

UNIVERSIDADE DO ESTADO DE SANTA CATARINA – UDESC
CENTRO DE CIÊNCIAS TECNOLÓGICAS – CCT
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL

JANINE GARCIA

DETERMINAÇÃO DOS FATORES INFLUENTES NO CONSUMO DE ÁGUA EM
RESIDÊNCIAS UNIFAMILIARES DE JOINVILLE

JOINVILLE – SC
2020

JANINE GARCIA

**DETERMINAÇÃO DOS FATORES INFLUENTES NO CONSUMO DE ÁGUA EM
RESIDÊNCIAS UNIFAMILIARES DE JOINVILLE**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Civil da Universidade do Estado de Santa Catarina – UDESC, como requisito parcial para obtenção do título de Mestre em Engenharia Civil.

Orientadora: Dra. Elisa Henning
Coorientadora: Dra. Andreza Kalbusch

**JOINVILLE – SC
2020**

**Ficha catalográfica elaborada pelo programa de geração automática da
Biblioteca Setorial do CCT/UDESC,
com os dados fornecidos pelo(a) autor(a)**

Garcia, Janine

Determinação dos fatores influentes no consumo de água em residências unifamiliares de Joinville / Janine Garcia. -- 2020.

158 p.

Orientadora: Elisa Henning

Coorientadora: Andreza Kalbusch

Dissertação (mestrado) -- Universidade do Estado de Santa Catarina, Centro de Ciências Tecnológicas, Programa de Pós-Graduação em Engenharia Civil, Joinville, 2020.

1. Fatores influentes. 2. Consumo de água. 3. Residências unifamiliares. 4. Joinville. I. Henning, Elisa. II. Kalbusch, Andreza. III. Universidade do Estado de Santa Catarina, Centro de Ciências Tecnológicas, Programa de Pós-Graduação em Engenharia Civil. IV. Título.

**Determinação dos Fatores Influentes no Consumo de Água em Residências
Unifamiliares de Joinville**

por

Janine Garcia

Esta dissertação foi julgada adequada para obtenção do título de

MESTRA EM ENGENHARIA CIVIL

Área de concentração em “Engenharia Urbana e da Construção Civil”
e aprovada em sua forma final pelo

CURSO DE MESTRADO ACADÊMICO EM ENGENHARIA CIVIL
DO CENTRO DE CIÊNCIAS TECNOLÓGICAS DA
UNIVERSIDADE DO ESTADO DE SANTA CATARINA.

Banca Examinadora:

ASSINADO DIGITALMENTE

Profa. Dra. Elisa Henning
CCT/UDESC
(Orientadora/Presidente)

VIA VIDEOCONFERÊNCIA

Prof. Dr. Enedir Ghisi
UFSC

VIA VIDEOCONFERÊNCIA

Profa. Dra. Inês Osório de Castro
Meireles
Universidade de Aveiro

Joinville, SC, 24 de julho de 2020.

AGRADECIMENTOS

Agradeço primeiramente a Deus por me conceder saúde, força e sabedoria para alcançar meus objetivos e não desanimar diante das dificuldades. Agradeço também à Nossa Senhora Aparecida pela sua intercessão e proteção divina e por estar sempre à minha frente me guiando pelo caminho correto.

Aos meus pais, Sônia Maria Garcia e Valdir Garcia, pelo amor incondicional, compreensão nos momentos de ausência e ensinamentos que jamais esquecerei.

Aos meus irmãos Jayne Garcia, Valdicéia Garcia e Vilson Garcia, pela fiel amizade e por sempre estarem ao meu lado me incentivando e dando suporte nos momentos difíceis.

Ao meu namorado Carlos Henrique Maria por todo amor, atenção e companheirismo. Aos meus sogros José Antônio Maria e Sandra Oesterreich Maria por todo apoio e conselhos recebidos.

À minha orientadora Elisa Henning e coorientadora Andreza Kalbusch, por compartilharem seus conhecimentos e pelo apoio no desenvolvimento da pesquisa. Aos membros da banca, pela disponibilidade em avaliar o presente trabalho.

