

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

VICENTE HENRIQUE DA COSTA SILVA

ANÁLISE DA RECUPERAÇÃO DE MATERIAIS RECICLÁVEIS: UMA APLICAÇÃO
DO PROPENSITY SCORE MATCHING NA BASE DE DADOS DO SNIS

Recife, 2023

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

**ANÁLISE DA RECUPERAÇÃO DE MATERIAIS RECICLÁVEIS: UMA APLICAÇÃO
DO PROPENSITY SCORE MATCHING NA BASE DE DADOS DO SNIS**

VICENTE HENRIQUE DA COSTA SILVA

Monografia apresentada à coordenação do curso de Bacharelado em Ciências Econômicas, sob a orientação da professora Chiara Natércia França Araújo, submetido à Universidade Federal Rural de Pernambuco como parte dos requisitos para a obtenção do título de Bacharel em Ciências Econômicas, de acordo com as exigências.

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)

S586a Silva, Vicente Henrique da Costa Silva
ANÁLISE DA RECUPERAÇÃO DE MATERIAIS RECICLÁVEIS: UMA APLICAÇÃO DO PROPENSITY
SCORE MATCHING NA BASE DE DADOS DO SNIS / Vicente Henrique da Costa Silva Silva. - 2023.
55 f. : il.

Orientadora: Chiara Natercia Franca Araujo.
Inclui referências e anexo(s).

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

1. PSM. 2. Clusterização. 3. SNIS. I. Araujo, Chiara Natercia Franca, orient. II. Título

CDD 330

VICENTE HENRIQUE DA COSTA SILVA

**ANÁLISE DA RECUPERAÇÃO DE MATERIAIS RECICLÁVEIS: UMA
APLICAÇÃO DO PROPENSITY SCORE MATCHING NA BASE
DE DADOS DO SNIS**

Monografia apresentada à coordenação do curso de Bacharelado em Ciências Econômicas, submetido à Universidade Federal Rural de Pernambuco como parte dos requisitos 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 (Orientadora)
Universidade Federal Rural de Pernambuco

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

Prof^ª. Dr^ª. GISLÉIA BENINI DUARTE (Examinadora Interna)
Universidade Federal de Pernambuco

Recife, 2023

Dedico este trabalho aos meus amados pais,
cuja influência e apoio foram essenciais para minha jornada acadêmica.

Ao meu pai, por ser um exemplo de mente inquisitiva.

À minha mãe,
agradeço por me ensinar a importância do esforço e da dedicação nos estudos.

Dedico também a Gabriela
por todo o suporte durante a necessária transformação.

AGRADECIMENTOS

Dedico este trabalho a todos que contribuíram para sua realização e para a conquista deste objetivo. Agradeço a minha família pelo amor e apoio incondicionais, aos amigos e colegas de estudo pelo encorajamento e suporte, e aos orientadores pela orientação e expertise. Expresso minha gratidão aos participantes da pesquisa, cuja colaboração foi fundamental. Este trabalho é resultado do esforço coletivo e do apoio de cada um de vocês. Que os resultados obtidos possam contribuir para avanços na área e para a busca de soluções que melhorem os serviços de saneamento.

"E ideias são à prova de balas."

Alan Moore

RESUMO

Este trabalho de conclusão de curso visa avaliar o impacto da implementação de Planos Municipais de Gestão Integrada de Resíduos Sólidos (PMGIRS), conforme estipulado na Lei nº 12.305/2010 - Política Nacional de Resíduos Sólidos. A abordagem metodológica central utiliza o conceito de *Propensity Score Matching* (PSM) para explorar o efeito causal da presença do PMGIRS sobre a Taxa de Recuperação de Materiais Recicláveis. Além disso, a pesquisa incorpora o uso da clusterização via k-médias com o objetivo de melhor caracterizar e classificar a amostra analisada. Isso permite identificar agrupamentos naturais dos dados e aprofundar a compreensão dos diferentes potenciais de gestão de resíduos sólidos em municípios distintos. O PSM é empregado para criar grupos comparáveis de municípios com e sem PMGIRS, equilibrando as características observáveis e controlando possíveis vieses de seleção. Os resultados obtidos a partir dessa metodologia contribuirão para uma análise embasada e objetiva do impacto dos PMGIRS na gestão municipal de resíduos sólidos. A combinação do PSM com a clusterização enriquecerá a análise, permitindo uma caracterização mais completa e uma classificação mais precisa da amostra estudada.

Palavras-chave: PMGIRS, *Propensity Score Matching*, Clusterização, K-médias.

ABSTRACT

This undergraduate thesis aims to assess the impact of the implementation of Municipal Integrated Solid Waste Management Plans (PMGIRS), as stipulated in Law No. 12,305/2010 - the National Solid Waste Policy. The central methodological approach utilizes the concept of Propensity Score Matching (PSM) to explore the causal effect of the presence of PMGIRS on the Recycling Material Recovery Rate. Additionally, the research incorporates the use of k-means clustering with the aim of better characterizing and classifying the analyzed sample. This allows for the identification of natural data clusters and a deeper understanding of different potentials for solid waste management in distinct municipalities. PSM is employed to create comparable groups of municipalities with and without PMGIRS, balancing observable characteristics and controlling for potential selection biases. The results obtained from this methodology will contribute to a well-founded and objective analysis of the impact of PMGIRS on municipal solid waste management. Combining PSM with clustering will enhance the analysis, enabling a more comprehensive characterization and precise classification of the studied sample.

Keywords: PMGIRS, Propensity Score Matching, Clusterization, K-means.

LISTAS DE FIGURAS

Figura 1	Matriz de Aderência à Gestão de Resíduos Sólidos dos municípios por região.....	37
-----------------	---	----

LISTAS DE TABELAS

Tabela 1	Resultado do pareamento pós PSM.....	37
Tabela 2	Resultados dos Sample Sizes em RStudio.....	41
Tabela 3	Resultados da Regressão.....	43

LISTA DE ABREVIATURAS E SIGLAS

PMGIRS	Planos Municipais de Gestão Integrada de Resíduos Sólidos
SNIS	Sistema Nacional de Informações de Saneamento
PSM	<i>Propensity Score Matching</i>
RSU	Resíduos Sólidos Urbanos
PNRS	Política Nacional de Resíduos Sólidos
PIB	Produto Interno Bruto
LNSB	Lei de Diretrizes Nacionais para o Saneamento Básico
PMSB	Plano Municipal de Saneamento Básico
VRC	Critério da Razão de Variância
MQO	Mínimos Quadrados Ordinário
IBGE	Instituto Brasileiro de Geografia e Estatística
RCC	Resíduos da Construção Civil
RDO	Resíduos Domiciliares Ordinários
RPU	Resíduos de Pequenos Estabelecimentos e Públicos

SUMÁRIO

1	INTRODUÇÃO.....	13
2	REVISÃO BIBLIOGRÁFICA.....	15
2.1	LEI Nº 12.305/2010 E OS PLANOS MUNICIPAIS DE GESTÃO INTEGRADA E RESÍDUOS SÓLIDOS (PMGIRS).....	15
3	METODOLOGIA.....	18
3.1	BASE DE DADOS.....	18
3.2	CLUSTERIZAÇÃO VIA K-MÉDIAS.....	24
3.3	<i>PROPENSITY SCORE MATCHING</i>	29
3.4	MÉTODO DOS MÍNIMOS QUADRADOS ORDINÁRIOS.....	32
4	APRESENTAÇÃO E ANÁLISE DOS RESULTADOS.....	35
4.1	ANÁLISE DOS RESULTADOS E CONSTRUÇÃO DA MATRIZ DE ADERÊNCIA À GESTÃO DE RESÍDUOS DE SÓLIDOS.....	35
4.2	APLICAÇÃO DO <i>PROPENSITY SCORE MATCHING</i>	38
4.3	REGRESSÃO LINEAR COM DADOS DO <i>MATCHING</i>	41
5	CONCLUSÃO.....	46
	REFERÊNCIAS.....	48
	ANEXOS.....	53

1 INTRODUÇÃO

A gestão adequada dos resíduos sólidos é um desafio premente na sociedade contemporânea, impulsionada pela necessidade de abordar questões ambientais, de saúde pública e sustentabilidade. A implementação de Planos Municipais de Gestão Integrada de Resíduos Sólidos (PMGIRS), conforme regulamentado pela Lei nº 12.305/2010 - Política Nacional de Resíduos Sólidos (Brasil, 2010a), representa um marco significativo nesse cenário, visando melhorar a qualidade da gestão dos resíduos em âmbito local.

A gestão sustentável de resíduos é, atualmente, um dos desafios mais significativos enfrentados, sobretudo em nações em desenvolvimento, devido ao crescimento na produção de Resíduos Sólidos Urbanos (RSU) (Halkos; Petrou, 2019). Esse aumento tem sido impulsionado nos últimos anos pelo crescimento da população e das áreas urbanas, juntamente com mudanças nos hábitos de consumo (Alzamora; Barros, 2020). Para ilustrar, entre 2013 e 2016, houve um aumento de aproximadamente 50% na geração de resíduos, totalizando 2,01 bilhões de toneladas, as projeções indicam que esse número pode chegar a cerca de 2,59 bilhões até 2030 (Hornweg; Bhada-Tata, 2012; Kaza *et al.*, 2018). No entanto, esse crescimento não foi acompanhado pela adequada infraestrutura de gerenciamento para lidar com o substancial aumento na quantidade de resíduos gerados (Chaves *et al.*, 2014), contribuindo então para uma maior complexidade e custos na gestão de resíduos (Conke, 2018).

No contexto brasileiro, dados obtidos do Sistema Nacional de Informações sobre Saneamento (SNIS) evidenciam que os gastos destinados à gestão de Resíduos Sólidos Urbanos (RSU) representaram aproximadamente 4,1% do total de despesas correntes dos municípios (SNIS, 2022), com potencial para ultrapassar 8% em cenários específicos, conforme destacado por Leal e Sampaio (2019). Dentro dessa parcela, entre 60% e 70% decorrem dos custos associados à coleta e ao transporte (Colvero *et al.*, 2020; D'Onza *et al.*, 2016). Adicionalmente, os encargos significativos relacionados ao descarte apropriado dos resíduos em aterros sanitários contribuem substancialmente para as despesas municipais (Dutra *et al.*, 2018). Segundo Kaza *et al.* (2018), os custos das fases de coleta e transporte variam entre

20 e 200 dólares por tonelada, enquanto a disposição final em aterro sanitário oscila entre 10 e 100 dólares por tonelada.

