

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
DEPARTAMENTO DE ECONOMIA
BACHARELADO EM CIÊNCIAS ECONÔMICAS

SAJAN PRYA CORREIA CIRILO

**PLANO MUNICIPAL DE SANEAMENTO BÁSICO E SEU IMPACTO NO
ÍNDICE DE ATENDIMENTO TOTAL DE ÁGUA, NO ÍNDICE DE
ATENDIMENTO TOTAL DE ESGOTO E NA TAXA DE COBERTURA
REGULAR DO SERVIÇO DE COLETA DE RESÍDUOS DOMICILIARES**

Recife

2023

SAJAN PRYA CORREIA CIRILO

**PLANO MUNICIPAL DE SANEAMENTO BÁSICO E SEU IMPACTO NO
ÍNDICE DE ATENDIMENTO TOTAL DE ÁGUA, NO ÍNDICE DE
ATENDIMENTO TOTAL DE ESGOTO E NA TAXA DE COBERTURA
REGULAR DO SERVIÇO DE COLETA DE RESÍDUOS DOMICILIARES**

Trabalho de conclusão de curso apresentado ao curso de Ciências Econômicas da Universidade Federal Rural de Pernambuco (UFRPE) como pré-requisito parcial para a obtenção do grau de Bacharel em Ciências Econômicas sob a orientação da Prof^a. Dr^a. Chiara Natércia França Araújo.

Recife

2023

Dados Internacionais de
Catalogação na Publicação
Universidade Federal Rural de
Pernambuco Sistema Integrado
de Bibliotecas
Gerada automaticamente, mediante os dados
fornecidos pelo(a) autor(a)

C578p Cirilo, Sajan Prya Correia
Plano Municipal de Saneamento Básico e seu impacto no Índice de Atendimento
Total de Água, Índice de Atendimento Total de Esgoto e na Taxa de Cobertura Regular
do serviço de Coleta de Resíduos Domiciliares / Sajan Prya Correia Cirilo. - 2023.
56 f.

Orientador: Chiara Natércia Franca Araujo.
Inclui referências.

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal
Rural de Pernambuco, Bacharelado em Ciências Econômicas, Recife, 2023.

1. Clusterização. 2. Propensity Score Matching. 3. Regressão linear. 4.
Saneamento Básico. 5. Plano Municipal de Saneamento Básico. I. ARAUJO, CHIARA
NATERCIA FRANCA, orient. II. Título

CDD 330

SAJAN PRYA CORREIA CIRILO

PLANO MUNICIPAL DE SANEAMENTO BÁSICO E SEU IMPACTO NO ÍNDICE DE ATENDIMENTO TOTAL DE ÁGUA, NO ÍNDICE DE ATENDIMENTO TOTAL DE ESGOTO E NA TAXA DE COBERTURA REGULAR DO SERVIÇO DE COLETA DE RESÍDUOS DOMICILIARES

TCC apresentado ao Curso de Graduação em Economia da Universidade Federal Rural de Pernambuco, como requisito parcial para a obtenção do título de Bacharel em Ciências Econômicas.

Aprovado em: 14/09/2023.

BANCA EXAMINADORA

Prof^a. Dr^a. CHIARA NATÉRCIA FRANÇA ARAÚJO – UFRPE (Orientadora)
Universidade Federal Rural de Pernambuco

Prof. Dr. LUIZ FLÁVIO ARREGUY MAIA – UFRPE (Examinador Interno)
Universidade Federal Rural de Pernambuco

Prof. Dr. LEONARDO FERRAZ XAVIER (Examinador Interna)
Universidade Federal Rural de Pernambuco

AGRADECIMENTOS

Gostaria de expressar minha sincera gratidão a todas as pessoas que tornaram possível a realização deste Trabalho de Conclusão de Curso. Sem o apoio, incentivo e compreensão de muitas pessoas próximas a mim, esta jornada acadêmica teria sido muito mais desafiadora.

Primeiramente, quero agradecer aos meus pais pelo amor incondicional e encorajamento constante ao longo de todos esses anos de estudo. Suas palavras de estímulo e crença em meu potencial foram fundamentais para que eu persistisse nos momentos mais difíceis. Sem vocês, nada disto seria possível.

Além disso, gostaria de estender meus agradecimentos a minha orientadora e a meus professores e colegas de curso, cujas orientações, ensinamentos e discussões enriqueceram meu conhecimento e me guiaram ao longo desta jornada acadêmica.

RESUMO

Este estudo visou examinar o efeito da implementação dos Planos Municipais de Saneamento Básico (PMSBs) em indicadores de desempenho dos serviços prestados referentes ao abastecimento de água, coleta de esgoto e manejo de resíduos sólidos. A análise se baseia em uma abordagem metodológica que combina o uso de clusterização (por meio de k-médias), Propensity Score Matching (PSM) e regressão linear. A técnica de clusterização através do algoritmo K-médias foi utilizada para melhor caracterizar e classificar a amostra analisada. O PSM foi a ferramenta empregada para criar conjuntos de municípios comparáveis, com e sem PMSB, assegurando que suas características observáveis estivessem equilibradas e controlando possíveis vieses de seleção. A regressão linear foi aplicada para avaliar as relações entre o PMSB e o comportamento do Índice de Atendimento Total de Água, o Índice de Atendimento Total de Esgoto e a Taxa de Cobertura Regular do Serviço de Coleta de Resíduos Domiciliares. Os resultados derivados deste trabalho forneceram uma análise sólida do efeito dos Planos Municipais de Saneamento Básico (PMSBs) na administração do saneamento básico em nível municipal.

Palavras-chave: Clusterização. Propensity Score Matching. Regressão linear. Saneamento Básico. Plano Municipal de Saneamento Básico.

ABSTRACT

This study aims to examine the impact of the implementation of Municipal Basic Sanitation Plans (PMSB) on performance indicators related to water supply, sewage collection, and solid waste management services. The analysis is based on a methodological approach that combines the use of clustering (k-means), Propensity Score Matching (PSM), and linear regression. The clustering technique using the k-means algorithm will be employed to better characterize and classify the analyzed sample. The use of PSM will serve as the tool to create comparable sets of municipalities, with and without PMSB, ensuring that their observable characteristics are balanced and controlling potential selection biases. Linear regression will be applied to assess the relationships between PMSB and the behavior of the Total Water Coverage Index, the Total Sewage Coverage Index, and the Regular Household Waste Collection Service Coverage Rate. The results derived from this study will provide a robust analysis of the impact of Municipal Basic Sanitation Plans (PMSB) on municipal-level basic sanitation management.

Keywords: Clustering, Propensity Score Matching, Linear Regression, Basic Sanitation, Municipal Basic Sanitation Plan

LISTA DE FIGURAS

- | | |
|--|----|
| Figura 1 – Matriz de Avaliação do Potencial de Gestão do Saneamento Básico | 42 |
| Figura 2 – Quantidade de observações de tratamento e controle | 44 |

LISTA DE QUADROS

Quadro 1 –	Variáveis para a clusterização	28
Quadro 2 –	Variáveis para o Propensity Score Matching	30
Quadro 3 –	Variáveis para a análise do Índice de Atendimento Total de Água na regressão linear	31
Quadro 4 –	Variáveis para a análise do Índice de Atendimento Total de Esgoto Referido aos Municípios Atendidos com Água na regressão linear	32
Quadro 5 –	Variáveis para a análise da Taxa de Cobertura Regular do Serviço de Coleta de RDO em Relação à População Total do Município na regressão linear	33
Quadro 6 –	Resultado da clusterização via K-médias nas 4 primeiras observações	41
Quadro 7 –	Resultado da regressão linear do Índice de Atendimento Total de Água	45
Quadro 8 –	Estatísticas globais da regressão linear do Índice de Atendimento Total de Água	45
Quadro 9 –	Resultado da regressão linear do Índice de Atendimento Total de Esgoto Referido aos Municípios Atendidos com Água	47
Quadro 10 –	Estatísticas globais da regressão linear do Índice de Atendimento Total de Esgoto Referido aos Municípios Atendidos com Água	48
Quadro 11 –	Resultado da regressão linear da Taxa de Cobertura Regular do Serviço de Coleta de RDO em Relação à População Total do Município	48
Quadro 12 –	Estatísticas globais da regressão linear da Taxa de Cobertura Regular do Serviço de Coleta de RDO em Relação à População Total do Município	49

LISTA DE SIGLAS E ABREVIATURAS

PMGIRS	Planos Municipais de Gestão Integrada de Resíduos Sólidos
SNIS	Sistema Nacional de Informações de Saneamento
PSM	Propensity Score Matching
RSU	Resíduos Sólidos Urbanos
ODS	Objetivos de Desenvolvimento Sustentável
PIB	Produto Interno Bruto
WHO	World Health Organization
PMSB	Plano Municipal de Saneamento Básico
ONU	Organização das Nações Unidas
MQO	Mínimos Quadrados Ordinário
IBGE	Instituto Brasileiro de Geografia e Estatística
Unicef	Fundo Internacional de Emergência das Nações Unidas para a Infância
PS	Propensity Score
RDO	Resíduos Domiciliares
RPU	Resíduos Sólidos Públicos

SUMÁRIO

1	INTRODUÇÃO	12
2	REFERENCIAL TEÓRICO	15
2.1	Saneamento básico	15
2.1.1	<i>Abastecimento de água</i>	15
2.1.2	<i>Esgotamento sanitário</i>	15
2.1.3	<i>Drenagem urbana</i>	15
2.1.4	<i>Manejo de resíduos sólidos</i>	16
2.2	Política de Saneamento Básico definida pela Lei 11.445/2007	16
2.3	Plano Municipal de Saneamento Básico (PMSB)	17
2.4	Clusterização, Propensity Score Matching e Regressão linear	20
3	METODOLOGIA	25
3.1	Descrição geral da pesquisa	25
3.2	Base de dados	26
3.2.1	<i>Considerações sobre a amostra</i>	26
3.2.2	<i>Variáveis</i>	27
3.3	Clusterização, PSM e regressão linear	34
3.4	Propensity Score Matching	35
3.5	Regressão linear	38
4	RESULTADOS	41
4.1	Clusterização	41
4.2	Propensity Score Matching	43
4.3	Regressão linear	44
4.3.1	<i>Índice de Atendimento Total de Água (IN055_AE)</i>	44
4.3.2	<i>Índice de Atendimento Total de Esgoto Referido aos Municípios Atendidos com Água (IN056_AE)</i>	47
4.3.3	<i>Taxa de Cobertura Regular do Serviço de Coleta de RDO em Relação à População Total do Município (IN015_RS)</i>	48
5	CONCLUSÃO	51
	REFERÊNCIAS	53

1 INTRODUÇÃO

O saneamento básico é uma questão fundamental para a qualidade de vida e a saúde pública em qualquer sociedade. No Brasil, embora tenha havido avanços significativos nas últimas décadas, o acesso universal aos serviços de água potável e esgoto ainda é um desafio. A falta de saneamento adequado não apenas compromete a saúde das populações, mas também impacta negativamente a qualidade do meio ambiente e a sustentabilidade econômica. Além disso, o saneamento básico adequado também é fundamental para reduzir a propagação de doenças contagiosas, a exemplo da COVID-19, pois a lavagem das mãos com sabão e água limpa é uma das medidas mais eficazes para prevenir a propagação do vírus.