À Companhia Águas de Joinville (CAJ) e Prefeitura Municipal de Joinville (PMJ) pela concessão de dados, além da empresa Docol Metais Sanitários e instituições de ensino Instituto Federal Catarinense (IFC), Instituto Federal de Santa Catarina (IFSC) e Universidade Federal de Santa Catarina (UFSC) pelo auxílio na aplicação dos questionários. Um agradecimento especial a cada pessoa que se propôs a responder o questionário.

Ao Programa de Pós-Graduação em Engenharia Civil da UDESC (PPGEC) pelas oportunidades de crescimento profissional e pessoal durante o mestrado como participação em congressos e visitas técnicas. À Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina - FAPESC pelo apoio financeiro por meio da bolsa de estudo concedida.

Por fim, agradeço a todos que direta ou indiretamente colaboraram para realização deste estudo.

RESUMO

A água é um recurso essencial à vida e a sua demanda cresceu consideravelmente nos últimos anos devido ao crescimento populacional, desenvolvimento econômico e mudanças nos padrões de consumo. Além disso, o abastecimento urbano é o segundo maior consumidor de água no Brasil. O objetivo deste estudo é avaliar os fatores que influenciam o consumo de água em residências unifamiliares na cidade de Joinville/SC, mediante a análise de variáveis construtivas, socioeconômicas e hábitos de consumo dos usuários. A primeira etapa do estudo consistiu em efetuar um planejamento amostral. Foi definida uma amostra de 400 residências com amostragem não probabilística por cotas, cuja característica de controle foi o número de unidades residenciais em cada região de abrangência das subprefeituras de Joinville. Os dados foram obtidos com auxílio de questionário abrangendo 46 questões relacionadas à localização do imóvel e características do entorno, aspectos socioeconômicos e demográficos, características construtivas e equipamentos instalados na edificação além dos hábitos de consumo dos usuários. Com o auxílio do software R, foi efetuada a verificação de sazonalidade determinística nas séries temporais de consumo de água, além de análise de correlação entre as variáveis. Os resultados indicaram que existe sazonalidade na série temporal de consumo médio de água, portanto o consumo de água nas residências unifamiliares avaliadas é afetado pelos diferentes meses, estações e épocas do ano. A análise de correlação apontou que, das 57 variáveis independentes, quatorze tiveram relação numérica com o consumo *per capita*. Essas variáveis foram utilizadas na elaboração de modelos de regressão linear múltipla, regressão robusta e regressão geograficamente ponderada. A amostra de 400 residências foi então dividida aleatoriamente em 70% (280 casas) para ajuste e 30% (120 casas) para validação. Os modelos foram validados pela estimativa do consumo *per capita* de água. Apenas 4,17% das observações ficou fora do intervalo de predição do modelo de regressão linear múltipla enquanto no caso da regressão robusta, esse percentual foi de 6,67%. As variáveis estatisticamente significativas na explicação do consumo *per capita* foram: número de moradores, propriedade da residência, área construída, presença de piscina e uso de balde para lavagem de áreas externas. Concluiu-se pela análise da regressão geograficamente ponderada que existe variabilidade espacial nas características construtivas e variáveis demográficas na área de estudo. O estudo dos padrões de consumo apresentados neste trabalho oferece suporte aos gestores públicos para a implementação de medidas e incentivos direcionados de acordo com as necessidades de cada região da cidade. Dessa forma, elaborar políticas públicas com base em dados científicos pode representar uma oportunidade valiosa de tornar as estratégias mais eficazes na consolidação do uso sustentável de água em edificações.

Palavras-chave: Fatores influentes. Consumo de água. Residências unifamiliares. Joinville.