A partir disso, Vianna e Amaral (2014) abordam a importância da avaliação de políticas públicas como uma ferramenta fundamental para melhorar a eficiência, eficácia e efetividade das ações governamentais, bem como para aprimorar a capacidade de gestão do Estado destacando a necessidade de fortalecer essa atividade avaliativa como parte integral da gestão pública.

Nesse sentido, outros estudos na área utilizaram a ferramenta do *Propensity Score Matching* (PSM), como o conduzido por Dias *et al.* (2023), quem avaliou os impactos da Política Nacional de Resíduos Sólidos (PNRS) no Brasil, a qual foi sancionada em 2010, ao utilizar métodos estatísticos como o PSM, analisando diversas variáveis relacionadas à gestão de resíduos sólidos.

Além do emprego do PSM, é importante notar que a metodologia de clusterização, utilizando o algoritmo K-médias, também tem sido aplicada em estudos relevantes, como visto em Teixeira *et al.* (2023) tal estudo analisou a importância das variáveis socioeconômicas na definição de áreas prioritárias de investimento em saneamento, utilizando o município de Niterói como estudo de caso por meio do método de agrupamento k-médias, realizando uma classificação multivariada para identificar estratos socioeconômicos distintos na área de estudo, resultando na discriminação de seis classes socioeconômicas diferentes.

Nesse sentido, o presente trabalho de conclusão de curso se propõe a avaliar o impacto da implementação dos PMGIRS na taxa de recolhimento de matéria residual. Para a obtenção dos dados utilizados, foram extraídas informações diretamente da base de dados do Sistema Nacional de Informações de Saneamento (SNIS). Cabe ressaltar que apenas foram consideradas na análise as variáveis apresentadas na base de dados do SNIS.

A metodologia de PSM permite uma avaliação mais precisa do efeito causal da presença dos PMGIRS sobre o indicador chave: a Taxa de Recuperação de Materiais Recicláveis. Ao equilibrar características observáveis e controlar possíveis vieses de seleção, tal abordagem proporciona uma compreensão mais sólida do verdadeiro impacto dos PMGIRS na gestão de resíduos.

2 REVISÃO BIBLIOGRÁFICA

2.1 LEI Nº 12.305/2010 E OS PLANOS MUNICIPAIS DE GESTÃO INTEGRADA E RESÍDUOS SÓLIDOS (PMGIRS)

Os Resíduos Sólidos Urbanos (RSU) constituem uma mistura variada de elementos, incluindo componentes inertes, minerais e orgânicos, resultantes tanto das atividades humanas quanto da natureza. No entanto, o manejo eficaz desses resíduos tem representado um desafio persistente no território nacional, ao longo de muitos anos. Conseqüentemente, observa-se a persistência de métodos inadequados de tratamento e disposição dos RSU, como é o caso dos lixões e vazadouros a céu aberto. Esses cenários podem ocasionar questões de cunho sanitário, econômico e ambiental, conforme identificado pelas autoridades (Brasil, 2007a).

A eficácia do controle dos resíduos sólidos emerge como uma variável estratégica, determinante para a viabilidade do sistema, otimização de sua sustentabilidade econômica e redução dos passivos ambientais emergentes. Em virtude disso, bem como de outros fundamentos, a Política Nacional de Resíduos Sólidos (PNRS) estimula a formação de consórcios intermunicipais, visando a gestão de resíduos sólidos urbanos por meio de aterros sanitários. Assim, a viabilidade deste empreendimento exige uma análise criteriosa da relação custo-benefício, assegurando um equilíbrio entre a qualidade ambiental do sistema e um custo acessível para o município (Brandão; Silva, 2011).

A Política Nacional de Resíduos Sólidos (PNRS), ancorada no planejamento como ferramenta central, demanda que esse planejamento defina os parâmetros para a realização dos objetivos da PNRS, tanto em relação aos entes federativos quanto aos geradores de resíduos. Conforme apontado por Toni (2003, p. 64):

A falta de planejamento por parte do governo, aliada à ausência de sistemas de responsabilização e direção eficaz, tende a sucumbir ao imediatismo, envolvendo a rotina burocrática diária. Isso acarreta agendas permeadas por rituais e superficialidades das dinâmicas sociais. (Toni, 2003, p. 64)

A integração entre a Política Nacional de Resíduos Sólidos (PNRS) e a Lei de Diretrizes Nacionais para o Saneamento Básico (LNSB) se manifesta na esfera da gestão dos resíduos sólidos urbanos. Segundo Ribeiro (2012), um ponto comum entre ambas as leis é a ênfase no planejamento da gestão dos resíduos sólidos urbanos. Enquanto a LNSB estipula o Plano Municipal de Saneamento Básico (PMSB)

como elemento central para a definição de contratos de gerenciamento dos resíduos sólidos urbanos, a PNRS aborda o Plano de Gestão Integrada de Resíduos Sólidos como um instrumento para identificar os responsáveis pela gestão ambientalmente adequada de todos os resíduos gerados no país (exceto os radioativos). Isso permite que municípios e estados criem regras mais detalhadas sobre a gestão dos resíduos.

A Lei nº 12.305/2010 (Brasil, 2010a), no Artigo 19, §1º, introduz uma inovação ao estabelecer que o Plano Municipal de Gestão Integrada de Resíduos Sólidos (PMGIRS) possa ser incorporado ao PMSB desde que os requisitos mínimos sejam atendidos, conforme previsto no Artigo 19 da Lei nº 11.445/2007 (Brasil, 2007b). Isso implica que o escopo do planejamento pode se estender além dos resíduos sólidos urbanos.

Com relação ao Plano Nacional de Resíduos Sólidos, o Decreto nº 7.404 (BRASIL, 2010b), o qual regula a PNRS, estipula um horizonte de planejamento de 20 anos nos Artigos 46 e 47, com revisões quinquenais. Para os planos municipais, uma vez que não há um prazo especificado, é indicado seguir a orientação fornecida pela PNRS em relação ao plano nacional.

A elaboração dos Planos Municipais de Gestão Integrada de Resíduos Sólidos (PMGIRS) não é de caráter obrigatório, mas se torna um requisito para que os municípios possam receber recursos da União, ou por ela controlados, destinados a projetos e serviços relacionados à gestão de resíduos sólidos. Também possibilita a elegibilidade para incentivos ou financiamentos provenientes de entidades federais de crédito ou fomento direcionados a essa finalidade (Brasil, 2010a). Além disso, os PMGIRS representam uma ferramenta de planejamento altamente eficaz para as administrações municipais ao lidar com seus resíduos, permitindo uma abordagem holística dentro do contexto local e/ou regional. Isso possibilita uma análise multidimensional, assim como o monitoramento contínuo do desenvolvimento nesse âmbito.

A Política Nacional de Resíduos Sólidos (PNRS), estabelecida pela Lei nº 12.305 (BRASIL, 2010a), reconhece a coleta seletiva como um instrumento fundamental para promover a reciclagem de materiais. Para fortalecer esse processo, a legislação prioriza a alocação de recursos provenientes da União para os municípios

que optarem, através dos seus Planos Municipais de Gerenciamento Integrado de Resíduos Sólidos (PMGIRS), por incorporar a coleta seletiva com a participação ativa de cooperativas. Com essa abordagem, a política incentiva a coleta seletiva e a inclusão dos catadores de materiais recicláveis como componentes essenciais na estruturação dos modelos de gestão adotados pelos municípios.

A responsabilidade pela elaboração dos PMGIRS recai sobre as prefeituras, que podem optar por contratar empresas terceirizadas para esse fim ou desenvolvê-los individualmente. Além disso, há a opção de elaboração conjunta com outros municípios por meio de consórcios, contudo, é crucial envolver a participação pública nesse processo (Brasil, 2010a). Tais planos devem ser construídos de acordo com o conteúdo mínimo estabelecido pela Política Nacional de Resíduos Sólidos, composto por 19 itens listados no Artigo 19º da lei correspondente. Também é viável integrar os PMGIRS ao Plano de Saneamento Básico, abrangendo os planos relacionados à água, esgoto, drenagem urbana e resíduos sólidos, conforme delineado na Lei nº 11.445/2007 (Brasil, 2007b), assegurando a observância do conteúdo mínimo regulamentado pela legislação (Brasil, 2010a).

3 METODOLOGIA

3.1 BASE DE DADOS

No âmbito deste estudo, a base de dados utilizada foi obtida do Sistema Nacional de Informações sobre Saneamento (SNIS). A fim de acessar e obter o conjunto de dados, os procedimentos seguidos foram explicitados abaixo:

Na interface de acesso às séries históricas, a guia "Município" foi selecionada, seguida da opção "Agrupamento dinâmico de indicadores e informações por ano de referência".

As seleções seguintes foram realizadas:

- ✓ 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", assinala-se a opção "marcar todos" em "Famílias de Informações e Indicadores".

O campo de interesse "*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?*" foi transformado em uma variável *dummy*. Neste contexto, os municípios que não apresentaram respostas válidas para o referido campo foram excluídos da análise, resultando em uma amostra final composta por 4900 municípios, cujas respostas foram classificadas como "SIM" ou "NÃO" para o campo "PO048".

É importante mencionar que a variável *dummy* é utilizada para codificar uma característica binária, com valores "1" indicando a presença da característica e "0" representando a ausência dela.

Os seguintes campos do SNIS foram empregados para a clusterização:

- CS001 - Existência de coleta seletiva formalizada pela prefeitura no município?

A variável "CS001 - Existência de coleta seletiva formalizada pela prefeitura no município" no Sistema Nacional de Informações sobre Saneamento (SNIS) indica a presença de um sistema oficial de coleta seletiva de resíduos sólidos devidamente estabelecido e gerenciado pela administração municipal em um dado município brasileiro. Tal sistema é destinado à separação e coleta de materiais recicláveis, como papel, vidro, plástico e metal, tendo como objetivo de os desviar do fluxo de resíduos sólidos urbanos comuns (lixo doméstico). A formalização desse programa é essencial para promover a reciclagem, reduzir impactos ambientais, gerar empregos na indústria da reciclagem, criar conscientização na comunidade e atender aos requisitos legais estabelecidos pela Política Nacional de Resíduos Sólidos (PNRS) do Brasil.