O saneamento básico desempenha um papel vital na promoção da saúde pública, da qualidade de vida e da sustentabilidade ambiental. No contexto brasileiro, apesar dos esforços significativos nas últimas décadas, os desafios persistem em relação ao acesso universal aos serviços de água potável e esgoto. Nesse cenário, os Planos Municipais de Saneamento Básico (PMSBs) surgem como instrumentos estratégicos para orientar ações e investimentos no setor com o objetivo de melhorar a infraestrutura sanitária e, conseqüentemente, elevar os índices de atendimento de água e esgoto nos municípios.

Estudos conduzidos por organizações internacionais, como a Organização das Nações Unidas (ONU), têm destacado repetidamente a importância crítica do saneamento básico como um dos pilares essenciais para o desenvolvimento sustentável. A ONU, por meio dos Objetivos de Desenvolvimento Sustentável (ODS), estabeleceu metas claras para garantir o acesso universal à água potável e ao saneamento adequado até 2030. No entanto, alcançar essas metas requer uma compreensão aprofundada dos fatores que influenciam o usufruto desses serviços essenciais. De acordo com a referida organização, em 2020, cerca de 4,2 bilhões de pessoas em todo o mundo não tinham acesso a instalações sanitárias básicas e em torno de 40% dos habitantes do globo viviam sem água e sabão para lavar as mãos. Também conforme a ONU, todos os dias, mais de 800 crianças morrem de doenças como diarreia e outras infecções causadas por falta de saneamento

e água contaminada.

Nos municípios brasileiros, os Planos Municipais de Saneamento Básico desempenham um papel crucial na definição das diretrizes e metas para o desenvolvimento de infraestrutura sanitária. Esses planos visam promover melhorias nos sistemas de abastecimento de água, tratamento de esgoto e manejo de resíduos sólidos.

Este estudo se propôs a investigar de maneira sistemática e empírica o impacto dos Planos Municipais de Saneamento Básico nos índices de atendimento de água e esgoto nos municípios do Brasil. Para alcançar esse objetivo, foram empregadas metodologias rigorosas, incluindo a clusterização (K-medias), o Propensity Score Matching (PSM) e a análise de regressão linear utilizando o método dos Mínimos Quadrados.

A clusterização, aplicada neste trabalho à análise de políticas públicas de saneamento básico, envolve a identificação regiões geográficas que compartilham características dos serviços de saneamento básico semelhantes. Isso permite que os formuladores de políticas identifiquem áreas com necessidades similares em termos de acesso a serviços de saneamento para que possam direcionar recursos e medidas específicas de maneira mais eficiente, garantindo um planejamento mais preciso e uma distribuição equitativa de recursos para melhorar a qualidade de vida das populações atendidas.

O Propensity Score Matching é uma abordagem estatística que visa equilibrar grupos de tratamento e controle, minimizando os vieses de seleção e permitindo uma comparação mais precisa dos resultados. Neste estudo, o PSM foi aplicado para selecionar municípios comparáveis, levando em consideração características socioeconômicas e demográfica.

A análise de regressão linear utilizando o método dos Mínimos Quadrados foi empregada para avaliar de forma quantitativa a relação entre a implementação dos PMSBs e os índices de atendimento de água e esgoto nos municípios. Através dessa análise, foi possível identificar os impactos diretos e indiretos dos planos no acesso à infraestrutura de saneamento básico.

Dessa forma, espera-se que este estudo contribua com a literatura sobre a eficácia dos PMSBs no Brasil, bem como com a utilização de uma

abordagem metodológica avançada objetivando a análise de políticas públicas no setor em questão. Os resultados obtidos têm o potencial de impactar tanto a formulação de políticas públicas quanto a alocação de recursos, a fim de otimizar os esforços em prol do desenvolvimento sustentável e da melhoria das condições de vida nos municípios brasileiros.

A relevância deste estudo se baseia na urgência de abordar as questões de saneamento básico que ainda persistem no Brasil. Apesar dos esforços em andamento, o país continua enfrentando desafios relacionados à falta de acesso universal à água potável e à infraestrutura de esgoto. Segundo o Instituto Trata Brasil, em 2022, quase 35 milhões de pessoas no Brasil vivem sem água tratada e cerca de 100 milhões não têm acesso à coleta de esgoto, resultando em doenças que poderiam ser evitadas, e que podem levar à morte por contaminação. Portanto, esta pesquisa visa contribuir para que ocorra uma compreensão mais clara dos efeitos dos Planos Municipais de Saneamento Básico no índice de atendimento de água e esgoto nos municípios do Brasil.

Espera-se que este trabalho possa oferecer *insights* valiosos para a formulação de políticas públicas mais eficazes, tendo em vista que, ao se analisarem os resultados alcançados pelos PMSBs em diferentes municípios, será possível identificar práticas bem-sucedidas, desafios comuns e lacunas que precisam ser superadas. Isso, por sua vez, pode influenciar as decisões de alocação de recursos e ações estratégicas para melhorar o saneamento básico em todo o país. A pesquisa também pode contribuir para o conhecimento acadêmico na área de Ciências Econômicas, pois aborda questões relacionadas à eficiência de investimentos públicos em saneamento básico e seus impactos socioeconômicos. Os resultados podem fornecer uma base para a realização de análises mais aprofundadas sobre a relação entre infraestrutura sanitária, desenvolvimento local e qualidade de vida.

2 REFERENCIAL TEÓRICO

2.1 Saneamento básico

O saneamento básico abrange quatro áreas fundamentais, que desempenham um papel vital na promoção da saúde pública, na qualidade de vida e na sustentabilidade ambiental. Cada uma dessas áreas é crucial para garantir o acesso a condições higiênicas e seguras para a população, e sua interconexão é essencial para alcançar o desenvolvimento sustentável.

2.1.1 Abastecimento de água

O abastecimento de água potável é uma necessidade fundamental para a vida humana e um dos pilares do saneamento básico. A garantia do acesso à água segura para consumo é um desafio global. A WHO (World Health Organization ou, em português, OMS – Organização Mundial da Saúde) e a Unicef têm abordado a importância do acesso à água potável como um direito humano fundamental, refletido nos Objetivos de Desenvolvimento Sustentável (ODS).

2.1.2 Esgotamento sanitário

O esgotamento sanitário é essencial para evitar a contaminação ambiental e proteger a saúde pública. A Lei 11.445/2007 (Brasil, 2007b), que estabelece diretrizes nacionais para o saneamento básico no Brasil, aborda a importância do esgotamento sanitário como parte integrante da política de saneamento. Autores como Tundisi (2015) enfatizam a necessidade de tratamento adequado dos esgotos para evitar doenças transmitidas pela água e degradação dos recursos hídricos.

2.1.3 Drenagem urbana

A gestão das águas pluviais é fundamental para prevenir inundações e alagamentos nas áreas urbanas. Autores como Oliveira (2021) discutem a

importância de sistemas de drenagem eficientes para lidar com as crescentes pressões urbanas e mudanças climáticas. Além disso, leis como o Estatuto da Cidade no Brasil (Brasil, 2001) destacam a necessidade de planejamento urbano integrado, incluindo a drenagem adequada.

2.1.4 Manejo de resíduos sólidos

O manejo adequado de resíduos sólidos é essencial para evitar a poluição do solo, água e ar. A Política Nacional de Resíduos Sólidos no Brasil, instituída pela Lei 12.305/2010 (Brasil, 2010), estabelece diretrizes para a gestão sustentável dos resíduos. Autores como Cossu e Lai (2007) discutem a importância da redução, reutilização e reciclagem de resíduos sólidos para minimizar os impactos ambientais.

A compreensão integrada dessas áreas é crucial para desenvolver políticas públicas eficazes de saneamento básico. A abordagem multidisciplinar, envolvendo aspectos de saúde, engenharia, planejamento urbano e legislação, é essencial para criar sistemas sustentáveis de saneamento que melhorem a qualidade de vida das comunidades e promovam a saúde ambiental. A interconexão dessas áreas é um reflexo do reconhecimento global da importância do saneamento básico para o alcance dos ODS e para um futuro mais saudável e sustentável.

2.2 Política de Saneamento Básico definida pela Lei 11.445/2007

Apesar de o saneamento ser um serviço público associado ao fornecimento de um direito social, a realidade atual das condições higiênicas no Brasil está longe do cenário ideal. Em grande parte das situações, essa situação emerge devido à falta de recursos financeiros disponíveis nas prefeituras, que têm a responsabilidade de oferecer tais serviços.

A Política de Saneamento Básico, definida pela Lei 11.445/2007 (Brasil, 2007b), representa um marco jurídico e institucional que estabeleceu as bases para a organização, regulação e implementação dos serviços de saneamento no Brasil. A legislação trouxe à tona a importância estratégica do acesso universal a serviços de água, esgoto, drenagem e manejo de resíduos sólidos,

ao mesmo tempo em que reconheceu o papel fundamental da gestão integrada, participativa e sustentável desses serviços para a promoção de saúde pública, qualidade de vida e preservação ambiental.

O enfoque central da Lei 11.445/2007 é a busca pela universalização dos serviços de saneamento básico com o intuito de atingir patamares mais elevados de saúde pública, meio ambiente equilibrado e desenvolvimento socioeconômico. A legislação define os princípios e diretrizes que norteiam a formulação e execução das políticas públicas de saneamento básico, promovendo a eficiência, transparência e equidade na prestação dos serviços.

A Lei de Saneamento Básico estabelece princípios econômicos e tarifários que visam à sustentabilidade financeira dos serviços, garantindo a manutenção, expansão e melhoria contínua das infraestruturas. A regulação dos serviços de saneamento é tratada de forma a incentivar a concorrência e promover a qualidade na prestação dos serviços.

Um dos elementos fundamentais da lei de saneamento básico é a exigência da elaboração de Planos Municipais de Saneamento Básico (PMSBs). Esses planos representam a estruturação da política nos níveis locais, permitindo que cada município desenvolva um conjunto de ações específicas para enfrentar seus desafios de saneamento. Os PMSBs devem ser participativos e contemplar diagnósticos detalhados, metas factíveis, programas, projetos e ações integradas que abordem as diferentes dimensões do saneamento.

2.3 Plano Municipal de Saneamento Básico (PMSB)

A Política de Saneamento Básico (Brasil, 2007b) reconheceu o direito ao saneamento como um direito humano essencial e estabeleceu a necessidade de assegurar à população o acesso a serviços de qualidade nas áreas de abastecimento de água, esgotamento sanitário, manejo de resíduos sólidos e drenagem urbana. Nesse contexto, o PMSB assume um papel preponderante, sendo concebido como um instrumento norteador para o planejamento, a implementação e a gestão das ações voltadas ao saneamento nos municípios. Cabe ressaltar que esses dois instrumentos, Plano e Política Municipal de Saneamento, são considerados essenciais tanto

pela lei quanto pelo decreto que a regulamenta. No entanto, um município pode ter um Plano Municipal de Saneamento Básico sem estabelecer a política e vice-versa.

Em muitas nações, a busca por padrões elevados de saneamento básico tem sido uma prioridade, alinhada com as metas estabelecidas por organizações internacionais como a Organização Mundial da Saúde e as Metas de Desenvolvimento Sustentável da ONU. Países de diferentes níveis de desenvolvimento reconhecem a importância do saneamento não apenas como um direito humano fundamental, mas também como um pilar essencial para o desenvolvimento social e econômico.

Por exemplo, a Índia adotou o "Swachh Bharat Abhiyan" (Campanha Limpa Índia) em 2014, um ambicioso programa nacional de saneamento que visa eliminar a defecação a céu aberto e promover a higiene. Esse programa é complementado pela elaboração de Planos de Saneamento nas áreas urbanas e rurais, abrangendo diversos aspectos da gestão de resíduos e do acesso à água potável.