ABSTRACT

Water is an essential resource for life and its demand has grown considerably in recent years due to population growth, economic development and changes in consumption patterns. In addition, urban supply is the second largest consumer of water in Brazil. This study aims to identify the factors that may influence water consumption in single-family households in the city of Joinville, Southern Brazil, by the analysis of constructive characteristics, socioeconomic variables and water-use habits. The first step was to perform a sample planning. Hence, a non-probabilistic sample of 400 households was defined by quota sampling, whose characteristic was the number of dwellings in each region of Joinville. Data were collected by applying a 46-question survey related to the surrounding infrastructure, socio-economic and demographic characteristics, constructive features, installed plumbing fixtures, and water-use habits. Seasonality and correlation analyses were carried out using the R software. The results indicated that there is seasonality in the average water consumption time series, therefore water consumption in single-family households is affected by seasons and periods of the year. Correlation analysis showed that 14 out of 57 independent variables had a numerical relationship with per capita water consumption. These variables were used in multiple linear regression, robust regression and geographically weighted regression (GWR) models. The sample of 400 residences was then randomly divided into 70% (280 houses) for model adjustment and 30% (120 houses) for validation. After obtaining the models, water consumption was estimated. Only 4.17% of the observations did not abide by the prediction intervals of multiple linear regression while in the case of robust regression, this percentage was 6.67%. The statistically significant variables that explain per capita water consumption in single-family households were: number of residents, household ownership, built area, swimming pool, and use of bucket for washing outside areas. GWR model showed that there is spatial variability in construction characteristics and demographic variables across the study area. The investigation of consumption patterns presented here supports public managers in the implementation of measures and incentives tailored to the needs of each region of the city. Thus, science-based public policies can represent a valuable opportunity to make strategies more effective in increasing sustainable water use in buildings.

Keywords: Regression models. Influential factors. Per capita water consumption. Single-family households. Joinville.

LISTA DE FIGURAS

Figura 1 – Resultados da regressão geograficamente ponderada.....	19
Figura 2 – Vista superior das edificações de cada grupo	24
Figura 3 – Fluxograma dos procedimentos metodológicos	52
Figura 4 – Localização geográfica do objeto de estudo	53
Figura 5 – Mapa das regiões de abrangência das subprefeituras de Joinville	57
Figura 6 – Mapa de Joinville com a localização das residências do estudo piloto	60
Figura 7 – Mapa de Joinville com a localização das 400 residências unifamiliares ..	71
Figura 8 – Tipologia das residências.....	74
Figura 9 – Renda familiar mensal da amostra.....	75
Figura 10 – Renda domiciliar em Joinville no ano de 2010	75
Figura 11 – Grau de escolaridade do responsável pela conta de água da amostra..	76
Figura 12 – Grau de escolaridade da população de Joinville com 25 anos ou mais de idade no ano de 2010.....	77
Figura 13 – Equipamentos economizadores de água	79
Figura 14 – Série temporal das médias e medianas do consumo <i>per capita</i> de água	80
Figura 15 – Boxplots do consumo médio e mediano de água em relação aos meses do ano	82
Figura 16 – Histograma das médias e medianas do consumo de água (litros/pessoa/dia)	83
Figura 17 – Variação no consumo <i>per capita</i> de água das 400 residências	86
Figura 18 – <i>Boxplot</i> do consumo <i>per capita</i> de água em relação às regiões de Joinville	87
Figura 19 – Boxplot da variável dependente consumo <i>per capita</i> (litros/pessoa/dia)	89
Figura 20 – Histograma da variável dependente consumo <i>per capita</i> (litros/pessoa/dia)	89
Figura 21 – Histograma da variável LN consumo <i>per capita</i>	90
Figura 22 – <i>Boxplot</i> e histograma da variável consumo <i>per capita</i> do conjunto de treinamento	93
Figura 23 – Resíduos <i>versus</i> valores ajustados da regressão linear múltipla	95
Figura 24 – Probabilidade normal dos resíduos da regressão linear múltipla	96
Figura 25 – Distância de Cook dos resíduos da regressão linear múltipla	97