- CO008 - Prestação de serviço de coleta noturna no município?

A variável "CO008 - Prestação de serviço de coleta noturna no município" denota a existência e execução de serviços de coleta de resíduos sólidos durante o período noturno em um determinado município. Estes serviços de coleta noturna envolvem a recolha de resíduos, como lixo doméstico, durante as horas da noite, após o horário comercial convencional. Essa variável é fundamental para avaliar a flexibilidade e a eficiência dos serviços de coleta de resíduos em relação à gestão do tráfego e às necessidades da comunidade local. A disponibilidade desses serviços pode variar de acordo com as políticas e a infraestrutura específicas de cada município.

- CC020 - Presença de empresas especializadas (caçambeiros) que realizam a coleta de Resíduos de Construção Civil (RCC) no município?

A variável "CC020 - Presença de empresas especializadas (caçambeiros) que realizam a coleta de Resíduos de Construção Civil (RCC) no município" se refere à existência de empresas especializadas, frequentemente chamadas de "caçambeiros", as quais oferecem serviços de coleta e transporte de Resíduos de Construção Civil (RCC) em um determinado município. Os RCC incluem materiais resultantes de demolições, construções e reformas, por exemplo concreto, tijolos, madeira e outros resíduos relacionados à construção civil. A presença dessas empresas especializa-

das indica a disponibilidade de serviços dedicados à gestão adequada desses resíduos, o que apresenta uma certa importância para o controle ambiental e a regulamentação do setor de construção civil em um município.

- CA004 - Presença de catadores de materiais recicláveis que trabalham de forma dispersa na cidade?

A variável "CA004 - Presença de catadores de materiais recicláveis que trabalham de forma dispersa na cidade" indica se há catadores de materiais recicláveis que atuam de maneira independente e sem organização formal em uma determinada cidade. Esses catadores geralmente coletam materiais recicláveis, como papel, vidro, plástico e metal, das ruas, caixas de coleta pública e locais de descarte para posterior venda ou reciclagem. A presença desse grupo de catadores dispersos se mostra relevante para a gestão de resíduos sólidos, uma vez que eles desempenham um papel na recuperação de materiais recicláveis e podem contribuir para a redução do desperdício e da poluição ambiental. No entanto, muitas vezes enfrentam condições de trabalho precárias e falta de apoio institucional, o que destaca a importância de políticas e iniciativas para apoiar e formalizar esse setor.

- GE202 - Existência de empresa com contrato de delegação para serviços de limpeza urbana no município?

A variável "GE202 - Existência de empresa com contrato de delegação para serviços de limpeza urbana no município" se refere à presença de uma empresa que detém um contrato formal com a administração municipal para realizar serviços de limpeza urbana na cidade. Tais serviços podem incluir a coleta de resíduos sólidos, a varrição de ruas, a limpeza de áreas públicas e outras atividades relacionadas à manutenção da limpeza e higiene urbana. A existência deste contrato indica a delegação de responsabilidades e obrigações para uma entidade privada ou pública especializada na gestão de resíduos e na limpeza da cidade em nome da prefeitura ou administração municipal. Esse tipo de contrato é comum em muitas cidades para melhorar a eficiência na prestação de serviços de limpeza urbana.

Para o processo de *Propensity Score Matching*, os seguintes campos do SNIS foram empregados:

- POP_TOT - População total do município
Obtida a partir do Instituto Brasileiro de Geografia e Estatística (IBGE).
- FN220 - Despesa total com serviços de manejo de RSU

A variável "FN220 - Despesa total com serviços de manejo de RSU" se refere ao valor total de despesas incorridas para a gestão de Resíduos Sólidos Urbanos (RSU) em um determinado contexto. Essas despesas podem incluir uma série de atividades relacionadas ao manejo de resíduos sólidos, como coleta, transporte, tratamento, disposição final adequada, programas de reciclagem, educação ambiental e manutenção de infraestrutura relacionada à gestão de resíduos. O valor indicado nessa variável representa o montante financeiro gasto pela entidade responsável, a qual pode ser uma prefeitura, administração municipal ou outra instituição, para realizar todas essas atividades relacionadas à gestão de RSU em um período específico. O acompanhamento dessa despesa é fundamental para avaliar o custo e a eficácia dos serviços de gestão de resíduos sólidos em uma determinada área.

- FN208 - Despesa total com o serviço de coleta de RDO e RPU

A variável "FN208 - Despesa total com o serviço de coleta de RDO (Resíduos Domiciliares Orgânicos) e RPU (Resíduos Domiciliares não Orgânicos)" representa o valor total de despesas financeiras associadas à execução dos serviços de coleta de resíduos domiciliares em uma área específica, os quais abrangem a coleta de resíduos gerados em domicílios, incluindo tanto os resíduos orgânicos quanto os não orgânicos. A variável abarca diversos custos relacionados à operação de coleta, como recursos humanos, equipamentos, manutenção, logística de rotas e combustível, entre outros, proporcionando uma medida abrangente do investimento financeiro necessário para assegurar a coleta eficiente de resíduos domiciliares na jurisdição considerada.

- FN206 - Despesas dos agentes públicos com o serviço de coleta de RDO e RPU

A variável "FN206 - Despesas dos agentes públicos com o serviço de coleta de RDO (Resíduos Domiciliares Orgânicos) e RPU (Resíduos Domiciliares não Orgânicos)" se refere aos gastos financeiros efetuados pelos órgãos ou agentes públicos, como prefeituras ou administrações municipais, no tocante à prestação dos serviços de coleta de resíduos domiciliares. Esses serviços envolvem a recolha de resíduos gerados nos lares, abrangendo tanto os resíduos orgânicos, como restos de alimentos, quanto os resíduos não orgânicos, como plástico, vidro e papel. A variável engloba todas as despesas associadas à operação dos serviços de coleta, incluindo custos de pessoal, aquisição e manutenção de equipamentos, gerenciamento de rotas, combustível e outros recursos necessários para assegurar a efetiva coleta de resíduos domiciliares. Ela fornece uma medida quantitativa dos gastos públicos relacionados a essa atividade específica de gestão de resíduos.

Por fim, a análise de impacto será realizada através do método dos mínimos quadrados ordinários (MQO), a variável dependente selecionada será:

- IN031_RS - Taxa de recuperação de materiais recicláveis (exceto matéria orgânica e rejeitos) em relação à quantidade total coletada (resíduos domésticos orgânicos + resíduos públicos urbanos)

A variável "IN031_RS - Taxa de recuperação de materiais recicláveis (exceto matéria orgânica e rejeitos) em relação à quantidade total coletada (resíduos domésticos orgânicos + resíduos públicos urbanos)" representa a proporção de materiais recicláveis, excluindo matéria orgânica e rejeitos, os quais são recuperados em relação à quantidade total de resíduos coletados, incluindo resíduos domésticos orgânicos e resíduos públicos urbanos. Essa taxa é um indicador-chave da eficiência do processo de coleta seletiva e reciclagem em uma determinada área geográfica. Quanto maior a taxa, maior a quantidade de materiais recicláveis desviados do fluxo de resíduos sólidos urbanos, o que contribui para a preservação de recursos naturais, redução de resíduos enviados para aterros sanitários e promoção da sustentabilidade ambiental.

As variáveis explicativas incluirão:

Variável *Dummy*:

- PO048: Presença ou ausência de Plano Municipal de Gestão Integrada de Resíduos Sólidos (PMGIRS) conforme a Lei nº 12.305/2010 que trata da Política Nacional de Resíduos

A variável "PO048: Presença ou ausência de Plano Municipal de Gestão Integrada de Resíduos Sólidos (PMGIRS) conforme a Lei nº 12.305/2010 que trata da Política Nacional de Resíduos" indica se um município possui ou não um Plano Municipal de Gestão Integrada de Resíduos Sólidos (PMGIRS) em conformidade com a Lei nº 12.305/2010, a qual estabelece a Política Nacional de Resíduos Sólidos no Brasil. O PMGIRS é um instrumento fundamental para orientar e regular a gestão de resíduos sólidos em âmbito municipal, delineando estratégias, metas e ações para a gestão adequada destes resíduos, incluindo a redução, a reciclagem, o tratamento e a disposição final ambientalmente adequada. A presença de um PMGIRS é um indicativo do comprometimento do município com a gestão sustentável dos resíduos sólidos e o cumprimento das diretrizes estabelecidas na legislação nacional.

Variáveis Quantitativas:

- IN023_RS: Custo unitário médio do serviço de coleta (resíduos domésticos orgânicos + resíduos públicos urbanos)

A variável "IN023_RS: Custo unitário médio do serviço de coleta (resíduos domésticos orgânicos + resíduos públicos urbanos)" é referenciada ao custo médio por unidade de serviço de coleta de resíduos, considerando tanto os resíduos domésticos orgânicos quanto os resíduos públicos urbanos. Esse custo é uma medida quantitativa, a qual representa o valor médio gasto para realizar a coleta de uma unidade de resíduo, como um saco de lixo ou um recipiente de resíduos, abrangendo todos os custos associados a esse serviço, como pessoal, veículos, manutenção, combustível e outros recursos. Essa variável é relevante para avaliar o custo opera-

cional da coleta de resíduos em uma determinada área e pode ser usada para análise de eficiência e planejamento orçamentário na gestão de resíduos sólidos.

- IN019_RS: Taxa de empregados (coletadores + motoristas) na coleta (resíduos domésticos orgânicos + resíduos públicos urbanos) em relação à população urbana

A variável "IN019_RS: Taxa de empregados (coletadores + motoristas) na coleta (resíduos domésticos orgânicos + resíduos públicos urbanos) em relação à população urbana" representa a proporção de trabalhadores empregados na coleta de resíduos, incluindo coletadores e motoristas, em relação à população urbana de uma determinada área geográfica. Essa taxa quantifica a força de trabalho dedicada à coleta de resíduos em relação ao tamanho da população urbana atendida por esse serviço. É uma medida que pode ser usada para avaliar a intensidade da mão de obra envolvida na coleta de resíduos em áreas urbanas e pode ser útil para planejar e dimensionar adequadamente a equipe de coleta em conformidade com a demanda da população.