Na França, a Lei para a Transição Energética (2015) também definiu objetivos ambiciosos para a gestão de resíduos e a promoção da economia circular. Esse quadro legal demanda a elaboração de Planos de Prevenção e Gestão de Resíduos Não Perigosos em níveis municipais e intermunicipais, demonstrando uma abordagem holística para o manejo dos resíduos sólidos.

O PMSB é um documento técnico que abarca um diagnóstico completo da situação atual dos serviços de saneamento no município. Sua elaboração envolve uma análise detalhada das demandas e das carências existentes, fornecendo um panorama preciso das condições de abastecimento de água, tratamento de esgoto, drenagem urbana e disposição final dos resíduos sólidos. Além disso, o diagnóstico realizado na construção do Plano leva em conta fatores sociais, econômicos, ambientais e institucionais, a fim de construir uma visão holística das necessidades locais.

De maneira abrangente, o Plano Municipal de Saneamento Básico deve englobar:

- Um levantamento detalhado da situação do saneamento e suas repercussões nas condições de vida, levando em conta diversos indicadores, como os sanitários, epidemiológicos, ambientais e

socioeconômicos;

- Estabelecimento de metas e objetivos para guiar o processo de desenvolvimento;
- Estratégias para abordar situações urgentes e imprevistas;
- Definição de métodos e procedimentos para avaliar a eficácia das ações planejadas.

É fundamental ressaltar que o processo de elaboração do PMSB deve ser participativo e inclusivo, envolvendo a sociedade civil, o setor público e o setor privado. A Lei 11.445/2007 enfatiza a importância da participação social como um dos pilares do PMSB, garantindo que os diversos atores interessados tenham voz nas decisões que impactam diretamente a qualidade de vida e o desenvolvimento sustentável do município.

O PMSB não se restringe ao diagnóstico, mas também estabelece metas e diretrizes que visam à ampliação e à melhoria dos serviços de saneamento básico. Essas metas devem ser realistas, mensuráveis e orientadas pela busca da sustentabilidade, considerando aspectos econômicos, sociais e ambientais. A lei prevê que os PMSBs devem ser revistos periodicamente para se adequarem às mudanças nas condições locais e aos avanços tecnológicos.

Conforme estipulado pela Lei, a existência do Plano Municipal de Saneamento Básico é uma pré-condição para que as cidades, e consequentemente as empresas responsáveis pelos serviços de água e esgoto, possam acessar recursos financeiros do governo federal destinados a projetos e empreendimentos de saneamento. Com esse propósito, o Decreto 10.203/2020 (Brasil, 2020) definiu um prazo até 31 de dezembro de 2022 para que as cidades elaborassem seus respectivos Planos de Saneamento Básico, garantindo, assim, a continuidade do acesso a financiamentos ou alocações orçamentárias públicas destinadas a esses serviços.

Dados de 2021 da pesquisa Perfil dos Municípios Brasileiros do Instituto Brasileiro de Geografia e Estatística (IBGE) indicam que menos de 41,5% das cidades brasileiras possuíam um Plano Municipal de Saneamento Básico, regulamentado ou não. Isso significa que menos da metade dos

municípios do país tem o documento estruturado para desenvolver os seus serviços de saneamento.

Diante desse contexto, não surpreende o cenário deficitário do acesso aos serviços no país, considerando que somente 41,5% dos municípios brasileiros têm um Plano Municipal de Saneamento Básico. A carência se reflete em números preocupantes: quase 35 milhões de brasileiros não têm acesso a abastecimento de água tratada, enquanto 100 milhões estão desprovidos de coleta de esgoto. A ausência desses serviços se correlaciona diretamente com a prevalência de doenças transmitidas pela água e acarreta problemas sociais significativos para as comunidades desfavorecidas por essa infraestrutura deficiente.

Ademais, a falta desses serviços impacta diretamente em indicadores de saúde, educação, renda e economia do país. Portanto, é de vital importância que a sociedade, como a principal interessada na melhoria das políticas de saneamento, participe ativamente nas tomadas de decisões relacionadas às ações e prioridades delineadas pelo PMSB, bem como no processo global de elaboração, avaliação e revisão desse plano.

A ausência de envolvimento social tem levado a um distanciamento entre as políticas governamentais e as reais necessidades da população. Portanto, é imperativo promover o estímulo à cooperação, através da ampla divulgação das propostas e da criação de canais efetivos de diálogo entre os partícipes da construção de um projeto que supra as demandas populacionais.

2.4 Clusterização, Propensity Score Matching e Regressão Linear

Tan *et al.* (2006) estabelecem que a finalidade primordial do agrupamento de dados reside na identificação de grupos de objetos, de modo que a similaridade entre os objetos internos de um grupo é enfatizada, enquanto os objetos pertencentes a diferentes grupos revelam dissimilaridades. A literatura que aborda a análise de agrupamento de dados é substancial, tendo suas origens remontando a 1954, conforme indicado por Jain (2010), em um estudo sobre dados antropológicos.

A análise de agrupamento, também denominada ***clustering***, engloba

um conjunto de técnicas computacionais voltadas para a segmentação de objetos em grupos distintos, denominados *clusters*, com base nas características individuais dos objetos, como mencionado por Linden (2009). Essa abordagem almeja alocar elementos com características semelhantes em *clusters* distintos dos elementos que apresentam características discrepantes, utilizando funções de dissimilaridade como critério.

Conforme definido por Everitt (1974), um grupo é conceituado como uma agregação de pontos no espaço, em que a distância entre qualquer par de pontos dentro do grupo é inferior à distância entre qualquer ponto dentro do grupo e qualquer ponto externo a ele. Ao longo do tempo, uma série de métodos e algoritmos foi desenvolvida para aprimorar a aplicação das técnicas de agrupamento de dados em contextos sociais e econômicos. Dentre esses métodos, destacam-se abordagens hierárquicas, k-médias, k-medoides, *fuzzy* e redes neurais. Todos eles partem do princípio fundamental da análise de grupo, que envolve a subdivisão de um conjunto de dados heterogêneo em subgrupos homogêneos.

Importante ressaltar que, ao contrário da técnica de classificação, na qual as classes são predefinidas, o agrupamento não requer classes preexistentes. Em vez disso, os elementos são agrupados com base em métricas de similaridade. De acordo com Zaiane (2003), uma análise de agrupamento eficaz demanda métodos que apresentem as seguintes características:

- Capacidade de manipular dados de elevada dimensionalidade;
- Adaptabilidade ao número de dimensões e à quantidade de elementos a serem agrupados;
- Flexibilidade para tratar distintos tipos de dados;
- Aptidão para estabelecer grupos de tamanhos e formas variáveis;
- Requisitos mínimos de conhecimento na determinação dos parâmetros de entrada;
- Sensibilidade à presença de ruído, com a capacidade de minimizar distorções causadas por valores atípicos;
- Obtenção de resultados consistentes independentemente da ordem de

apresentação dos dados.

Os algoritmos de agrupamento, que constituem uma categoria de técnicas analíticas, são frequentemente empregados na criação de *clusters*. Dentre essas técnicas, destacam-se os métodos hierárquicos que possuem a característica distintiva de permitir a fusão de *clusters* em etapas específicas do algoritmo, resultando na formação de múltiplos agrupamentos. Esses métodos organizam os dados com base na proximidade entre indivíduos, culminando na construção de uma árvore binária conhecida como **dendrograma**, como evidenciado por Doni (2004). O dendrograma retrata o conjunto total de dados na raiz, com as folhas representando os indivíduos finais.

Os métodos hierárquicos se subdividem em duas categorias principais: **métodos aglomerativos** e **divisivos**. No método aglomerativo, cada elemento ou grupo inicia como uma entidade independente e, ao longo do processo, une-se a outras entidades ou grupos com base em semelhanças, culminando na formação de um grupo único com todos os elementos. Por contraste, na abordagem divisiva, um grupo inicial contém todos os elementos do conjunto de dados e, progressivamente, é dividido em subgrupos, de modo que os elementos de um subgrupo apresentem uma determinada distância em relação aos elementos de outros subgrupos.

Um dos métodos não-hierárquicos mais notáveis é o **k-médias**, que será examinado em maior detalhe adiante e que representa o cerne desta investigação. Trata-se de um algoritmo de agrupamento de dados não hierárquico que emprega uma abordagem iterativa para particionar um conjunto de dados. Embora tenha sido proposto por S. Lloyd em 1957, sua publicação oficial só ocorreu em 1982. De maneira fundamental, esse algoritmo visa a minimizar a distância entre os elementos de um conjunto de dados e k-centros por meio de uma abordagem iterativa.

Ademais, o Método Propensity Score Matching (PSM) oferece uma abordagem alternativa para estimar os efeitos causais associados ao recebimento de tratamento em uma amostra de indivíduos. Essa técnica visa identificar indivíduos tanto no grupo tratado quanto no não tratado (grupo de controle) que possuam características próximas.

A introdução do método PSM na literatura ocorreu em 1983 por Paul Rosenbaum e Donald Rubin. A essência dessa abordagem reside em atribuir uma probabilidade de receber o tratamento a cada indivíduo da amostra, conhecida como **escore de propensão** (ou, do inglês, *propensity scores* – PS), enquanto se controlam suas características observadas. Posteriormente, as unidades de ambos os grupos são pareadas com *propensity scores* semelhantes, permitindo a comparação dos resultados obtidos entre esses pares. Esse modelo tem o potencial de abordar o problema de viés de seleção, proporcionando estimativas imparciais do efeito médio do tratamento.

Dehejia e Wahba (1999) aplicaram o PSM na economia para estimar o impacto causal de programas de treinamento em empregos, demonstrando que essa ferramenta é eficaz no controle de vieses de seleção em análises observacionais e ressaltando a importância de verificar a robustez dos resultados.

Na área das ciências médicas, Wolfe e Michaud (2004) usaram o PSM para avaliar os efeitos de tratamentos médicos em estudos observacionais. Eles destacaram que esse método pode ser crucial para controlar o viés de seleção e fornecer estimativas mais confiáveis dos efeitos causais de intervenções médicas.

Campello, Graham e Harvey (2010) salientaram que o PSM é uma técnica valiosa para isolar os efeitos específicos de uma política ou ação sobre os resultados financeiros de uma empresa. Nas finanças, exploraram tal recurso metodológico para analisar os impactos de decisões corporativas em empresas.

Para avaliar os efeitos de políticas e intervenções sociais, Gangl (2006) e Grodsky (2007) utilizaram o PSM, observando a sua eficácia para lidar com viés de seleção e melhorar a validade interna das análises em estudos observacionais.

Arceneaux, Gerber e Green (2006) utilizaram o PSM para investigar os efeitos de campanhas políticas e, em seu estudo, enfatizaram a relevância dessa ferramenta para controlar fatores de confusão em estudos observacionais e permitir inferências mais sólidas sobre causalidade.

É preciso, também, comentar acerca da regressão linear, método que tem suas raízes nos trabalhos de Francis Galton e Karl Pearson no final do

século XIX. No entanto, foi o trabalho de Sir Ronald A. Fisher que formalizou a metodologia estatística da regressão linear na década de 1920. Ao longo dos anos, diversos pesquisadores, como George W. Snedecor e Herman Wold, contribuíram para o desenvolvimento da técnica, levando a uma ampla gama de variantes e aplicações.

Hair *et al.* (2009, p. 176) destacam que a “análise de regressão múltipla é uma ferramenta estatística que permite investigar a ligação entre uma variável dependente única e múltiplas variáveis independentes (também conhecidas como variáveis preditoras)”. Nesse sentido, por meio da regressão, é viável avaliar a extensão da associação entre a variável Y , que é a que depende do estudo, e o conjunto de variáveis independentes X_i .