Figura 26 – Resíduos <i>versus</i> valores ajustados da regressão robusta	99
Figura 27 – Probabilidade normal dos resíduos da regressão robusta	100
Figura 28 – <i>Boxplot</i> e histograma da variável consumo <i>per capita</i> do conjunto de validação	101
Figura 29 – Relação entre existência de coleta de esgoto e consumo <i>per capita</i> de água	104
Figura 30 – Relação entre número de banheiros e consumo <i>per capita</i> de água ...	106
Figura 31 – Relação entre idade do imóvel e consumo <i>per capita</i> de água.....	106
Figura 32 – Relação entre número de moradores e consumo <i>per capita</i> de água..	107
Figura 33 – Relação entre propriedade da residência e consumo <i>per capita</i> de água	108
Figura 34 – Relação entre área do imóvel e consumo <i>per capita</i> de água	109
Figura 35 – Relação entre piscina e consumo <i>per capita</i> de água.....	110
Figura 36 – Relação entre equipamentos para lavagem de áreas externas e consumo <i>per capita</i>	110
Figura 37 – GWR Coeficientes de determinação ao longo da área de estudo.....	112
Figura 38 – Resultado GWR: número de moradores	114
Figura 39 – Resultado GWR: Propriedade da residência.....	114
Figura 40 – Resultado GWR: Área total do imóvel.....	115
Figura 41 – Resultado GWR: Presença de piscina	116

LISTA DE TABELAS

Tabela 1 – Estudos relativos à modelagem espacial e fatores influenciadores no consumo de água.....	34
Tabela 2 – Análise de variância (ANOVA)	44
Tabela 3 – Distribuição da amostra por região e cota populacional	57
Tabela 4 – Distribuição da amostra por cotas	71
Tabela 5 – Análise exploratória das 400 residências	73
Tabela 6 – Análise das variáveis dicotômicas	78
Tabela 7 – Estatística descritiva da variável consumo <i>per capita</i> de água	81
Tabela 8 – Probabilidades de significância da ANOVA e dos testes Bartlett, Levene e Kruskal-Wallis.....	84
Tabela 9 – Estatística descritiva da variável consumo <i>per capita</i>	87
Tabela 10 – Estatística descritiva da variável LN consumo <i>per capita</i>	91
Tabela 11 – Variáveis independentes correlacionadas com o Consumo <i>Per Capita</i>	92
Tabela 12 – Estatística descritiva da variável consumo per capita do conjunto de treinamento	93
Tabela 13 – Estatísticas e coeficientes do modelo de regressão linear múltipla.....	94
Tabela 14 – Estatísticas e coeficientes do modelo de regressão robusta	97
Tabela 15 – Estatística descritiva do peso das observações no modelo de regressão robusta	98
Tabela 16 – Observações com menor peso no modelo de regressão robusta	98
Tabela 17 – Estatística descritiva da variável consumo <i>per capita</i> do conjunto de validação	101
Tabela 18 – Indicadores da validação dos modelos de regressão linear múltipla e regressão robusta	102
Tabela 19 – Estimativa dos coeficientes do modelo GWR.....	113
Tabela 20 – Valores mínimo e máximo de <i>VIF – Variance Inflation Factor</i> do modelo GWR	113
Tabela 21 – Indicadores da validação do modelo GWR	116

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
AIC	<i>Akaike Information Criterion</i> (Critério de Informação de Akaike)
ANA	Agência Nacional de Águas
ANOVA	<i>Analysis of Variance</i> (Análise de Variância)
CAJ	Companhia Águas de Joinville
GWR	<i>Geographically Weighted Regression</i> (Regressão Geograficamente Ponderada)
IBGE	Instituto Brasileiro de Geografia e Estatística
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
IPPUJ	Instituto de Pesquisa e Planejamento Urbano de Joinville
MAE	<i>Mean Absolute Error</i> (Erro absoluto médio)
MAPE	<i>Mean Absolute Percentage Error</i> (Erro absoluto médio percentual)
OLS	<i>Ordinary Least Squares</i> (Mínimos quadrados ordinários)
RMSE	<i>Root Mean Squared Error</i> (Raiz quadrada do erro médio)
SEPUD	Secretaria de Planejamento Urbano e Desenvolvimento Sustentável
SM	Salário Mínimo
SNIS	Sistema Nacional de Informações sobre Saneamento