- POP_TOT: População total do município.

3.2 CLUSTERIZAÇÃO VIA K-MÉDIAS

A análise de *clusters* é um dos principais métodos estatísticos empregados para agrupar dados com base em sua homogeneidade. Esse método é amplamente utilizado na formulação de classificações e indicadores, proporcionando uma maneira de identificar padrões e estruturas subjacentes nos dados. A análise de clusters permite a formulação de hipóteses sobre a distribuição interna dos dados, possibilitando uma análise exploratória das informações.

Para Mingoti (2007), a análise de grupos pode ser realizada de maneira hierárquica ou não-hierárquica. Assim, segundo o autor, procedimentos hierárquicos geram estruturas do tipo árvore, no qual cada elemento inicialmente constitui um grupo, e em cada etapa, os dois grupos mais próximos são combinados em um novo grupo. Já os métodos não-hierárquicos visam particionar um conjunto de n elemen-

tos em k grupos, atendendo a dois requisitos cruciais: coesão interna e isolamento dos grupos.

Entre os métodos não-hierárquicos, o método k -médias se destaca como um dos mais utilizados. Tal método, inicialmente descrito por MacQueen (1967), é caracterizado por fixar o número de grupos (k) e estabelecer um centro para cada grupo, cujo valor é a média das observações do grupo. A alocação de cada membro aos grupos é determinada de forma a minimizar uma função de erro (Hartigan; Wong, 1979).

De acordo com as pesquisas de Samoilenko e Osei-Bryson (2008), Johnson e Wichern (1998) e Witten e Frank (2005), a técnica de agrupamento não-hierárquica conhecida como k -médias é amplamente utilizada para dividir um conjunto de elementos em uma coleção de k grupos. O método k -médias busca otimizar a alocação de casos a grupos, minimizando a função de erro. Esse processo é iterativo e baseado nas distâncias entre os casos e os centroides dos grupos.

O cálculo das distâncias entre casos e centroides é frequentemente realizado usando a distância euclidiana¹. Para cada caso, o grupo cujo centroide está mais próximo é atribuído a esse caso. Em seguida, os centroides são atualizados, tomando a média das observações em cada grupo. O processo de atribuição e atualização de centroides é repetido até que os centroides não mudem significativamente ou um critério de convergência seja alcançado.

O processo envolve as seguintes etapas detalhadas:

- Inicialização: Inicialmente, k centroides são selecionados como os centros dos k grupos;
- Atribuição: Cada elemento é alocado ao grupo cujo centroide está mais próximo, com base na distância euclidiana entre o elemento e os centroides;
- Atualização dos Centroides: Um novo centroide é calculado para cada grupo, utilizando as médias das coordenadas dos elementos alocados nesse grupo;

¹ A distância euclidiana é uma métrica amplamente utilizada na análise de dados multivariados para medir a distância entre dois pontos em um espaço euclidiano. Ela é calculada como a raiz quadrada da soma dos quadrados das diferenças entre as coordenadas dos pontos. A distância euclidiana é fundamental em diversas áreas, incluindo estatística, aprendizado de máquina e reconhecimento de padrões.

- Realocação: Os elementos são novamente atribuídos aos grupos cujos centróides estão mais próximos, com base nas novas posições dos centróides;
- Iteração: As etapas de atualização de centróides e realocação são repetidas até que não ocorram mais mudanças nas atribuições dos elementos aos grupos.

A determinação do número ideal de grupos, k , é crucial no método k -médias. Vários procedimentos são empregados para definir esse valor de forma eficaz e avaliar a qualidade dos agrupamentos resultantes. De acordo com Milligan e Cooper (1985) e Mooi e Sarstedt (2011), um método amplamente utilizado é o critério da razão de variância (VRC). Esse critério envolve a análise da relação entre a variância intra-grupo (variabilidade dos elementos dentro de cada grupo) e a variância total dos dados. Em termos de fórmulas:

$$VRC = (SSB / (k - 1)) / (SSW / (n - k))$$

No qual:

- SSB: Soma dos quadrados entre os grupos, mede a variabilidade entre os centróides dos grupos
- SSW: Soma dos quadrados dentro dos grupos, mede a variabilidade dos elementos dentro dos grupos
- k : Número de grupos
- n : Número total de objetos

O objetivo é encontrar o valor de k que maximize a razão de variância, indicando uma separação eficaz dos grupos com relação à variabilidade interna e externa. Isso proporciona uma abordagem estatística para determinar o número ótimo de agrupamentos na análise k -médias.

O critério de razão de variância (VRC) é uma ferramenta fundamental na determinação do número ideal de grupos. O VRC compara a variação entre grupos com a variação dentro dos grupos, visando identificar um equilíbrio entre coesão interna e separação entre grupos. O valor de k que maximiza o VRC é frequentemente considerado o número ótimo de grupos, representando uma configuração que captura de forma eficaz as nuances dos dados.

A eficácia da técnica k-médias na análise de clusters é evidenciada pela capacidade de encontrar agrupamentos naturais nos dados. O método permite identificar padrões que podem ser usados para fins de tomada de decisão, segmentação de mercado, análise de clientes, entre outras aplicações.

A metodologia de clusterização via k-médias desempenha um papel eficiente na criação de estruturas de classificação e indicadores. Nesse sentido, com o objetivo de melhor caracterizar a amostra resultante dos tratamentos aplicados na etapa anterior, os quatro mil e novecentos municípios cujo campo "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?" está preenchido, seja com 'sim' ou com 'não', serão classificados em uma matriz de potencial de gestão de resíduos sólidos.

A matriz será construída através de um *rating* de aderência à gestão de resíduos sólidos. O *rating* é elaborado com base nos resultados apresentados nos campos de informações específicos fornecidos ao SNIS, que permitem mapear os principais aspectos que refletem a aderência à gestão de resíduos sólidos de um município, são eles: CS001, CO008, CC020, CA004, GE202 e PO048, os quais foram previamente citados.

No contexto deste estudo, as informações acima obtidas do SNIS, serão utilizadas de maneira binária, no qual a resposta "SIM" será representada por 1 e a resposta "NÃO" será representada por 0. Tal abordagem visa criar uma estrutura que facilite a quantificação dos critérios selecionados e permita a geração de um *rating* final e individual de avaliação por município.

Cada critério será tratado como uma variável binária, em que a presença ou ausência da característica específica será codificada como 1 ou 0, respectivamente. Por exemplo, se um município atender a determinado critério, esse critério receberá o valor 1 (SIM), indicando a presença da condição. Caso contrário, o valor será 0 (NÃO), indicando a ausência.

Ao somar os valores binários de todos os critérios selecionados para avaliação, é possível obter uma pontuação cumulativa que reflete a conformidade do município com os critérios estabelecidos. Essa pontuação final é construída de forma a

fornecer uma medida quantitativa da adequação do município em relação aos aspectos avaliados. Dessa maneira, a soma das respostas binárias transforma-se em uma avaliação numérica, no qual a nota final reflete o grau da aderência às condições consideradas relevantes.

A abordagem binária e de soma cumulativa oferece uma maneira clara e objetiva de avaliar e comparar diferentes municípios com base nos critérios estabelecidos. Além disso, ela simplifica a interpretação dos resultados, permitindo a identificação das áreas de excelência e possíveis lacunas na gestão de resíduos sólidos.

No escopo deste estudo, as notas resultantes do *rating* atribuído aos municípios representam a métrica central que será submetida à análise de clusterização utilizando o algoritmo k-médias, por meio da plataforma RStudio. Essa abordagem visa discernir padrões intrínsecos na distribuição das notas, agrupando os municípios de acordo com similaridades em sua aderência à gestão de resíduos sólidos.

O algoritmo k-médias é uma técnica consagrada na análise de agrupamentos, projetada para identificar estruturas subjacentes em conjuntos de dados. No presente contexto, a clusterização será realizada com base nas notas de *rating* atribuídas aos municípios. A clusterização é uma ferramenta poderosa para revelar agrupamentos naturais ou segmentações nos dados, proporcionando uma compreensão aprofundada dos diferentes potenciais de gestão de resíduos sólidos presentes nos municípios analisados.

O cálculo do número de *clusters* é um passo crucial na determinação da estrutura de agrupamento ideal. O presente estudo aborda a análise da aderência à gestão de resíduos sólidos por meio da criação de uma matriz tripartite de avaliação (alto, médio e baixo). O propósito é proporcionar uma estrutura clara e comunicativa para a classificação da aderência à gestão, permitindo uma tomada de decisão embasada e orientada à ação. Nesse contexto, a justificativa para a adoção de três *clusters* na análise ganha relevância tanto em termos teóricos quanto práticos.

A escolha de três clusters reflete a natureza tridimensional do potencial de gestão de resíduos sólidos, possibilitando uma representação eficaz dos níveis de desempenho. A divisão em alto, médio e baixo é intuitiva, alinhando-se naturalmente com as expectativas das partes interessadas e facilitando a interpretação dos resul-

tados. Isso é crucial, uma vez que a adoção de uma matriz tridimensional visa promover a ação e a melhoria contínua nas estratégias de gestão.

Além disso, a validade teórica desse enfoque é respaldada pela existência de características distintivas nos diferentes níveis de gestão. A capacidade do método de agrupamento escolhido, no caso, o k-médias, de discernir entre esses níveis é um aspecto crítico. A análise deve identificar atributos distintos, os quais permitam a segregação eficaz das instâncias em clusters representativos dos níveis da aderência à gestão.

A importância da escolha de três *clusters* é reforçada ainda pela visualização clara que ela proporciona. Através de gráficos, por exemplo, a matriz de aderência à gestão de resíduos sólidos, as características específicas de cada cluster podem ser apresentadas de maneira tangível e acessível, facilitando a comunicação dos resultados a diferentes públicos.

Em termos estatísticos, a justificativa encontra suporte na validação dos clusters por meio de métricas objetivas, como o coeficiente de *Silhouette*. Se essas métricas indicarem uma separação consistente e bem definida entre os clusters, a escolha de três *clusters* ganha validade estatística.

No âmbito do contexto do projeto, a decisão de adotar três clusters para a análise da aderência à gestão de resíduos sólidos é influenciada pela necessidade de criar uma matriz de avaliação que seja tanto cientificamente embasada quanto facilmente interpretável pelos atores envolvidos. A abordagem tridimensional oferece um meio para atingir tais objetivos, se consolidando como uma estratégia sólida e adequada para a análise proposta.