O método dos mínimos quadrados é a técnica usada na regressão linear para encontrar essa linha ou hiperplano que melhor se ajusta aos dados observados. Para Gujarati (2011), ele é utilizado para determinar os coeficientes da regressão que minimizam a soma dos quadrados das diferenças entre os valores observados da variável dependente e os valores previstos pela linha de regressão. A ideia por trás do método dos mínimos quadrados é escolher os coeficientes da regressão de forma que a soma dos quadrados dos resíduos (ou seja, as diferenças entre os valores observados e os valores previstos) seja a menor possível. Isso garante que a linha de regressão se aproxime o máximo possível dos pontos de dados observados.

3 METODOLOGIA

3.1 Descrição geral da pesquisa

Conforme Gerhardt e Silveira (2009) argumentam, a pesquisa científica pode ser categorizada com base em quatro características distintas: abordagem, natureza, objetivos e procedimentos.

Quanto à abordagem, este estudo é identificado como uma pesquisa quantitativa. De acordo com as autoras, essa abordagem se destaca pelo uso de raciocínio dedutivo, princípios da lógica e atributos mensuráveis da experiência humana. Além disso, a pesquisa quantitativa concentra-se em um número limitado de conceitos, é objetiva e emprega análise de dados por meio de métodos estatísticos. Ela parte do pressuposto de que já existem ideias preconcebidas sobre a relação entre os conceitos.

Em relação à natureza, uma pesquisa pode ser classificada como básica ou aplicada. Neste estudo, a classificação é de pesquisa aplicada, conforme definido por Gil (2002). Isso significa que seu propósito principal é gerar conhecimento que possa ser aplicado na prática para resolver problemas específicos dentro do contexto em questão. Isso difere da pesquisa básica, que não tem uma aplicação prática imediata prevista e lida com verdades e interesses universais.

Quanto aos objetivos, este trabalho se enquadra na categoria de pesquisa descritiva. De acordo com Prodanov e Freitas (2013), a pesquisa descritiva busca descrever as características de uma determinada população ou fenômeno, ou estabelecer relações entre variáveis — neste caso, as variáveis explicativas relacionadas ao índice de atendimento dos serviços de saneamento básico. Para alcançar esses objetivos, foi empregado um método quantitativo para examinar as relações entre as variáveis do modelo, caracterizando-se, assim, como uma pesquisa quantitativa descritiva, conforme conceituado por Mattar (1994).

Do ponto de vista dos procedimentos técnicos, uma pesquisa pode ser classificada de várias maneiras, incluindo pesquisa bibliográfica, documental, experimental (mais comum nas ciências biológicas e tecnológicas), estudo de caso, levantamento, entre outras abordagens metodológicas. O procedimento

técnico deste estudo pode ser caracterizado como um estudo de caso, que envolve a coleta e análise de informações sobre um indivíduo, uma família, um grupo ou uma comunidade específica, com o propósito de estudar diversos aspectos de suas vidas, de acordo com o tema da pesquisa, conforme descrito por Gerhardt e Silveira (2009).

3.2 Base de dados

Nesta seção, serão delineadas todas as variáveis empregadas na formulação do modelo, descrevendo suas particularidades, relevância no contexto do estudo e a fonte pela qual os dados relacionados a cada variável foram obtidos.

Todos os dados são divulgados pelo Ministério da Integração do Desenvolvimento Regional através do Sistema Nacional de Informações sobre Saneamento (SNIS).

3.2.1 Considerações sobre a amostra

Dentro do escopo deste estudo, obtivemos os dados do Sistema Nacional de Informações sobre Saneamento (SNIS). Para coletá-los, na interface de acesso às séries históricas, selecionamos a guia "Município" e, em seguida, optamos por "Agrupamento dinâmico de indicadores e informações por ano de referência". Então, foram feitas as seguintes seleções:

- Todos os Municípios do Brasil;
- Ano de referência: 2021;
- Todas as regiões;
- Todas as regiões metropolitanas;
- Todos os municípios.

Após confirmar a seleção, ao clicar em "continuar", assinalou-se a opção "marcar todos" em "Famílias de Informações e Indicadores".

A variável de interesse, intitulada "PO028 – O município possui Política

de Saneamento Básico Conforme a Lei 11.445/2007?” (será utilizado o apenas termo PO028 para simplificar) será a base para o *propensity score match* e será o foco da análise dos resultados da regressão linear, para entender se/como ela interfere nos índices de desempenhos dos serviços de saneamento básico nos municípios

Dos 5.570 municípios do Brasil, 4.900 municípios possuem respostas válidas na variável de interesse PO028. No decorrer deste trabalho, utilizaremos apenas os municípios que tiveram alguma resposta (sim ou não) para essa variável, de forma que os 670 municípios nos quais o campo da variável PO001 estava vazio foram excluídos. Desses 5.570 municípios com respostas válidas, 2692 municípios tiveram como resposta “Sim” (possuem plano municipal de saneamento básico) na variável PO028, sendo assim, 2.208 municípios tiveram como resposta “Não” (não possuem o plano municipal de saneamento básico) na variável PO028.

Para este estudo, transformamos a variável PO028 em uma variável *dummy*, ou seja, os municípios com plano municipal de saneamento terão no campo PO028 o número 1 para indicar que o município possui o plano, e os municípios que não dispõem dele terão no campo PO028 o número 0, indicando que a ausência do plano. A variável *dummy* é empregada para representar uma característica categórica ou binária, de modo que "1" denota a presença dessa característica, e "0" denota a ausência dela.

3.2.2 Variáveis

Durante o presente trabalho foram utilizadas diversas variáveis para executar a clusterização, o *propensity score matching* e a regressão linear. Em todos os métodos citados, foi utilizado o campo PO028 da tabela do SNIS, que é a variável que avalia se um município possui um plano municipal de saneamento básico que atenda aos requisitos estabelecidos na Lei Federal nº 11.445/2007.

As variáveis utilizadas na clusterização foram selecionadas a partir das variáveis *dummy* disponibilizadas pelo SNIS que indicam a adoção de boas práticas para a gestão do saneamento básico, ou seja, variáveis que impactam o potencial da gestão no municípios. Na tabela 1, são apresentadas

as variáveis utilizadas na clusterização.

Quadro 1 – Variáveis para a clusterização

Variáveis na tabela do SNIS	Siglas
PO028 – O município possui Plano Municipal de Saneamento Básico elaborado nos termos estabelecidos na Lei nº 11.445/2007?	PO028
PO001 – O município possui Política de Saneamento Básico Conforme a Lei nº 11.445/2007?	PO001
PO042 – O município é integrante de algum consórcio intermunicipal regulamentado pela Lei nº 11.107/2005 que tenha entre suas atribuições específicas a gestão ou prestação de um ou mais serviços de manejo de RSU (serviços de coleta de resíduos domiciliares ou públicos, operações de aterro sanitário etc.)	PO042
PO048 – O Município possui Plano Municipal de Gestão Integrada de Resíduos Sólidos (PMGIRS) conforme a Lei nº 12.305/2010 que trata da Política Nacional de Resíduos Sólidos?	PO048
QD001 – Tipo de atendimento da portaria sobre qualidade da água	QD001

Fonte: Elaborado pelo autor.

A variável PO001 se refere à conformidade de um município com a legislação brasileira que estabelece diretrizes nacionais para o saneamento básico. A Lei nº 11.445/2007, também conhecida como Lei do Saneamento Básico, é uma legislação federal que estabelece os princípios, diretrizes e normas para a prestação de serviços de saneamento básico no Brasil. Se o município possui a política de saneamento básico, é contabilizado 1 ponto no *rating* para o município.

A variável PO042 refere-se à participação de um município em um consórcio intermunicipal, que é uma forma de cooperação entre municípios no Brasil, regulamentada pela Lei nº 11.107/2005. Essa participação ocorre quando o município faz parte de uma entidade conjunta que tem como uma de suas atribuições específicas a gestão ou a prestação de serviços relacionados ao manejo de Resíduos Sólidos Urbanos (RSU). Se o município participa de um consórcio municipal, é contabilizado 1 ponto no *rating* para o município.

A variável PO048 refere-se à existência do Plano Municipal de Gestão Integrada de Resíduos Sólidos (PMGIRS), em conformidade com a Lei nº 12.305/2010, que estabelece a Política Nacional de Resíduos Sólidos no Brasil (Brasil, 2010). O PMGIRS é um instrumento fundamental para orientar

e sistematizar as ações relacionadas à gestão de resíduos sólidos em nível municipal. Ele deve conter diretrizes, metas, estratégias e ações a serem implementadas pelo município para lidar com questões como coleta seletiva, disposição final de resíduos, reciclagem, educação ambiental e outras medidas relacionadas à gestão ambientalmente adequada dos resíduos sólidos. Se o município possui PMGIRS, é contabilizado 1 ponto no *rating* para o município.

A variável QD001 refere-se ao modo como as diretrizes e regulamentações estabelecidas em uma portaria relacionada à qualidade da água estão sendo implementadas e seguidas. Ela se concentra na maneira como as autoridades ou organizações responsáveis estão atendendo às normas e padrões estabelecidos para garantir a qualidade da água. Essa variável pode ter diferentes valores, que indicam os diferentes tipos de atendimento, os quais estão descritos a seguir.

- **Atendimento Integral:** significa que todas as diretrizes e regulamentações da portaria estão sendo plenamente seguidas e implementadas de acordo com os padrões estabelecidos. Isso implica que todas as medidas necessárias estão sendo tomadas para garantir a qualidade da água;
- **Atendimento Parcial:** nesse caso, parte das diretrizes da portaria está sendo implementada, mas não há uma instauração integral. Pode haver alguns aspectos em que o atendimento não esteja completo ou plenamente em conformidade com os padrões;
- **Não Atendimento:** indica que as diretrizes e regulamentações da portaria não estão sendo seguidas ou implementadas. Pode ser o caso de que nenhuma medida está sendo tomada para garantir a qualidade da água de acordo com os padrões estabelecidos na portaria.

Essa variável é importante para avaliar o cumprimento das regulamentações de qualidade da água e garantir que a água fornecida à população atenda aos padrões de segurança e qualidade estabelecidos pelas autoridades reguladoras. O tipo de atendimento pode variar de acordo com a região, a infraestrutura disponível e outros fatores que afetam a gestão da qualidade da água. Se o município possui atendimento integral é contabilizado

2 pontos no rating do município, se o município possui atendimento parcial é contabilizado 1 ponto no rating do município, e se o município não possui atendimento não é adicionado ponto rating.

No Quadro 2, são apresentadas as variáveis utilizadas para o Propensity Score Matching (PSM).

Quadro 2 – Variáveis para o Propensity Score Matching

Variáveis na tabela do SNIS	Siglas	Tipo
POP_TOT – População total do município (Fonte: IBGE)	POP_TOT	Contínua
PO028 – O município possui Plano Municipal de Saneamento Básico, elaborado nos termos estabelecidos na Lei nº 11.445/2007?	PO028	Dummy
FN017 – Despesas totais com os serviços (DTS)	FN017	Contínua
FN220 – Despesas totais com serviços de manejo de RSU (Antigo campo GE007)	FN220	Contínua
FN223 – Despesas correntes da Prefeitura durante o ano com TODOS os serviços do município (saúde, educação, pagamento de pessoal, etc.) (Antigo campo GE010)	FN223	Contínua

Fonte: Elaborado pelo autor.

