo fiscal e financiamento da política de saúde:** um estudo dos municípios do Brasil**.** 2019. P. 129 f. Tese (Doutorado em Saúde Pública) - Fundação Oswaldo Cruz, Instituto Aggeu Magalhães, Recife, 2019.

FREITAS, Igor Wescley Silva de. **Um estudo comparativo de técnicas de detecção de outliers no contexto de classificação de dados**. 2019. 97 f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal Rural do Semi-árido, Mossoró-RN, 2019. Disponível em: https://repositorio.ufersa.edu.br//handle/prefix/1093. Acesso em: 4 out. 2022.

GONTIJO, Tiago Silveira; REIS, Ilka Afonso. Os determinantes da eficiência na Atenção Primária à Saúde dos municípios paulistas: um modelo georreferenciado. **Physis:** Revista de Saúde Coletiva, Rio de Janeiro, v. 31, n. 1, p. e310132, 2021.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, Estados Unidos, v. 33, n. 1, p. 159–174, 1977.

NEPOMUCENA, Tania Miranda; CIRILLO, Marcelo Angelo. Identificação de outliers por alocação das observações através do modelo Poisson aplicado ao número de casos de Aids diagnosticados no Brasil. **Semina cienc. biol. saude**, Londrina, v. 31, n. 1, p. 29–35, 2012.

OLIVEIRA, C. D.; CAROLI, A. A.; AMARAL, A. de S.; VILCA, O. L. Detecção de Fraudes, Anomalias e Erros em Análise de Dados Contábeis: Um Estudo com Base em Outliers. **Redeca, Revista Eletrônica do Departamento de Ciências Contábeis & Departamento de Atuária e Métodos Quantitativos**, São Paulo, v. 1, n. 1, p. 102–127, 2015. DOI: 10.23925/2446-9513.2014v1i1p102-127. Disponível em: https://revistas.pucsp.br/index.php/redeca/article/view/23377. Acesso em: 5 set. 2022.

RODRIGUES, Alexandre de Cássio; GONTIJO, Tiago Silveira; GONÇALVES, Carlos Alberto. Eficiência do gasto público em atenção primária em saúde nos municípios do Rio de Janeiro, Brasil: escores robustos e seus determinantes. **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 26, p. 3567–3579, 2021.

SILVA, Philippe Barbosa *et al*. Identificação de outliers em dados de acidentes de trânsito no Brasil: análise de cluster versus métodos estatísticos. **Revista de Engenharia e Tecnologia**, Ponta Grossa, v. 11, n. 1, p. páginas 103-114, 2019.

SOUZA, Thiago I. A.; MAGALHAES, Deborah; AQUINO, Andre L. L.; GOMES, Danielo G. Um Método para Detecção e Diagnóstico de Outliers em Dados Urbanos via Análise Multidimensional. *In*: SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES E SISTEMAS DISTRIBUÍDOS (SBRC), 36., 2018, Campos do Jordão. **Anais** [...]. Porto Alegre: Sociedade Brasileira de Computação, 2018. p. 295-308. ISSN 2177-9384. DOI: https://doi.org/10.5753/sbrc.2018.2423.

TEIXEIRA, Hugo Vocurca; TEIXEIRA, Marcelo Gouvêa. Financiamento da saúde pública no Brasil: a experiência do Siops. **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 8, p. 379–391, 2003.

VENTURA, Thiago *et al*. **Detecção de Outliers em Dados Micrometeorológicos**. [*s. l.: s. n*.], 2013.