Essa avaliação estratégica auxilia na definição adequada do número de clusters, otimizando a capacidade do algoritmo de classificação k-médias para revelar padrões significativos nos dados.

3.3 PROPENSITY SCORE MATCHING

O método *Propensity Score Matching* (PSM), desenvolvido por Paulo Rosenbaum e Donald Rubin (1983), possui como objetivo a avaliação da probabilidade

condicional de adesão a um tratamento ou programa. Esse método estatístico busca estimar o efeito de uma intervenção, política ou tratamento, levando em consideração as covariáveis que influenciam a seleção para receber tal tratamento.

A metodologia envolve a definição de um vetor de variáveis observadas para identificar indivíduos que possuem características indicativas de tratamento. A teoria do *propensity score* visa mitigar vieses potenciais, mesmo em amostras com tamanho reduzido. Os autores Rosenbaum e Rubin (1983) delinearam o método de amostragem que combina os escores de propensão, o PSM opera como uma generalização da análise discriminante, incorporando a subclassificação baseada no escore de propensão. Subclasses equivalentes são utilizadas para estimar os efeitos do tratamento em todas as variáveis das subpopulações.

De acordo com Khandker *et al.* (2010), o PSM cria um grupo de comparação que é confrontado com o grupo tratado, ele é composto por indivíduos que, embora compartilhem características observáveis com o grupo tratado, não foram contemplados pelo tratamento, os indivíduos sem pares comparativos são excluídos da análise.

Apesar de reduzir a amostra, eliminando indivíduos sem pares, o PSM permanece válido, respaldado por dois princípios fundamentais: Independência Condicional e Suporte Comum. A hipótese de Independência Condicional assegura que as variáveis pré-tratamento (X) sejam independentes do tratamento, garantindo a independência do grupo de tratamento e os impactos potenciais do programa. A hipótese de suporte comum estabelece que indivíduos do grupo tratado com características X possam encontrar correspondentes no grupo de controle, permitindo a comparação justa entre beneficiados e não beneficiados pelo programa.

Ribeiro e Cacciamali (2012) corroboram essa ideia ao afirmar que o PSM se assemelha aos métodos de correspondência. A estratégia visa estabelecer um grupo de controle contrafactual que se assemelhe ao grupo de tratamento em termos de características observáveis. Cada participante é emparelhado com um não participante com características semelhantes observadas e a diferença média dos resultados entre esses grupos é comparada para estimar o efeito do tratamento. Essas diferenças médias entre os grupos tratados e de controle em relação às variáveis de resposta são, portanto, calculadas.

No entanto, ao compor o grupo de controle, é essencial lidar com o viés de seleção, o qual pode surgir de características observáveis e não observáveis distintas entre os grupos. Para mitigar esse problema é crucial identificar grupos semelhantes com precisão, embora encontrar indivíduos exatamente equiparados nos dois grupos exija rigor (Khandker *et al.*, 2010).

Portanto, a estratégia de Pareamento pelo Escore de Propensão surge como uma solução em potencial para atenuar o viés de seleção. Nesse contexto, cada indivíduo participante seria comparado com um não participante, usando um escore de propensão único, o qual encapsula a probabilidade condicional de adesão, considerando suas distintas características observadas representadas por X . Esse enfoque reflete as duas premissas fundamentais do modelo: a independência condicional e o suporte comum.

Conforme argumentado por Ribeiro e Cacciamali (2012), a variável a ser explicada, denotada por Y , encapsula a participação do indivíduo no programa. As variáveis observadas, representadas por X , abrangem as informações sobre os candidatos antes de sua participação no programa. A fase inicial do método *Propensity Score Matching* (PSM) visa identificar a adesão dos indivíduos ao programa.

O objetivo central do método *Propensity Score Matching* (PSM) é gerar um sistema de contagem de pontos, escore, que corresponde à probabilidade de receber tratamento, considerando tanto os indivíduos tratados quanto os não tratados, com base em um conjunto específico de covariáveis predefinidas, as quais servem para identificar a elegibilidade (probabilidade) de participação no programa (Cameron; Trivedi, 2005). Embora teoricamente qualquer técnica de regressão binária (logit, probit ou mínimos quadrados ordinários) possa ser empregada para estimar essa probabilidade, as regressões logísticas têm sido as mais amplamente utilizadas para esse propósito (Becker; Ichino, 2002).

Após a obtenção da estimativa do modelo de escolha binária que incorpora tanto os participantes quanto os não participantes - e sob a suposição de que três condições, formalmente demonstradas por Rosenbaum e Rubin (1983), são satisfeitas:

- 1) Independência Condicional;
- 2) Suporte Comum;

O passo subsequente é construir o grupo de controle por meio dos escores de propensão (probabilidades estimadas). Isso resulta na criação de uma situação contrafactual, permitindo a análise do impacto do programa.

Na literatura, há diversas técnicas de emparelhamento (*matching*) disponíveis para a geração do grupo de controle usando os *propensity scores*, e a escolha entre elas pode depender da natureza dos dados, sem necessariamente haver uma técnica que seja superior às demais (Caliendo; Kopeinig, 2008). Diante disso, a recomendação dada por Caliendo e Kopeinig (2008) é a utilização de diversos algoritmos como uma forma de verificar a robustez do impacto estimado do programa em análise. Neste estudo, serão adotadas algumas das técnicas² mais comuns, listadas a seguir:

- 1) vizinho mais próximo (nearestneighbor);
- 2) vizinho mais próximo com distância pré-definida através de um calibre (RadiusMatchingwithcaliper);
- 3) regressão não paramétrica (Kernel Matching);
- 4) regressão linear local (local linear regression);
- 5) emparelhamento através das variáveis pré-determinadas (covariate-matching).

3.4 MÉTODO DOS MÍNIMOS QUADRADOS ORDINÁRIOS

O modelo de regressão linear desempenha um papel de destaque na análise de dados, segundo os autores Hair *et al.* (2009, p. 176), "a análise de regressão múltipla é uma técnica estatística que pode ser usada para examinar a relação entre uma única variável dependente e várias variáveis independentes (preditoras)". Sendo assim, através da regressão, é possível avaliar o grau de associação entre a variável dependente Y e o conjunto de variáveis independentes Xi.

² Formações detalhadas sobre esses procedimentos podem ser encontradas em Cameron e Trivedi (2005) e Caliendo e Kopeinig (2008).

A representação de dados numéricos por meio de gráficos e tabelas é uma abordagem frequentemente utilizada para visualização, contribuindo para antecipar interpretações e análises. A exploração mais profunda dos dados é iniciada com duas técnicas estatísticas: 1) a análise de medidas de tendência central; e, conseqüentemente 2) a avaliação das medidas de dispersão. No contexto da econometria, a adoção dos conceitos de média e desvio-padrão para descrever o comportamento na regressão surge do interesse em determinar se há influências causais. Caso sejam identificadas, a análise de regressão poderia oferecer projeções mais precisas da média, lembrando que os modelos econométricos frequentemente buscam testar teorias econômicas (Gujarati; Porter, 2011).

No contexto da pesquisa econométrica, o objetivo central é oferecer medições concretas que sustentem a teoria econômica, utilizando a inferência estatística como elo. Os autores citados anteriormente enfatizam que a econometria teórica se dedica ao desenvolvimento de métodos apropriados para mensurar relações econômicas especificadas em modelos econométricos. É crucial ressaltar que a econometria mantém uma relação intrínseca com a estatística matemática. Neste cenário, um dos métodos amplamente empregados é o dos Mínimos Quadrados Ordinários (MQO).

De forma empírica, os métodos estatísticos de análise de dados, fundamentados nos princípios da econometria, têm se firmado como uma abordagem robusta dentro da metodologia. Tal fato se justifica pela capacidade da econometria em fornecer estruturas sólidas para analisar dados econômicos, abrangendo tópicos como produção, consumo, distribuição, demanda, oferta, juros, renda nacional e balança comercial. Conseqüentemente, a contribuição da economia tem sido notável na melhoria das métricas da informação, alinhando-as a essa perspectiva.

No contexto brasileiro, a adoção dessas práticas é observada de forma moderada, especialmente quando comparada à realidade nos Estados Unidos. De acordo com Soares (2005), no campo das ciências sociais no Brasil, existe uma notável "resistência em relação a métodos quantitativos e estatística". A pesquisa de Valle e Silva (1999) e Santos e Coutinho (2000) reforça essa avaliação, indicando que a aplicação de abordagens básicas de estatística descritiva e inferencial ainda é limitada nas Ciências Sociais, independentemente do formato da produção (sejam artigos, dissertações ou teses).

O desdobramento disso é a diminuição generalizada da robustez do conhecimento científico. Curiosamente, mesmo os métodos qualitativos não parecem escapar dessa situação. Assim, conforme apontado por Soares (2005), a ausência de métodos quantitativos não implica automaticamente a prevalência de métodos qualitativos. Em geral, o cenário predominante é a carência de método.

4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

4.1 ANÁLISE DOS RESULTADOS E CONSTRUÇÃO DA MATRIZ DE ADERÊNCIA À GESTÃO DE RESÍDUOS SÓLIDOS

Espera-se obter um número de *clusters* distintos após a aplicação do algoritmo, cada um representando um nível diferente de aderência à gestão de resíduos sólidos. Esses *clusters* serão interpretados como categorias representativas dos seguintes potenciais: alto, médio e baixo. A clusterização permitirá discernir quais municípios apresentam características semelhantes em termos de potenciais de gestão, e a divisão em três *clusters* distintos possibilitará uma classificação clara dos municípios em diferentes níveis de aptidão.

A aplicação do algoritmo k-médias, juntamente com a interpretação dos *clusters* resultantes, proporcionará uma compreensão mais profunda das diferentes dinâmicas e tendências relacionadas à gestão de resíduos sólidos em âmbito municipal. A identificação desses *clusters* pode ser um passo significativo na formulação de estratégias mais eficazes e direcionadas para aprimorar a gestão de resíduos sólidos em diferentes regiões.