PO028 é a variável indicadora de tratamento, ao passo que POP_TOT, FN017, FN220 e FN223 são as variáveis que representam características que podem influenciar a probabilidade de o município receber o tratamento.

POP_TOT é uma medida que indica o número total de habitantes que residem na área geográfica correspondente ao município. Essa informação é coletada e fornecida pelo Instituto Brasileiro de Geografia e Estatística (IBGE), uma instituição responsável por coletar, analisar e divulgar dados estatísticos sobre o Brasil.

A variável FN017 se refere às despesas totais incorridas pelo município no fornecimento e na gestão dos serviços públicos relacionados ao abastecimento de água e esgotamento sanitário.

A variável FN220 se refere aos custos totais relacionados aos serviços de manejo de Resíduos Sólidos Urbanos (RSU) realizados pelo prestador de serviços em um determinado município ou região. Os RSU são os resíduos sólidos gerados nas áreas urbanas, incluindo o lixo doméstico e outros tipos

de resíduos produzidos pela população e pelas atividades urbanas.

FN223 é uma medida financeira que representa as despesas correntes totais realizadas pela Prefeitura do município ao longo do ano. Essas despesas englobam todos os serviços e atividades que a administração pública municipal é responsável por fornecer à comunidade. O intuito é selecionar municípios economicamente semelhantes para os grupos de controle e de tratamento.

Para a análise do comportamento do Índice de Atendimento Total de água na regressão linear, teremos as seguintes variáveis:

Quadro 3 – Variáveis para a análise do Índice de Atendimento Total de Água na regressão linear

Variáveis na tabela do SNIS	Siglas	Tipo	Explicada/Explicativa	Sinal esperado
IN055_AE – Índice de Atendimento Total de Água	IN055_AE	Contínua	Explicada	xxxxxxxx
PO028 – O município possui Plano Municipal de Saneamento Básico elaborado nos termos estabelecidos na Lei nº 11.445/2007?	PO028	Dummy	Explicativa	+
FN023 – Investimento realizado em abastecimento de água pelo prestador de serviços	FN023	Contínua	Explicativa	+
IN005_AE - Tarifa média de água	IN005_AE	Contínua	Explicativa	+

Fonte: Elaborado pelo autor.

A variável IN055_AE é uma métrica que avalia a extensão do acesso da população de um município aos serviços de abastecimento de água. Ela é o percentual da população total atendida com abastecimento de água pela população total residente do(s) município(s) com abastecimento de água. Essa métrica é essencial para entender a disponibilidade desse serviço básico para uma determinada população.

A variável FN023 é uma medida que avalia os recursos financeiros destinados ao setor de abastecimento de água por parte da entidade prestadora de serviços, seja ela pública ou privada.

A variável IN005_AE é uma métrica que avalia o custo médio que os

consumidores pagam pelo fornecimento de água em um determinado município. É a receita operacional dividida pelo volume de água.

Para a análise do comportamento do Índice de Atendimento Total de Esgoto Referido aos Municípios Atendidos com Água na regressão linear, temos as seguintes variáveis:

Quadro 4 – Variáveis para a análise do Índice de Atendimento Total de Esgoto Referido aos Municípios Atendidos com Água na regressão linear

Variáveis na tabela do SNIS	Siglas	Tipo	Explicada/ Explicativa	Sinal esperado
IN056_AE – Índice de Atendimento Total de Esgoto Referido aos Municípios Atendidos com Água	IN056_AE	Contínua	Explicada	xxxxxxx
PO028 – O município possui Plano Municipal de Saneamento Básico elaborado nos termos estabelecidos na Lei nº 11.445/2007?	PO028	Dummy	Explicativa	+
FN024 - Investimento realizado em esgotamento sanitário pelo prestador de serviços	FN024	Contínua	Explicativa	+
IN006_AE - Tarifa média de esgoto	IN006_AE	Contínua	Explicativa	+

Fonte: Elaborado pelo autor.

A variável IN056_AE é uma métrica que avalia a extensão do serviço de esgotamento sanitário em relação àqueles municípios que já têm acesso ao serviço de abastecimento de água. É a razão da população total atendida com esgotamento sanitário pela população total residente dos municípios com abastecimento de água. Essa métrica é relevante para compreender o quão eficaz é o serviço de coleta de esgoto em termos de cobertura.

A variável FN024 é o montante de investimentos feitos pelo prestador de serviços de saneamento básico em esgotamento sanitário em um determinado período, sendo ele feito pelo setor público ou privado.

A variável IN006_AE refere-se ao valor médio da tarifa cobrada dos usuários pelos serviços de coleta e tratamento de esgoto sanitário. É a receita operacional dividida pelo volume de esgoto coletado.

Para a análise do comportamento da Taxa de Cobertura Regular do

Serviço de Coleta de Resíduos Domiciliares (RDO) em relação à população total do município na regressão linear, temos as seguintes variáveis:

Quadro 5 – Variáveis para a análise da Taxa de Cobertura Regular do Serviço de Coleta de RDO em Relação à População Total do Município na regressão linear

Variáveis na tabela do SNIS	Siglas	Tipo	Explicada/ Explicativa	Sinal esperado
IN015_RS – Taxa de cobertura regular do serviço de coleta de RDO em relação à população total do município	IN015_RS	Contínua	Explicada	xxxxxxxx
PO028 – O município possui Plano Municipal de Saneamento Básico elaborado nos termos estabelecidos na Lei nº 11.445/2007?	PO028	Dummy	Explicativa	+
IN019_RS – Taxa de Empregados (Coletadores + Motoristas) na Coleta (RDO + RPU) em Relação à População Urbana	TB001	Discreta	Explicativa	+
IN023_RS – Custo Unitário Médio do Serviço de Coleta (RDO + RPU)	TB002	Discreta	Explicativa	+

Fonte: Elaborado pelo autor.

A variável IN015_RS é a proporção da população total de um município que é atendida regularmente pelo serviço de coleta de resíduos domiciliares. Essa variável é fundamental para avaliar o acesso da população a um serviço essencial de saneamento básico, que é a coleta adequada de resíduos domésticos.

A variável IN019_RS é a proporção de trabalhadores (motoristas e coletadores) envolvidos na coleta de Resíduos Domiciliares (RDO) e resíduos públicos urbanos (RPU) em relação à população do município.

A variável IN023_RS é a razão das despesas dos agentes públicos e privados com o serviço de coleta de RDO e RPU pela quantidade total de RDO e RPU coletadas.

3.3 Clusterização, PSM e regressão linear

A técnica de agrupamento de dados, comumente referida como *clustering*, constitui um procedimento analítico direcionado à caracterização de grupos ou conjuntos de dados, fundamentado na identificação de similaridades intrínsecas. Essa abordagem envolve a consolidação das informações obtidas, por meio da aplicação de uma função discriminativa, com o intuito de segmentar os dados em elementos que compartilhem semelhanças e distinções.

A clusterização elaborada no presente trabalho foi utilizada para mapear o potencial de gestão de saneamento básico dos municípios do Brasil. A clusterização consistirá em agrupar os municípios com gestão de saneamento básico semelhantes. Isso ajuda os formuladores de políticas a identificar áreas com necessidades comuns em termos de acesso a serviços de saneamento. A aplicação do algoritmo k-médias e a interpretação dos grupos resultantes permitirão obter uma compreensão mais aprofundada das diversas dinâmicas e tendências relacionadas à gestão de saneamento básico em nível municipal. Identificar esses grupos representa um passo significativo na formulação de estratégias mais eficientes e direcionadas para aprimorar a gestão de saneamento básico em várias regiões, permitindo uma alocação mais eficiente de recursos e a implementação de medidas específicas para melhorar a qualidade de vida das comunidades atendidas.

O processo de clusterização começou com a etapa de preparação dos dados, na qual os dados foram adequadamente formatados, incluindo a extração da coluna de notas de *rating* dos municípios, que é a variável principal para a clusterização. Em seguida, foi decidido o número de *clusters*, optando por três grupos distintos, para analisar os três níveis de capacidade de gestão (Alto, Médio e Baixo).

O algoritmo k-médias foi, então, aplicado aos dados de notas de *rating*, começando com a seleção aleatória de três pontos como centros iniciais dos *clusters*, denominados **médias**. Cada ponto de dado foi atribuído ao *cluster* cujo centro estava mais próximo. Após isso, ocorreu a atualização das médias, na qual o algoritmo calculou a média de todos os pontos em cada *cluster* e atualizou os centros correspondentes. Em seguida, houve a reatribuição de

pontos, de forma que cada ponto de dado foi novamente atribuído ao *cluster* cujo centro estava mais próximo após a atualização.

Esses passos foram repetidos iterativamente até que não houvesse mudanças significativas nas atribuições dos pontos aos *clusters* ou até que um número máximo de iterações fosse atingido. O resultado final foi uma solução em que os pontos foram agrupados em *clusters* de forma a minimizar a soma dos quadrados das distâncias entre os pontos e os centros dos *clusters*, resultando em uma lista de *clusters*, com cada ponto de dado associado a um *cluster* específico.

3.4 Propensity Score Matching

Neste trabalho, o Propensity Score Matching foi a técnica de análise utilizada para buscar igualar as características das unidades de observação (neste caso, os municípios) que receberam o tratamento (presença do PMSB) e aquelas que não receberam, a fim de criar grupos comparáveis para, somente assim, realizar a regressão linear e avaliar o impacto do PMSB em indicadores de desempenho dos serviços prestados de saneamento básico. Isso é feito calculando um escore de propensão, que representa a probabilidade de cada unidade receber o tratamento com base em suas características observáveis, na qual foi utilizada o pacote MatchIt da linguagem R (R FOUNDATION, 2016) e o método de correspondência pelo “vizinho mais próximo” (Do inglês, *nearest neighbor matching* – NNM).

O *propensity score* é uma probabilidade condicional de que um participante em um estudo receba o tratamento, dadas as variáveis observadas. Portanto, essa probabilidade é atribuída não apenas aos participantes que efetivamente receberam o tratamento, mas também àqueles que não o receberam. Adicionalmente, ao condicionar pelo escore de propensão, cada participante apresenta a mesma probabilidade de ser alocado para o tratamento, semelhante a um experimento randomizado. Neste trabalho, utilizamos o PSM para definir o grupo de tratamento e o grupo de controle, de tal forma que características predeterminadas entre os grupos (controle e de tratamento) sejam parecidas, para reduzir o viés de seleção.

O *propensity score* pode ser considerado um "*balancing score*" (pontuação de equilíbrio), que representa um conjunto de covariáveis. Nesse sentido, um par de indivíduos, um do grupo de controle e outro do grupo de tratamento, que possuam um propensity score semelhante, podem ser considerados comparáveis, mesmo que apresentem diferenças nas covariáveis específicas. Isso possibilita uma comparação mais justa entre os grupos, controlando as influências das variáveis observadas e permitindo uma avaliação mais precisa do efeito causal do tratamento.

São vantagens do PSM:

- Equivalência aproximada – o PSM possibilita a criação de grupos tratados e de controle comparáveis em relação às características observáveis, contribuindo para uma comparação mais justa e controlada;
- Utilização de dados observacionais – em situações em que experimentos controlados não são viáveis ou éticos, o PSM é uma ferramenta valiosa para analisar o impacto de políticas públicas usando dados observacionais do mundo real;
- Mitigação de viés de seleção – o PSM aborda o desafio do viés de seleção, comum em análises de políticas públicas, ao considerar as características que influenciam a escolha de aderir a um tratamento específico.