A descrição detalhada da abordagem de clusterização, utilizando o algoritmo k-médias para segmentar as notas de *rating*, em diferentes grupos representativos de aderência à gestão de resíduos sólidos nos municípios pode ser vista a seguir:

O algoritmo k-médias funciona no contexto do R conforme as etapas abaixo:

1. Preparação dos Dados: Os dados foram preparados, incluindo a extração da coluna de notas de *rating* dos municípios, a qual é a variável de interesse para a clusterização.
2. Definição do Número de Clusters: Foi definido três, como o número de *clusters* desejado, conforme estabelecido no tópico anterior. Isso determina quantos grupos distintos o algoritmo tentará identificar, para análise do estudo 3 níveis de capacidade de gestão, são eles: alto, médio e baixo.
3. Execução do Algoritmo K-médias: O algoritmo k-médias foi aplicado aos dados de notas de *rating*. Ele inicia selecionando aleatoriamente três pontos como centros iniciais dos *clusters* (esses centros são chamados de "médias").

Em seguida, ele atribui cada ponto de dado ao cluster cujo centro está mais próximo.

4. Atualização das Médias: Com os pontos atribuídos aos clusters, o algoritmo calcula a média de todos os pontos em cada cluster e atualiza os centros dos clusters com essas médias.
5. Reatribuição de Pontos: O algoritmo então reatribui cada ponto de dado ao cluster cujo centro é mais próximo após a atualização.
6. Iteração: Os passos 4 e 5 são repetidos iterativamente até que não haja mais mudanças significativas nas atribuições dos pontos aos clusters ou até que um número máximo de iterações seja alcançado.
7. Resultado: O algoritmo converge para uma solução na qual os pontos estão agrupados em clusters de forma a minimizar a soma dos quadrados das distâncias entre os pontos e os centros dos clusters. O resultado é uma lista de clusters, com cada ponto de dado atribuído a um cluster específico.

Com base nos resultados da clusterização, é possível interpretar os grupos formados e atribuir significado a eles, os três *clusters* representam diferentes níveis de aderência à gestão de resíduos sólidos nos municípios.

Os *clusters* obtidos após a aplicação do algoritmo k-médias representam diferentes níveis de aderência à gestão de resíduos sólidos nos municípios analisados. Nesse contexto, conforme definido, o *cluster* '1' reflete um alto nível de capacidade de gestão, o *cluster* '2' indica uma capacidade de gestão de nível intermediário e o *cluster* '3' está associado a um baixo nível de capacidade de gestão.

Utilizando a própria planilha de resultado, realizaremos a segmentação desses níveis de gestão por região, o que resultará na criação de uma matriz de aderência à gestão de resíduos sólidos para os municípios do Brasil, organizados por região.

Para realizar essa etapa, estes foram os passos executados:

1. Organização dos Dados por Região: Primeiro, os dados são organizados de acordo com as regiões geográficas do Brasil. Essa organização permitirá a criação de uma matriz que agregará as informações de capacidade de gestão de resíduos sólidos por região.

2. Agrupamento por *Clusters*: Em seguida, os resultados da clusterização são utilizados para atribuir cada município a um dos clusters identificados (1, 2 ou 3), com base na aderência à gestão associado.
3. Matriz de Aderência à Gestão de Resíduos Sólidos: Com os municípios agrupados por região e *clusters*, é possível criar uma matriz em que as linhas representarão as diferentes regiões do Brasil e as colunas corresponderão aos *clusters* de aderência à gestão. A matriz conterá informações sobre quantos municípios de cada região pertencem a cada um dos *clusters*.

Com a matriz criada, é possível visualizar a distribuição dos municípios por região e seus respectivos níveis de aderência à gestão. Isso permitirá uma análise mais abrangente das dinâmicas regionais em termos de gestão de resíduos sólidos.

Figura 1 – Matriz de Aderência à Gestão de Resíduos Sólidos dos municípios por região.

Norte	285	45	299
Nordeste	86	53	1122
Centro - Oeste	75	34	296
Sudeste	423	124	1003
Sul	285	53	799
	Alto Potencial	Médio Potencial	Baixo Potencial

Fonte: Elaboração Própria, a partir dos dados compilados a partir do relatório do SNIS.

Com base nas informações obtidas na matriz de aderência à gestão, é possível identificar regiões com necessidades específicas de aprimoramento e direcionar estratégias mais eficazes para o desenvolvimento da gestão de resíduos sólidos em diferentes partes do Brasil. Essa primeira caracterização das amostras, as quais serão comparadas, proporciona uma visão ampla e estratégica da capacidade de gestão de resíduos sólidos em níveis regionais, permitindo uma alocação mais direcionada de recursos e ações para melhoria das práticas de gestão.

4.2 APLICAÇÃO DO *PROPENSITY SCORE MATCHING*

Na segunda etapa da análise, o objetivo principal é avaliar o impacto da presença do Plano Municipal de Gestão Integrada de Resíduos Sólidos (PMGIRS), conforme definido na Lei nº 12.305/2010, o qual estabelece a Política Nacional de Resíduos, sobre a Taxa de Recuperação de Materiais Recicláveis. Para atingir esse objetivo, será empregada a ferramenta estatística do *Propensity Score Matching*.

O *Propensity Score Matching* é uma técnica de análise que visa equilibrar as características das unidades de observação (neste caso, os municípios) que possuem o tratamento (presença do PMGIRS) e as que não possuem, de modo a criar grupos comparáveis. Isso é alcançado por meio do cálculo de um escore de propensão, que reflete a probabilidade de cada unidade receber o tratamento com base em suas características observáveis.

Dessa forma, para cada município é calculada a probabilidade de ter um PMGIRS com base nas características presentes nos campos existentes na base de dados do SNIS, neste caso, os campos FN208, FN206, FN220, POP_TOT já mencionados na seção 3.1.

Através da metodologia do *Propensity Score Matching*, são selecionados municípios que não possuam PMGIRS, mas que compartilhem as características acima descritas mais aproximadas com aqueles que possuem. Esses municípios de controle formam o grupo de comparação.

É relevante destacar que, para a realização do algoritmo *Matchit*, são essenciais algumas verificações na base de dados. Estas incluem:

- Verificação de Campos Vazios: É imperativo assegurar que os campos envolvidos no processo não contenham lacunas ou valores ausentes. Isso garante que a análise seja conduzida com informações completas e confiáveis;
- Validação da Variável *Dummy*: A variável *dummy* considerada na base de dados precisa ser efetivamente binária em todos os registros. A presença de quaisquer valores que não sejam estritamente 0 ou 1 pode comprometer a precisão da análise;

- Exclusão de Linhas Vazias: Qualquer linha que contenha informações faltantes deve ser excluída da análise. Isso evita distorções nos resultados e assegura que apenas os dados completos e relevantes sejam considerados.

Essas verificações são essenciais para garantir a qualidade dos resultados obtidos por meio do algoritmo *MatchIt*. Uma base de dados consistente e bem-preparada é fundamental para obter conclusões sólidas e confiáveis na análise de impacto do Plano Municipal de Gestão Integrada de Resíduos Sólidos.

Com os grupos comparáveis, é possível comparar a Taxa de Recuperação de Materiais Recicláveis e o Custo Unitário Médio do Serviço de Coleta entre os municípios com e sem PMGIRS. Essa comparação permite avaliar o efeito da presença do plano nesses indicadores.

Tabela 1 – Resultado do pareamento pós PSM

Summary of Balance for All Data:		
	Means Treated	Means Control
distance	0,5526	0,5455
FN206	596087,9706	314819,3043
FN208	3972473,659	1586541,56
FN220	9023951,124	3454266,297
POP_TOT	62243,1823	32496,8733
Summary of Balance for Matched Data:		
	Means Treated	Means Control
distance	0,5532	0,5455
FN206	611602,8717	314819,3043
FN208	4368878,635	1586541,56
FN220	9865721,015	3454266,297
POP_TOT	66730,5142	32496,8733

Fonte: RStudio. Elaboração própria.

A análise dos resultados comparativos do *MatchIt* envolve a avaliação das características das amostras tratadas e não tratadas após a aplicação do método de pareamento. O *MatchIt* é uma ferramenta estatística usada para criar grupos de tra-

tamento e controle comparáveis em estudos observacionais, minimizando o viés de seleção.

Entre as informações apresentadas no resumo, duas métricas importantes são *sample sizescontrol* e *sample sizestreated*, as quais se referem aos tamanhos das amostras nos grupos de controle e tratamento, respectivamente.

A métrica *sample sizescontrol* indica o tamanho da amostra no grupo de controle após o pareamento ter sido aplicado. Esse número representa quantos indivíduos ou unidades pertencentes ao grupo de controle foram retidos após o processo de pareamento, ou seja, quantos foram considerados semelhantes o suficiente aos indivíduos ou unidades no grupo de tratamento para formar pares. Uma amostra controlada de tamanho adequado é crucial para garantir que as comparações entre os grupos sejam estatisticamente robustas e que os resultados sejam confiáveis.

Por outro lado, a métrica *sample sizestreated* representa o tamanho da amostra no grupo de tratamento após o pareamento. Isso indica quantos indivíduos ou unidades originalmente no grupo de tratamento foram mantidos após o processo de pareamento. Similar à métrica *sample sizescontrol*, o tamanho da amostra no grupo de tratamento possui um papel crítico na determinação da força estatística e na validade das conclusões. Um grupo de tratamento suficientemente grande assegura que as estimativas de efeito ou associação sejam calculadas com precisão.

A comparação desses tamanhos de amostra controlados e tratados após o pareamento é crucial para avaliar a eficácia do processo de pareamento em equilibrar as características entre os grupos. A convergência dos tamanhos de amostra nos grupos de controle e tratamento sugere que o pareamento foi bem-sucedido em criar grupos comparáveis. Se houver uma disparidade significativa entre os tamanhos de amostra, pode indicar desequilíbrio nas características ou que o processo de pareamento não foi capaz de criar pares suficientemente semelhantes.

No contexto deste estudo, os resultados da aplicação do método de *Propensity Score Matching* (PSM) visam avaliar o equilíbrio entre grupos tratados e de controle após o processo de emparelhamento com base nas pontuações de propensão. Inicialmente, antes do emparelhamento, se observou que algumas variáveis apresentavam desequilíbrio significativo entre os grupos. Essas variáveis incluíam

FN206, FN208, FN220 e POP_TOT, com médias consideravelmente diferentes entre os grupos.