Por outro lado, como desvantagens do PSM, destacam-se:

- Suposições críticas – O PSM depende da suposição de ignorabilidade condicional, ou seja, de que todas as variáveis relevantes que influenciam a seleção no tratamento estão incluídas no modelo de propensão;
- Viés residual e não observáveis – mesmo após o PSM, podem persistir diferenças não observáveis entre os grupos tratado e de controle, levando a um viés residual;
- Número de variáveis – em cenários com muitas variáveis explicativas, a estimativa do escore de propensão pode se tornar menos estável.

Alguns motivos para utilizar o Propensity Score Matching em análises de políticas públicas são destacados a seguir:

- Ausência de experimentos controlados – em muitos casos, a realização de experimentos controlados não é viável, tornando o PSM uma alternativa viável para avaliar o impacto de políticas;
- Seleção não aleatória – políticas públicas frequentemente não são aplicadas de forma aleatória, o que pode gerar viés de seleção, e o PSM ajuda a controlar essa seleção não aleatória;
- Avaliação causal – o PSM permite estimar causalmente o efeito de uma política ou intervenção, considerando características observáveis e controlando viés de seleção;
- Uso de dados disponíveis – em muitos casos, os dados observacionais são a única fonte viável de informação, e o PSM se mostra eficaz porque oferece uma abordagem rigorosa para extrair conclusões causais desses dados.

No PSM, é necessário definir a variável que representa o tratamento (neste trabalho será a variável *dummy* referente a se o município possui o PMSB) e as variáveis que representam características que podem influenciar a probabilidade de receber o tratamento. São variáveis observáveis que ajudam a equilibrar os grupos de tratamento e controle.

A abordagem do Propensity Score Matching, ao controlar o viés de seleção, permitir a comparação entre grupos comparáveis e contribuir para a avaliação causal de políticas públicas, desempenha um papel fundamental na análise de intervenções em cenários do mundo real. Ao lidar com dados observacionais e possibilitar a comparação de grupos tratados e de controle de maneira mais precisa, o PSM fornece *insights* valiosos sobre os efeitos das políticas, auxiliando na tomada de decisões informadas e eficazes.

Para cada município, foi calculada a probabilidade de ter um PMSB com base nas informações contidas nos campos disponíveis na base de dados do SNIS, os quais são, neste trabalho, os campos FN017, FN208, FN223 e POP_TOT, como mencionado na seção 3.2.

Com o Propensity Score Matching, são escolhidos municípios que não tenham o PMSB, mas que apresentem características semelhantes às descritas anteriormente, de modo a criar um grupo de comparação. Esses municípios de controle são selecionados com base na proximidade de suas características (FN017, FN208, FN223 e POP_TOT) com os municípios que possuem o PMSB. Dessa forma, são definidos o grupo de controle (municípios com o PMSB) e o grupo de controle da regressão linear.

3.5 Regressão Linear

A regressão linear é uma técnica estatística utilizada para modelar a relação entre uma variável dependente e uma ou mais variáveis independentes. Ela é amplamente empregada para entender como as mudanças nas variáveis independentes afetam a variável dependente e para fazer previsões com base nesses relacionamentos.

Tem-se como propósito, no presente trabalho, usar a regressão linear para estabelecer uma relação matemática entre as variáveis independentes e a variável dependente apresentadas na seção 3.2. Por meio da análise estatística, o modelo de regressão busca identificar como as variáveis independentes afetam a variável dependente e quantificar essa relação por meio dos coeficientes de regressão. Isso permite realizar previsões ou inferências sobre como as variáveis independentes impactam a variável dependente.

O objetivo da regressão linear é encontrar uma linha reta (ou hiperplano, em casos de múltiplas variáveis independentes) que melhor se ajusta aos dados observados. Essa linha é chamada de **linha de regressão** e é definida pelos coeficientes da regressão. Tais coeficientes indicam a magnitude e a direção do impacto das variáveis independentes na variável dependente. Isso inclui identificar relações causais entre variáveis, prever valores futuros com base em padrões históricos e avaliar o impacto de intervenções políticas. Ao permitir a quantificação dos efeitos das variáveis independentes sobre a variável dependente, a regressão linear oferece uma estrutura analítica sólida para entender os fatores que influenciam as políticas públicas.

Existem dois tipos principais de regressão linear: simples e múltipla. Na regressão linear simples, há apenas uma variável independente, enquanto, na regressão linear múltipla, várias variáveis independentes são consideradas. A regressão linear é comumente utilizada em análises estatísticas, econômicas, científicas e sociais para modelar relações, fazer previsões e testar hipóteses sobre os efeitos das variáveis independentes na variável dependente.

Apesar de sua ampla aplicabilidade, a regressão linear tem limitações. Ela pressupõe uma relação linear entre as variáveis, o que pode não refletir adequadamente a complexidade das relações do mundo real. Além disso, a presença de multicolinearidade entre as variáveis independentes pode levar a resultados imprecisos. A regressão linear também pode ser sensível a *outliers* e não captura relações não lineares.

A principal vantagem da regressão linear é sua interpretabilidade. Os coeficientes estimados permitem avaliar diretamente o tamanho e a direção das relações entre variáveis. Além disso, a regressão linear fornece intervalos de confiança e testes estatísticos que auxiliam na avaliação da significância dos resultados. É uma ferramenta valiosa para análises exploratórias e modelagem de políticas.

As desvantagens da regressão linear incluem a rigidez na modelagem de relações não lineares, bem como a vulnerabilidade a suposições sobre a distribuição dos erros e a independência entre observações. Em situações em que os dados não atendem a essas suposições, os resultados da regressão linear podem ser distorcidos e enganosos.

Em uma análise de regressão linear, as variáveis dependentes e as variáveis independentes desempenham papéis fundamentais na compreensão e modelagem das relações entre diferentes conjuntos de dados. Essas variáveis têm funções distintas e são essenciais para construir e interpretar um modelo de regressão.

Variáveis dependentes ou explicadas, em uma análise de regressão linear, são aquelas que representam os resultados, observações ou valores que você deseja entender, prever ou explicar por meio do modelo de regressão. São o foco principal da análise e o que se procura relacionar com as variáveis independentes. Em outras palavras, a variável dependente é aquela que se está tentando explicar ou entender com base nas mudanças

nas variáveis independentes.

Variáveis independentes ou explicativas, em uma análise de regressão linear, são aquelas que representam os fatores, características ou influências que são usados para explicar ou prever as variações na variável dependente. Elas são as variáveis que se acredita terem um impacto sobre a variável dependente e são manipuladas ou observadas para determinar como afetam os resultados.

Na regressão linear, pudemos analisar o comportamento de 3 variáveis dependentes: Índice de Atendimento Total de Água, Índice de Atendimento Total de Esgoto referido aos municípios atendidos com água e a Taxa de Cobertura Regular do Serviço de Coleta de RDO em relação à População Total do Município. Por meio delas, tornou-se possível avaliar os serviços prestados referente ao abastecimento de água, coleta de esgoto e gestão de resíduos sólidos respectivamente, utilizando variáveis independentes apresentadas nos quadros 3, 4 e 5.

A regressão linear é especialmente útil em análises de políticas públicas devido à sua simplicidade e capacidade de fornecer resultados quantitativos interpretáveis. Ela permite avaliar o impacto das intervenções governamentais, controlando outros fatores relevantes. Quando usada com cuidado e considerando suas limitações, a regressão linear pode oferecer *insights* valiosos para embasar a tomada de decisões informadas em políticas públicas.

4 RESULTADO

4.1 Clusterização

Os agrupamentos resultantes após a utilização do algoritmo k-médias denotam distintos graus de potencial no que diz respeito à administração do saneamento básico nos municípios sob análise. Conforme estabelecido, o Cluster 1 representa um elevado nível de aptidão na gestão, o Cluster 2 sinaliza um potencial intermediário, enquanto o Cluster 3 está vinculado a um potencial de gestão reduzido. Uma parte desse resultado é apresentada no quadro a seguir:

Quadro 6 – Resultado da clusterização via K-médias nas 4 primeiras observações

Município	Região	PO001	PO028	PO042	PO048	QD001	Rating	Cluster
Brasília	Centro-Oeste	1	1	0	1	2	5	2
Abadia de Goiás	Centro-Oeste	0	1	0	1	1	3	3
Abadiânia	Centro-Oeste	0	0	0	0	0	0	3
Acreúna	Centro-Oeste	0	0	0	0	1	1	3

Fonte: Elaborado pelo autor.

A partir da planilha de resultados disponível, procedemos à divisão desses diferentes níveis de administração por região, o que culminou na formulação de uma matriz que representa o potencial de gerenciamento do saneamento básico para os municípios brasileiros, estruturados de acordo com suas respectivas regiões geográficas.

Para realizar essa fase, realizaram-se os seguintes procedimentos:

- Organização geográfica dos dados – inicialmente, os dados são estruturados de acordo com as diversas regiões do Brasil, facilitando a formação de uma matriz consolidando as informações sobre a competência na gestão de saneamento básico em cada região;
- Agrupamento com base em *clusters* – em seguida, os resultados da

clusterização são empregados para classificar cada município em um dos *clusters* identificados (1, 2 ou 3), com base na sua capacidade de gestão;

- Matriz de avaliação do potencial de gestão do saneamento básico – com os municípios categorizados por região e clusters, foi possível criar uma matriz na qual as linhas representariam as distintas regiões do Brasil, e as colunas corresponderiam aos *clusters* que indicam o potencial de gestão. A matriz conteria informações sobre quantos municípios de cada região se encontrariam em cada um dos *clusters*.

Mediante a criação dessa matriz, tornou-se possível visualizar a distribuição dos municípios por região e os respectivos níveis de habilidade na gestão do saneamento básico. Isso habilitou uma análise abrangente das dinâmicas regionais relacionadas à gestão do saneamento.

Figura 1 – Matriz de Avaliação do Potencial de Gestão do Saneamento Básico

Norte	9	137	235
Nordeste	48	323	1056
Centro - Oeste	17	120	268
Sudeste	52	552	946
Sul	37	566	534
	Alto Potencial	Médio Potencial	Baixo Potencial

Fonte: Elaborada pelo autor.

A partir dos dados presentes na matriz de potencial de gestão, tornou-se factível identificar regiões que demandam melhorias específicas e, assim,

orientar estratégias mais eficazes para aprimorar a gestão do saneamento em diferentes partes do Brasil. Essa análise inicial das amostras a serem comparadas fornece uma visão estratégica e abrangente da capacidade de gerenciamento do saneamento básico em âmbito regional, possibilitando uma distribuição mais precisa de recursos e a implementação de ações direcionadas para o desenvolvimento das práticas de gestão.

A clusterização permitiu verificarmos que em todas as regiões há problemas quanto ao potencial de gestão de saneamento pelo municípios. A tabela pôde orientar o potencial de gestão e orientar o poder público onde estão os municípios com piores resultados e aplicar medidas para melhorar seu desempenho no saneamento básico.

Com os resultados obtidos através da clusterização, tornou-se viável compreender os conjuntos formados e dar sentido a eles; nesse contexto, os três *clusters* representam distintos patamares de capacidade na administração do saneamento básico nos municípios. É possível notar que todas as regiões possuem a maioria dos municípios associados a um baixo potencial, com exceção da região Sul. A região com pior avaliação é a região Nordeste, com 74% dos seus municípios em uma região de baixo potencial enquanto a região Sul tem melhores resultados com 47% dos municípios em uma região de baixo potencial.

Para alcançar melhores resultados os municípios podem adotar Política de Saneamento Básico conforme a Lei nº 11.445/2007, possuir plano municipal de saneamento básico, elaborado nos termos estabelecidos na Lei 11.445/2007, ser integrante de algum consórcio intermunicipal regulamentado pela Lei nº 11.107/2005, possuir Plano Municipal de Gestão Integrada de Resíduos Sólidos (PMGIRS) conforme a Lei nº 12.305/2010 e que as diretrizes e regulamentações estabelecidas em uma portaria relacionada à qualidade da água sejam atendidas integralmente.

4.2 Propensity Score Matching

Antes da regressão linear, foi feito um tratamento no grupo de controle e de tratamento através do PSM. Havia 1731 observações no grupo de

controle e 3169 no grupo de tratamento, mas, após o PSM, restaram 1731 observações no grupo de controle e 1731 no grupo de tratamento, excluindo as 1438 observações de tratamento com menor *propensity score*.

É importante destacar a transformação realizada pelo PSM no tamanho dos grupos de controle e tratamento. Antes do PSM, havia um desequilíbrio significativo entre o número de observações nos dois grupos, o que poderia levar a vieses nos resultados das regressões devido à falta de comparabilidade entre esses grupos.

Figura 2 – Quantidade de observações de tratamento e controle

Sample Sizes:		
	Control	Treated
All	1731	3169
Matched	1731	1731
Unmatched	0	1438
Discarded	0	0

Fonte: Elaborada pelo autor.

Essa metodologia nos permite avaliar de forma causal o efeito da implementação do PMSB nos indicadores definidos na seção 3.2, pois, ao mitigar possíveis vieses de seleção, adquirem-se evidências mais robustas em relação à eficácia do PMSB nos municípios. A uniformidade nos tamanhos de amostra nos grupos de controle e tratamento indica que o processo de pareamento alcançou sucesso na formação de grupos equiparáveis. Para cada elemento no grupo de controle há um elemento no grupo de tratamento associado a ele. Com resultado do PSM, são definidos os grupos de controle e de tratamento que serão utilizados na regressão linear.

4.3 Regressão linear

4.3.1 Índice de Atendimento Total de Água (IN055_AE)

O comportamento do Índice de Atendimento Total de Água na regressão linear é apresentado a seguir:

Quadro 7 – Resultado da regressão linear do Índice de Atendimento Total de Água

Variáveis Explicativas	Coefficiente	P Valor
PO028	$4,49 * 10$	$2 * 10^{-16}$
FN023	$8,5 * 10^{-7}$	$3,64 * 10^{-12}$
IN005_AE	4,67	$2 * 10^{-16}$

Fonte: Elaborado pelo autor.

A equação que representa o Índice de Atendimento Total de Água (IN055_AE) é:

$$IN055_AE = 44,87 * PO028 + 8,5 * 10^{-0,7} * FN023 + 4,67 * IN005_AE$$

O modelo de regressão mostra que todas as variáveis independentes (PO028, FN023 e IN005_AE) são significativas para explicar o índice de atendimento de água.

O coeficiente para PO028 é positivo e significativo (p-valor < 0,05), o que sugere que a presença de um Plano Municipal de Saneamento Básico (conforme a Lei nº 11.445/2007) está associada a um aumento no índice de atendimento de água.

O coeficiente para FN023 também é positivo e significativo (p-valor < 0,05), o que indica que maiores investimentos em serviços de saneamento básico estão relacionados a um maior índice de atendimento de água.

O coeficiente para IN005_AE também é positivo e significativo (p-valor < 0,05), o que sugere que uma tarifa média mais alta de água está associada a um aumento no índice de atendimento de água.

Quadro 8 – Estatísticas globais da regressão linear do Índice de Atendimento Total de Água

Estatísticas globais	Resultado
Multiple R-squared	0,6436
Adjusted R-squared	0,6432
F-statistic	1729
P Valor	$2,2 * 10^{-16}$

Fonte: Elaborado pelo autor.

- Multiple R-squared: O valor é 0,6436. Quanto mais próximo a 1, significa que as variáveis independentes do modelo explicam muito bem a variabilidade da variável dependente, o que sugere que o modelo é eficaz na explicação dos dados.
- Adjusted R-squared: O valor é 0,6432. É uma variação do R-quadrado múltiplo (Multiple R-squared), que leva em consideração o número de variáveis independentes no modelo de regressão. Enquanto o R-quadrado múltiplo simplesmente mede o quão bem as variáveis independentes explicam a variabilidade da variável dependente, o R-quadrado ajustado penaliza o modelo quando inclui variáveis independentes desnecessárias. A principal diferença entre o R-quadrado múltiplo e o R-quadrado ajustado é que o R-quadrado ajustado tende a diminuir à medida que mais variáveis independentes são adicionadas ao modelo, a menos que essas variáveis adicionais realmente melhorem a capacidade do modelo de explicar os dados. Isso o torna uma métrica mais robusta para avaliar a qualidade do modelo, especialmente quando você está considerando modelos com diferentes números de variáveis independentes.
- F-statistic: O valor é 1729. É uma medida usada em análise de regressão para avaliar a significância global de um modelo de regressão. Um valor alto indica que pelo menos uma das variáveis independentes é significativa.
- P Valor: O valor é muito baixo (p -valor $< 0,05$), o que indica que o modelo é globalmente significativo. O valor- p associado à estatística F é usado para determinar a significância estatística do modelo. Um valor- p baixo (geralmente menor que 0,05) indica que o modelo é estatisticamente significativo.

Os resultados desta regressão indicam um bom ajuste do modelo aos dados. O R-squared ajustado de 0,6432 sugere que cerca de 64,32% da variabilidade no IN055_AE é explicada pelas variáveis independentes. O baixo p -valor ($2,2 * 10^{-16}$) indica que pelo menos uma das variáveis independentes é estatisticamente significativa na explicação da variabilidade no IN055_AE. Indica que o modelo é estatisticamente significativo.

4.3.2 Índice de Atendimento Total de Esgoto Referido aos Municípios Atendidos com Água (IN056_AE):

O comportamento do Índice de Atendimento Total de Esgoto Referido aos Municípios Atendidos com Água na regressão linear é apresentado a seguir:

Quadro 9 – Resultado da regressão linear do Índice de Atendimento Total de Esgoto Referido aos Municípios Atendidos com Água

Variáveis Explicativas	Coefficiente	P Valor
PO028	$5,00 * 10$	$2 * 10^{-16}$
FN024	$3,51 * 10^{-7}$	$2,95 * 10^{-3}$
IN006_AE	2,15	$2 * 10^{-16}$

Fonte: Elaborado pelo autor.

A equação que representa o Índice de Atendimento Total de Esgoto Referido aos Municípios Atendidos com Água (IN056_AE) é:

$$IN056_AE = 50,02 * PO028 + 3,508 * 10^{-07} * FN024 + 2,147 * IN006_AE$$

O modelo de regressão mostra que todas as variáveis independentes (PO028, FN024 e IN006_AE) são significativas para explicar o índice de atendimento de esgoto.

O coeficiente para PO028 é positivo e significativo (p-valor < 0,05), o que indica que a presença de um Plano Municipal de Saneamento Básico está associada a um aumento no índice de atendimento de esgoto.

O coeficiente para FN024 também é positivo e significativo (p-valor < 0,05). Isso sugere que investimentos adicionais com esgotamento sanitário estão associados a um aumento no índice de atendimento de esgoto.

O coeficiente para IN006_AE é positivo e significativo (p-valor < 0,05), indicando que uma tarifa média mais alta de esgoto está relacionada a um aumento no índice de atendimento de esgoto.

Quadro 10 – Estatísticas globais da regressão linear do Índice de Atendimento Total de Esgoto Referido aos Municípios Atendidos com Água

Estatísticas globais	Resultado
Multiple R-squared	0,5477
Adjusted R-squared	0,5465
F-statistic	460,5
P Valor	$2,2 * 10^{-16}$

Fonte: Elaborado pelo autor.

Esta regressão também mostra um bom ajuste do modelo aos dados. O R-squared ajustado de 0.5465 sugere que cerca de 54.65% da variabilidade no IN056_AE é explicada pelas variáveis independentes. O baixo p-valor ($2,2 * 10^{-16}$) indica que pelo menos uma das variáveis independentes é estatisticamente significativa na explicação da variabilidade no IN056_AE. Isso indica que o modelo é estatisticamente significativo.

4.3.3 Taxa de Cobertura Regular do Serviço de Coleta de RDO em Relação à População Total do Município (IN015_RS)

O comportamento da Taxa de Cobertura Regular do Serviço de Coleta de RDO em relação à População Total do Município na regressão linear é apresentado a seguir:

Quadro 11 – Resultado da regressão linear da Taxa de Cobertura Regular do Serviço de Coleta de RDO em Relação à População Total do Município

Variáveis explicativas	Coefficiente	P Valor
PO028	$5,05 * 10$	$2 * 10^{-16}$
IN023_RS	$8,21 * 10^{-2}$	$2 * 10^{-16}$
IN019_AE	7,14	$2 * 10^{-16}$

Fonte: Elaborado pelo autor.

A equação que representa o índice de atendimento total de água (IN015_RS) é:

$$IN015_RS = 50,47 * PO028 + 8,211 * 10^{-02} * IN023_RS + 2,147 * IN019_RS$$

O modelo de regressão mostra que todas as variáveis independentes (PO028, IN023_RS e IN019_RS) são significativas para explicar a Taxa de Cobertura do Serviço de Coleta de Resíduos Sólidos.

O coeficiente para PO028 é positivo e significativo (p-valor < 0,05), o que sugere que a presença de um Plano Municipal de Saneamento Básico está associada a uma maior Taxa de Cobertura do Serviço de Coleta de Resíduos Sólidos.

O coeficiente para IN023_RS (custo unitário médio do serviço de coleta) é positivo e significativo (p-valor < 0,05), o que indica que maiores custos unitários médios do serviço de coleta estão relacionados a uma taxa de cobertura mais alta.

O coeficiente para IN019_RS (taxa de empregados na coleta em relação à população urbana) também é positivo e significativo (p-valor < 0,05), sugerindo que uma maior taxa de empregados na coleta em relação à população urbana está associada a uma taxa de cobertura mais alta.

Quadro 12 – Estatísticas globais da regressão linear da Taxa de Cobertura Regular do Serviço de Coleta de RDO em Relação à População Total do Município

Estatísticas globais	Coefficiente
Multiple R-squared	0,6715
Adjusted R-squared	0,671
F-statistic	4,67
P-Valor	$2,2 * 10^{-16}$

Fonte: elaboração própria.

Os resultados desta regressão mostram um ajuste excelente do modelo aos dados. O R-squared ajustado de 0,671 indica que cerca de 67,10% da variabilidade no IN015_RS é explicada pelas variáveis independentes. O baixo p-valor (< 2,2e-16) indica que pelo menos uma das variáveis independentes é estatisticamente significativa na explicação da variabilidade no IN015_RS, além de também evidenciar que o modelo é

estatisticamente significativo.

O Plano Municipal de Saneamento Básico (PMSB) é uma política pública significativa e positivamente relacionada com todos os indicadores de desempenhos estudados neste trabalho (índice de atendimento total de água, índice de atendimento total de esgoto referido aos municípios atendidos com água e taxa de cobertura regular do serviço de coleta de RDO). Ou seja: estatisticamente, existe a relação positiva entre o PMSB e um melhor desempenho do saneamento básico, orientando que os municípios que ainda não possuem um PMSB podem alocar recursos para a elaboração e acompanhamento de um PMSB com o intuito de aumentar a qualidade dos serviços de saneamento básico prestado, melhorando as condições de sua população, e contribuir para um desenvolvimento mais sustentável.