No entanto, após a implementação do PSM, foi observado uma melhoria substancial no equilíbrio dessas variáveis. As médias das variáveis FN206, FN208, FN220 e POP_TOT no grupo tratado se aproximaram das médias correspondentes no grupo de controle, indicando uma redução significativa no viés. Esses resultados sugerem que o método PSM foi eficaz em tornar os grupos mais comparáveis e reduzir a disparidade inicial entre eles, o que é fundamental para análises subsequentes em estudos observacionais. Assim, é perceptível que o resultado do pareamento criou dois grandes grupos, o primeiro com 1579 municípios para o tratamento e o segundo 1579 para o controle.

Tabela 2 – Resultados dos *Sample Sizes* em RStudio.

Sample Sizes:	Control	Treated
All	1579	1925
Matched	1579	1579
Unmatched	0	346
Discarded	0	0

Fonte: Elaboração Própria. RStudio.

Ao controlar possíveis vieses de seleção, obtemos evidências mais sólidas sobre a eficácia do PMGIRS nos municípios.

4.3 REGRESSÃO LINEAR COM DADOS DO *MATCHING*

A amostragem final compreende uma base de dados composta por 3.158 municípios, dos quais metade possui Planos Municipais de Gestão Integrada de Resíduos Sólidos (PMGIRS) e a outra metade não, tal arranjo viabiliza a aplicação de uma análise de Mínimos Quadrados Ordinários (MQO) para efetuar uma avaliação mais precisa do efeito causal resultante da presença dos PMGIRS sobre a Taxa de Recuperação de Materiais Recicláveis.

Em uma análise de regressão linear, as variáveis explicadas e as variáveis explicativas desempenham papéis fundamentais para entender e modelar as rela-

ções entre diferentes conjuntos de dados. Tais variáveis têm funções distintas e são essenciais para a construção e interpretação de um modelo de regressão.

1. Variáveis Dependentes: são aquelas que se deseja entender, prever ou explicar por meio do modelo de regressão³. São os resultados ou as observações que buscamos analisar em relação às mudanças nas variáveis independentes. No estudo proposto, a variável dependente será a *IN031_RS - Taxa de recuperação de materiais recicláveis (exceto matéria orgânica e rejeitos)* em relação à quantidade total Resíduos Domiciliares Orgânicos (RDO) + Resíduos Perigosos Urbanos (RPU) coletada.
2. Variáveis Independentes⁴: As variáveis explicativas, também conhecidas como variáveis independentes, são aquelas utilizadas para explicar ou prever as variações na variável explicada. São as características, fatores ou influências que estão associados às mudanças na variável dependente. No estudo proposto, as variáveis explicadas utilizadas serão:
 - a. IN023_RS - Custo unitário médio do serviço de coleta (rdo + rpu);
 - b. IN019_RS - Taxa de empregados (coletadores + motoristas) na coleta (rdo + rpu) em relação à população urbana;
 - c. 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.

Desse modo, o objetivo da regressão linear é encontrar um relacionamento matemático entre as variáveis explicativas e a variável explicada que melhor descreva os dados observados. Através da análise estatística dos dados, o modelo de regressão procura identificar a influência das variáveis independentes sobre a variável dependente e quantificar essa relação por meio dos coeficientes de regressão. Isso permite fazer previsões ou inferências sobre os efeitos das variáveis independentes sobre a variável dependente.

A execução do código no ambiente RStudio é efetuada mediante a utilização do comando *lm()*. Nesse contexto, a modelagem de regressão é implementada por

³ Para uma compreensão aprofundada sobre variáveis dependentes em análise de regressão, bem como sua relação com variáveis independentes, consulte Gujarati e Porter (2011).

⁴ Para uma compreensão aprofundada sobre variáveis independentes em análise de regressão, bem como sua relação com variáveis dependentes, consulte Gujarati e Porter (2011).

meio do comando $lm(IN31_RS \sim -1 + PO048 + IN023_RS, data = Modelo)$, no qual o termo Modelo se refere à base de dados derivada do armazenamento da tabela resultante do procedimento de Pareamento por Escore de Propensão (PSM) previamente executado.

Os resultados da regressão podem ser observados no quadro abaixo:

Tabela 3 – Resultados da Regressão

Call:
lm(formula=IN031_RS~-1+PO048+IN023_RS+IN019_RS,data = Modelo)

Residuals:

	Min	1Q	Median	3Q	Max
	-26.622	-4.315	-1.755	3.493	51.453

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
PO048	2.405.334	0.410303	5.862	5.77e-09***
IN023_RS	0.013343	0.001385	9.631	< 2e-16 ***
IN019_RS	1.872.618	0.231625	8.085	1.40e-15***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ''

Residual standard error: 8777 on 1314 degrees of freedom
(1841 observations deleted due to missingness)

Multiple R-squared: 0.4216, Adjusted R-squared: 0.4203
F-statistic: 319.3 on 3 and 1314 DF, p-value: < 2.2e-16

Fonte: Elaboração própria, através do ambiente RStudio.

A equação associada ao resultado da regressão é dada por:

$$IN031_{RS} = 2,381 * PO048 + 0,015 * IN023_{RS} + 1,778 * IN019_{RS}$$

Para os resultados acima apresentados, é possível considerar que:

- As variáveis PO048, IN023_RS e IN019_RS são todas significativas, com p-valores muito baixos;
- A estatística F é alta e o valor p associado é baixo (< 2.2e-16), indicando que o modelo como um todo é significativo.

Verifica-se, com significância estatística, que a implementação de diretrizes e estratégias delineadas pelo PMGIRS resulta em um aumento estatisticamente significativo na eficiência da recuperação de materiais recicláveis. O impacto estimado de

2.38% na taxa de recuperação confirma a relevância prática do enfoque utilizado, sugerindo um aumento expressivo na proporção de materiais recicláveis resgatados em relação ao volume total coletado.

Além disso, é observado que a influência não se restringe somente ao PMGIRS. A análise expõe que o custo unitário médio do serviço de coleta (RDO + RPU) inerente ao IN023_RS e a taxa de empregados (coletadores + motoristas) na coleta (RDO + RPU) relativo ao IN019_RS, em relação à população urbana, também apresentam impactos de significância sobre a taxa de recuperação de materiais recicláveis (IN031_RS). O estudo aponta a importância da consideração dessas variáveis inter-relacionadas na definição de políticas e estratégias que aspiram otimizar a recuperação de recursos valiosos a partir dos resíduos sólidos.

O propósito da clusterização utilizando o método k-médias consistiu na categorização dos municípios brasileiros de acordo com seu potencial na gestão de resíduos sólidos, com base nas informações disponíveis na base de dados do Sistema Nacional de Informações sobre Saneamento (SNIS). Uma análise da matriz de resultados revelou uma notável concentração de municípios na categoria de baixo potencial, o que reflete desafios relacionados aos indicadores utilizados na composição da classificação ou à ausência de dados, um problema de igual relevância.

Em aprofundamento a análise do mapeamento, como parte da investigação do impacto desses indicadores em resultados pertinentes para práticas contemporâneas de tratamento de resíduos sólidos, como a reciclagem, a variável *PO048 - O Município possui Plano Municipal de Gestão Integrada de Resíduos Sólidos (PMGIRS) conforme a Lei nº 12.305/2010*, a qual aborda a Política Municipal de Resíduos, foi selecionada como variável explicada.

Devido à concentração significativa de municípios na categoria de baixo potencial, vista na matriz, e com o objetivo de mitigar possíveis vieses de seleção, foi aplicada uma abordagem de *Propensity Score Matching* (PSM) aos municípios que compõem a matriz. O objetivo era identificar, dentro das diversas faixas de potencial, municípios com características semelhantes para formar grupos de controle e tratamento. Por meio dessa metodologia, foi possível selecionar, dentre os municípios

listados na matriz de aderência à gestão de resíduos sólidos, aqueles mais similares para posterior análise por regressão linear.

Os resultados obtidos na regressão indicam que a presença e a análise do impacto do PMGIRS sobre a taxa de recuperação de materiais recicláveis são positivas. O propósito fundamental deste projeto é apresentar uma abordagem eficaz para a análise de impacto dos componentes de uma matriz de aderência à gestão de resíduos sólidos, a fim de identificar insights que possam informar as estratégias de tomada de decisão dos municípios em relação à gestão de resíduos sólidos.

5 CONCLUSÃO

No contexto deste estudo, são apresentadas conclusões importantes para a gestão e aprimoramento dos resíduos sólidos municipais. Através da utilização de uma análise de regressão, é possível mensurar o impacto da presença do Plano Municipal de Gestão Integrada de Resíduos Sólidos (PMGIRS) sobre a taxa de recuperação de materiais recicláveis em relação à quantidade total (RDO + RPU).

A promoção e execução efetiva do PMGIRS emergem como ferramentas primordiais para impulsionar a eficiência da recuperação de materiais recicláveis, contribuindo para a atenuação dos impactos ambientais decorrentes da disposição inadequada desses resíduos. Concomitantemente, a consideração dos custos associados à coleta e à taxa de empregados possibilita uma compreensão holística das dinâmicas envolvidas na maximização da recuperação de recursos, orientando tomadas de decisão informadas e embasadas em evidências empíricas. Dessa forma, esta pesquisa incita a reflexão contínua sobre as abordagens de gestão sustentável dos resíduos sólidos municipais em busca de soluções cada vez mais eficazes e alinhadas com as exigências da preservação ambiental e desenvolvimento socioeconômico.

Por outro lado, mediante a instância de uma matriz dedicada à avaliação da aderência à gestão de resíduos sólidos, operando por meio da assimilação de indicadores e traços provenientes do Sistema Nacional de Informações sobre Saneamento (SNIS), permite uma categorização que aponta o atual estado de municípios sob esta perspectiva.

Os resultados obtidos ostentam uma aglomeração proeminente de municípios, estendendo por todas as regiões do país, situados na faixa de aderência à gestão de resíduos sólidos reduzido. Este fato implica que uma proporção substancial de localidades enfrenta desafios intrínsecos à otimização de suas estratégias para a gestão de resíduos sólidos.

Fundamental destacar que a construção do *rating* em si incorporou uma amalgamação de critérios distintos, abraçando a execução de atividades, o grau de especialização das operações, o porte da força laboral e, não menos importante, a consolidação de contratos de delegação, além do próprio Plano Municipal de Gestão Integrada de Resíduos Sólidos (PMGIRS) como norteador principal.