5 CONCLUSÃO

Neste trabalho, foi utilizada a técnica de clusterização k-means para gerar uma matriz de potencial de gestão de saneamento básico, identificando grupos de municípios com características semelhantes em termos de gestão. Essa abordagem nos permitiu categorizar os municípios em *clusters* com diferentes níveis de eficiência na gestão de saneamento.

Exploramos a eficácia das políticas de saneamento básico nos municípios brasileiros, com foco em três indicadores-chave de desempenho: o Índice de Atendimento Total de Água, o Índice de Atendimento Total de Esgoto e a Taxa de Cobertura Regular do Serviço de Coleta de Resíduos Sólidos em relação à população total do município. Utilizando uma abordagem metodológica que incorporou o pareamento por escore de propensão (PSM), conseguimos obter evidências mais sólidas e robustas sobre o impacto das políticas de saneamento.

Uma das principais contribuições deste estudo foi a capacidade de controlar potenciais vieses de seleção, permitindo-nos avaliar de forma causal o impacto da presença do Plano Municipal de Saneamento Básico (PMSB) sobre os indicadores de interesse. Os resultados obtidos demonstram que a presença do PMSB está positivamente associada a melhorias significativas nos indicadores de atendimento de água e esgoto, bem como na taxa de cobertura regular de coleta de resíduos sólidos.

Como resultado, este estudo fornece uma base sólida para a formulação de políticas públicas relacionadas ao saneamento básico no Brasil. Destacamos a importância de políticas que promovam não apenas a expansão dos serviços, mas também a eficiência na gestão e a qualidade na prestação dos serviços de água, esgoto e resíduos sólidos. Além disso, destacamos a utilidade da abordagem do PSM na avaliação de políticas públicas, especialmente quando se trata de variáveis dependentes complexas e sujeitas a vieses de seleção.

Este estudo, no entanto, não esgota o assunto, e há espaço para futuras pesquisas que explorem ainda mais os determinantes e os impactos das políticas de saneamento. No entanto, acredita-se que os resultados apresentados aqui oferecem uma contribuição valiosa para a compreensão

das dinâmicas do saneamento básico no Brasil e fornecem *insights* importantes para a melhoria contínua desses serviços essenciais à qualidade de vida da população.

REFERÊNCIAS

ARCENEUX, K.; GERBER, A. S.; GREEN, D. P. Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment. **Political Science & Politics**, v. 39, n. 4, p. 727-731, 2006.

ARCHANA, S.; YADAV, A.; RANA, A. K-means with Three different Distance Metrics. **International Journal of Computer Applications**, v. 67, n. 10, p. 13–17, abr. 2013.

BECKER, S. O.; ICHINO, A. Estimation of average treatment effects based on propensity scores. **The Stata Journal**, v. 2, n. 4, p. 358-377, 2002.

BRASIL. Lei nº 10.257, de 10 de julho de 2001. **Regulamenta os arts. 182 e 183 da Constituição Federal, estabelece diretrizes gerais da política urbana e dá outras providências**. Brasília, 2001, *on-line*. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/leis_2001/l10257.htm. Acesso em: 20 jul. 2023.

BRASIL. Fundação Nacional de Saúde. **Manual de Saneamento**. 3. ed. rev. Brasília: Fundação Nacional de Saúde, 2007a.

BRASIL. Lei nº 11.445, de 5 de janeiro de 2007. **Estabelece diretrizes nacionais para o saneamento básico; altera as Leis nos 6.766, de 19 de dezembro de 1979, 8.036, de 11 de maio de 1990, 8.666, de 21 de junho de 1993, 8.987, de 13 de fevereiro de 1995; revoga a Lei no 6.528, de 11 de maio de 1978; e dá outras providências**. Brasília, 2007b, *on-line*. Disponível em: <http://www.planalto.gov.br/ccivil03/ato20072010/2007/lei/L11445compilado.htm>. Acesso em: 20 jul. 2023.

BRASIL. Lei nº 12.305, de 2 de agosto de 2010. **Institui a Política Nacional de Resíduos Sólidos; altera a Lei nº 9.605, de 12 de fevereiro de 1998; e dá outras providências**. Brasília, 2010, *on-line*. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2010/lei/l12305.htm. Acesso em: 20 jul. 2023.

CALIENDO, M.; KOPEINIG, S. Some practical guidance for the implementation of propensity score matching. **Journal of Economic Surveys**, v. 22, n. 1, p. 31-72, 2008.

CAMERON, A. Colin; TRIVEDI, Pravin K. **Microeconometrics: Methods and Applications**. Cambridge University Press, 2005.

CAMPELLO, M.; GRAHAM, J.; HARVERY, C. (2010). The real effects of financial constraints: Evidence from a financial crisis. **Journal of Financial Economics**, v. 97, p. 470-487, 2010. Disponível em: <https://doi.org/10.1016/j.jfineco.2010.02.009>. Acesso em: 14 jul. 2023.

CHAKRABARTI, A.; GHOSH, J. K. AIC, BIC and recent advances in model

selection. **Philosophy of statistics**, p. 583-605, 2011.

CHRISTENSEN, T.; COSSU, R.; STEGMANN, R. (Eds.) **Sanitary Landfilling: Process, Technology and Environmental Impact**. Londres: Academic Press, 2012.

DEHEJIA, R. H.; WAHBA, S. Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. **Journal of the American Statistical Association**, v. 94, p. 1053-1062, 1999.

DONI, M. V. **Análise de cluster: métodos hierárquicos e de particionamento**. São Paulo, 2004.

FRASER, G. E.; STRAM, D. O. Regression calibration in studies with correlated variables measured with error. **American Journal of Epidemiology**, v. 154, p. 836–844, 2023.

FONSECA, J. S.; MARTINS, G. A.; TOLEDO, G. L. **Estatística aplicada**. 2. ed. São Paulo: Atlas, 1991.

GANGL, M. Scar Effects of Unemployment: An assessment of institutional complementarities. **American Sociological Review**, v. 71, n. 6, p. 986–1013, 2006.

GERHARDT, T. E.; SILVEIRA, D. T. **Métodos de pesquisa**. Plageder, 2009.

GIL, A. C. Como elaborar projetos de pesquisa. São Paulo: Atlas, 2002.

GRODSKY, E. Compensatory sponsorship in higher education. **American Journal of Sociology**, v. 112, n. 6, p. 1662–1712, 2007.

GUO, S.; FRASER, M. W. **Propensity score analysis: Statistical methods and applications**. Thousand Oaks: SAGE Publications, 2014.

GUJARATI, D. N.; PORTER, D. C. **Econometria**. 5. ed. Porto Alegre: AMGH, 2011.

HAIR JR., J. F. *et al.* **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009.

HIRANO, K.; IMBENS, G. W. **The Propensity Score with Continuous Treatments**. Hoboken: John Wiley & Sons, 2005.

JAIN, A. K. Data Clustering: 50 Years Beyond K-Means. **Pattern Recognition Letters**, v. 1, n. 8, p. 651–666, 2010.

JAMES, D. E.; THROSBY, D. C. **Métodos quantitativos aplicados à economia**. Tradução de Carlos Roberto Vieira Araújo, São Paulo: Atlas, 1977.

KAUFMANN, L.; ROUSSEEUW, P. J. **Finding groups in data: an introduction to cluster analysis**. Nova York: Wiley, 1990.

KHANDKER, S. R. *et al.* **Handbook on Impact Evaluation: Quantitative Methods and Practices**. Washington, D.C.: The World Bank, 2010.

KMENTA, J. **Elementos de econometria**. Tradução de Carlos Roberto Vieira de Araújo. São Paulo: Atlas, 1978.

LEITE, W. **Practical propensity score methods using R**. Thousand Oaks: SAGE Publications, 2016.

LEITE, W.; AYDIN, B. (Eds.). **A comparison of methods for imputation of missing covariate data prior to propensity score analysis**. Washington, DC: American Education Research Association Conference, 2016.

LINDEN, R. Técnicas de Agrupamento. **Revista de Ssistemas de Informação da FSMA**, n. 4, p 18–36, 2009.

MATTAR, F. N. **Pesquisa de marketing: metodologia, planejamento, execução e análise**, 2. ed. São Paulo: Atlas, 1994, 2v., v.2.

NAÇÕES UNIDAS. Mais de 4,2 bilhões de pessoas vivem sem acesso a saneamento básico. **Nações Unidas**, 2020, *on-line*. Disponível em: <https://brasil.un.org/pt-br/101526-mais-de-42-bilhoes-de-pessoas-vivem-sem-acesso-saneamento-basico>. Acesso em: 12 set. 2023.

OLIVEIRA, T. H. **Sistema de drenagem urbana sustentável como estratégia para resiliência aos impactos das mudanças climáticas**. 2021. Dissertação (Mestrado em Ciências: Engenharia Hidráulica e Saneamento) – Universidade de São Paulo, São Carlos, 2021. Disponível em: <https://www.teses.usp.br/teses/disponiveis/18/18138/tde-02122021-094412/>. Acesso em: 08 out. 2023.

PRODANOV, C. C.; FREITAS, E. C. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2. ed. Novo Hamburgo: Editora Feevale, 2013.

R FOUNDATION FOR STATISTICAL COMPUTING. The R project for statistical computing. **R-Project**, *on-line*. Disponível em: <http://www.r-project.org>. Acesso em: 03 fev. 2023.

REBOUÇAS, A. C.; BRAGA, B.; TUNDISI, J. G. **Águas Doces no Brasil: Capital Ecológico, Uso e Conservação**. São Paulo: Escrituras, 2004.

ROSENBAUM, P.; RUBIN, D. The Central Role of the Propensity Score in Observational Studies for Causal Effects. **Biometrika**, v. 70, n. 1, p. 41-55, 1983.

SAIANI, C. C. S. **Restrições à expansão dos investimentos em**

saneamento básico no Brasil: déficit de acesso e desempenho dos prestadores. 2007. 315 f. Dissertação (mestrado em economia aplicada) — Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2007.

SENADO FEDERAL. Estudo aponta que falta de saneamento prejudica mais de 130 milhões de brasileiros. **Senado**, 2022, *on-line*. Disponível em: <https://www12.senado.leg.br/noticias/infomaterias/2022/03/estudo-aponta-que-falta-de-saneamento-prejudica-mais-de-130-milhoes-de-brasileiros>. Acesso em: 12 set. 2023.

SOUZA, C. M. N. S.; COSTA, A. M.; MORAES, L. R. S.; FREITAS, C. M. **Saneamento: promoção da saúde, qualidade de vida e sustentabilidade ambiental.** Rio de Janeiro: Fiocruz, 2015.

SNIS (SISTEMA NACIONAL DE INFORMAÇÕES EM SANEAMENTO). Disponível em: www.snis.gov.br. Acesso em: 13 maio 2023.

TAN, P.; STEINBACH, M.; KUMAR, V. **Introduction to Data Mining.** Boston: Addison–Wesley, 2006.

TUNDISI, J. G. **Água no século XXI: enfrentamento a escassez.** São Carlos: RiMa, IIE, 2009.

WOLFE, F.; MICHAUD, K. Heart failure in rheumatoid arthritis: rates, predictors, and the effect of anti-tumor necrosis factor therapy. **The American Journal of Medicine**, v. 116, n. 5, p. 305–11, 2004.

WONG, Y. D.; SUN, D. D.; LAI, D. Value-added utilisation of recycled concrete in hot-mix asphalt. **Waste Management**, v. 27, n. 2, p. 294–301, 2007.

ZAIANE, O.; OLIVEIRA, S. **Geometric data transformation for privacy preserving Clustering.** Edmonton, Alberta, Canada, 2003.