A matriz direciona a atenção para as regiões Nordeste, Sudeste e Sul do país, devido maior densidade demográfica, convocando, portanto, estratégias específicas para abordar os desafios complexos tangíveis à gestão de resíduos sólidos. Diante disso, os PMGIRS e seus desdobramentos estratégicos ratificam a essencialidade de tais planos como orientadores para a proposição de direcionamentos condizentes com as necessidades de cada município, direcionados para a redução dos obstáculos diagnosticados no contexto da gestão de resíduos sólidos.

Tal panorama acentua a necessidade de articular estratégias de gestão perspicazes, abarcando a formulação de planos estratégicos alinhados aos ditames da Matriz de Aderência à Gestão de Resíduos Sólidos, com o fito de enfrentar essas adversidades de forma focalizada e eficaz.

REFERÊNCIAS

ALZAMORA, B. R.; BARROS, R. T. de V. Review of municipal waste management charging methods in different countries. *Waste Management*, v. 115, p. 47–55, 2020.

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.

BRANDÃO, A. de O.; SILVA, G. N. da. Impactos econômicos da implantação de aterros sanitários individuais nos municípios. *Revista Holos*, Natal, v. 3, 2011.

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

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.

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.

BRASIL. Decreto nº 7.404, de 23 de dezembro de 2010. Regulamenta a Lei nº 12.305, de 2 de agosto de 2010, que institui a Política Nacional de Resíduos Sólidos, cria o Comitê Interministerial da Política Nacional de Resíduos Sólidos e o Comitê Orientador para a Implantação dos Sistemas de Logística Reversa, e dá outras providências.

BRASIL. (2007a) 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. *Diário Oficial da União*, Brasília, 8 de janeiro de 2007.

BRASIL. (2007b) Manual de Saneamento. 3. ed. rev. - Brasília: Fundação Nacional de Saúde.

BRASIL. (2010a) 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.

BRASIL. (2010b) Decreto nº 7.404, de 23 de dezembro de 2010. Regulamenta a Lei nº 12.305, de 2 de agosto de 2010, que institui a Política Nacional de Resíduos Sólidos, cria o Comitê Interministerial da Política Nacional de Resíduos Sólidos e o Comitê Orientador para a Implantação dos Sistemas de Logística Reversa, e dá outras providências.

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. C.; TRIVEDI, P. K. *Microeconometrics: Methods and Applications*. Cambridge University Press, 2005.

CASADO, E. S. *Las ciencias de La documentación*. Barcelona: Editorial Ariel, 1994.

CHAVES, G. L. D.; SANTOS, J. L. dos; ROCHA, S. M. S. The challenges for solid waste management in accordance with Agenda 21: A Brazilian case review.

COLVERO, D. A.; RAMALHO, J.; GOMES, A. P. D.; MATOS, M. A. A. de; TARELHO, L. A. da C. (2020). Economic analysis of a shared municipal solid waste management facility in a metropolitan region. *Waste Management*, v. 102, p. 823–837, agosto 2014.

CONKE, L. S. (2018). Barriers to waste recycling development: Evidence from Brazil. *Resources, Conservation and Recycling*, v. 134, p. 129–135, outubro 2017.

DIAS, J. D. S. *et al.* Avaliação da Política Nacional dos Resíduos Sólidos (PNRS): Uma análise para as regiões brasileiras. Acesso em: setembro de 2023. Disponível em: https://www.anpec.org.br/sul/2019/submissao/files_l/i4-64541614065b646e1f636df92c8b7bc8.pdf.

DUTRA, R. M. S.; YAMANE, L. H.; SIMAN, R. R. Influence of the expansion of the selective collection in the sorting infrastructure of waste pickers' organizations: A case study of 16 Brazilian cities. *Waste Management*, v. 77, 2018, p. 50–58.

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.

HAIR JR, J. F. *et al.* *Análise multivariada de dados*. 6. ed. Porto Alegre: Bookman, 2009, p. 176.

HALKOS, G.; PETROU, K. N. (2019). Assessing 28 EU member states' environmental efficiency in national waste generation with DEA. *Journal of Cleaner Production*, v. 208, p. 509–521.

HARTIGAN, P.; WONG, M. A. A k-means clustering algorithm: algorithm AS 1366. *Applied Statistics*, n. 28, p. 126-130, 1979.

JOHNSON, R. A.; WICHERN, D. W. *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall, 1998.

KAZA, S.; YAO, L.; BHADA-TATA, P.; VAN WOERDEN, F. (2018). *What a waste 2.0: A Global snapshot of solid waste Management to 2050* Word Bank; World Bank.

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

LEAL, T. L. M.; SAMPAIO, R. J. (2019). Solid waste management: The case of the sustainable development consortium of alto sertão in Bahia. *Revista Brasileira de Gestão Urbana*, v. 13, p. 1–13.

MACQUEEN, E. Some methods for classification and analysis of multivariate observations. In: *Proceedings of the 5th Berkeley Symposium on mathematical statistics and probability*, v. 1. p. 281-297. Berkeley, CA: University of California Press, 1967.

MARTINS, J. P. *et al.* Responsabilidade Ambiental, Custo de Capital, Risco e endividamento. *Advances in Scientific & Applied Accounting*, v. 12, n. 2, 2019.

MILLIGAN, G. W.; COOPER, M. C. An Examination of Procedures for Determining the Number of Clusters in a Data Set. *Psychometrika*, 50, 159-179, 1985.

MINGOTI, S. A. Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada. Minas Gerais: Editora UFMG, 2007.

MOOI, E.; SARSTEDT, M. A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistic. Heidelberg: Springer, 2011.

RAVICHANDRA RAO, I. Measures of Association. New Delhi: Wiley Eastern Limited, 1986.

RIBEIRO, A. C. M.; CACCIAMALI, M. C. Educação Corporativa e Gestão do Conhecimento. *Revista de Administração*, v. 47, n. 4, p. 671-683, 2012.

RIBEIRO, W. A. A relação entre os marcos regulatórios do saneamento básico e dos resíduos sólidos. In: PHILIPPI JUNIOR, A. (Coord.) Política nacional, gestão e gerenciamento de resíduos sólidos. São Paulo: Manole. p. 541-560, 2012.

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.

SAMOILENKO, S.; OSEI-BRYSON, K. Increasing the discriminatory Power of DEA in the presence of the sample heterogeneity with cluster analysis and decision trees. *Expert Systems with Applications*, v. 34, p. 1568-1581, 2008.

SILVA, N. V. (1999), Relatório de Consultoria sobre Melhoria do Treinamento em Ciência Social quantitativa e Aplicada no Brasil, Rio de Janeiro, Laboratório Nacional de Computação Científica, 15 de abril de 1999, 22 pág.

SNIS. SNIS - Série Histórica Recuperado em 09 de janeiro, 2021 de Sistema Nacional de Informações Sobre Saneamento, 2022.

SOARES, G. O Calcanhar Metodológico da Ciência Política no Brasil. *Sociologia, Problemas e Práticas*, n. 48, p. 27-52, 2005.

TEIXEIRA, A. L. *et al.* Análise espacial de indicadores sócio-econômicos aplicada à gestão na área de saneamento. Acesso em: 03 setembro 2023. Disponível em: <https://files.abrhidro.org.br/Eventos/Trabalhos/155/301.pdf>.

TONI, J. D. (2003) Planejamento e elaboração de projetos: um desafio para a gestão no setor público. Porto Alegre: FDRH.

VIANNA, I. L.; AMARAL, E. F. de L. Utilização de metodologias de avaliação de políticas públicas no Brasil. Aplicações de técnicas avançadas de avaliação de políticas públicas, p. 15, 2014.

WITTEN, I. H.; FRANK, E. *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco: Elsevier, 2005.

ANEXOS

RStudio 2023.06.0 Build 421

#Clusterização via Kmeans

Importando as bibliotecas

```
library(tidyverse)
library(factoextra)
library(openxlsx)
```

Carregar a tabela a partir do arquivo CSV

```
Rating <- "C:\Users\Vicente\Desktop\TCC\Rating.xlsx"
```

Verificar a estrutura da tabela

```
glimpse(Rating)
```

Selecionar a coluna "Rating" para a clusterização

```
coluna_clusterizacao <- Rating$Rating
```

Definir o número de clusters desejado

```
num_clusters <- 3
```

Executar o algoritmo k-means para a clusterização

```
kmeans_result <- kmeans(coluna_clusterizacao, centers =
num_clusters)
```

Obter os rótulos de cluster para cada amostra

```
rotulos_clusters <- kmeans_result$cluster
```

Adicionar os rótulos de cluster à tabela original

```
Rating$rotulo_cluster <- rotulos_clusters
```

Visualizar as primeiras linhas da tabela com os rótulos de cluster

```
head(Rating)
```

Salvar a tabela com os rótulos de cluster em um novo arquivo Excel

```
caminho_novo_arquivo_excel <-
"C:/Users/Vicente/Desktop/TCC/metodologia.xlsx"
```

Salvar a tabela com os rótulos de cluster em um novo arquivo Excel

```
write.xlsx(Rating, caminho_novo_arquivo_excel)
```

#Propensity Score Matching

#Importando as bibliotecas

```
library(MatchIt)
library(AER)
library(readxl)
```

Importando a base de dados

```
base_de_trabalho <- read_excel
("C:/Users/Vicente/Desktop/TCC/base de trabalho1.xlsx")
```

Verificar valores ausentes na variável de tratamento

```
base <- base_de_trabalho
[complete.cases(base_de_trabalho$PO048), ]
```

Fazendo o matching com o método "nearest"

```
matchtest <- matchit(PO048 ~ FN206 + FN208 + FN220 + POP_TOT,
data = base, method = "nearest", m.order = "random", ratio =
1)
```

Verificando o processo de matching

```
summary(matchtest)
```

Criando nova base apenas com as observações com matching

```
Modelo <- match.data(matchtest)
```

#Regressão Linear

Regressão Linear com os dados do matching

```
reg1 <- lm( IN031_RS ~ -1 + PO048 + IN023_RS + IN019_RS, data
= Modelo)
```

Exibindo resultados da regressão

```
summary(reg1)
```