SUMÁRIO

1	INTRODUÇÃO	12
1.1	JUSTIFICATIVA.....	13
1.2	OBJETIVOS	14
1.2.1	Objetivo Geral	14
1.2.2	Objetivos Específicos	15
1.3	SELEÇÃO DA ÁREA DE ESTUDO	15
2	REVISÃO DE LITERATURA.....	16
2.1	FATORES QUE INFLUENCIAM O CONSUMO DOMÉSTICO DE ÁGUA....	16
2.1.1	Aspectos demográficos e características construtivas	16
2.1.2	Status socioeconômico dos moradores.....	25
2.1.3	Fontes alternativas de abastecimento	27
2.1.4	Equipamentos economizadores de água	29
2.1.5	Hábitos de consumo e atitudes de conservação de água	31
2.1.6	Modelagem espacial.....	34
2.2	ANÁLISE ESTATÍSTICA	35
2.2.1	Estatística descritiva dos dados	36
2.2.2	Sazonalidade determinística em séries temporais	37
2.2.3	Diferença de médias de duas populações independentes	38
2.2.4	Verificação da hipótese de normalidade	39
2.2.5	Análise de correlação.....	40
2.2.6	Regressão linear múltipla	41
2.2.7	Regressão robusta	47
2.2.8	Regressão <i>stepwise</i>	48
2.2.9	Regressão geograficamente ponderada	49
2.2.10	Validação do modelo.....	50
2.3	SÍNTESE DA REVISÃO E SUA INFLUÊNCIA SOBRE O MÉTODO ADOTADO.....	50
3	MÉTODO.....	52
3.1	DESCRIÇÃO DO OBJETO DE ESTUDO.....	52
3.2	TAMANHO DA AMOSTRA	55
3.3	AMOSTRAGEM.....	56
3.4	ELABORAÇÃO DO QUESTIONÁRIO	58
3.5	ESTUDO PILOTO.....	59
3.6	COLETA DE DADOS.....	60
3.7	MODELAGEM ESTATÍSTICA	61
3.7.1	Variáveis dependentes e independentes.....	62
3.7.2	Análise exploratória, sazonalidade e correlação	62

3.7.3	Modelos de regressão	65
4	RESULTADOS E DISCUSSÃO	70
4.1	CARACTERIZAÇÃO DA AMOSTRA.....	72
4.2	ANÁLISE DO CONSUMO DE ÁGUA	80
4.2.1	Verificação da sazonalidade determinística.....	80
4.2.2	Comparação entre consumo médio de água de 24 e 48 meses	84
4.2.3	Variação espacial.....	85
4.2.4	Análise exploratória	87
4.3	FATORES RELEVANTES NO CONSUMO PER CAPITA DE ÁGUA.....	91
4.3.1	Correlação entre variáveis	91
4.3.2	Modelos de regressão linear múltipla e robusta.....	92
4.3.3	Discussão dos fatores	103
4.4	PADRÕES ESPACIAIS NO CONSUMO PER CAPITA DE ÁGUA	111
5	CONCLUSÕES	118
5.1	LIMITAÇÕES DA PESQUISA.....	120
5.2	SUGESTÕES PARA TRABALHOS FUTUROS.....	121
	REFERÊNCIAS	122
	APÊNDICES.....	135
	APÊNDICE A – MODELO DE QUESTIONÁRIO	135
	APÊNDICE B – DICIONÁRIO DE VARIÁVEIS	146
	APÊNDICE C – VALIDAÇÃO DOS MODELOS.....	153

1 INTRODUÇÃO

A água é um dos recursos vitais mais importantes da Terra: não há vida, produção econômica e meio ambiente sem ela; não há atividade humana que não dependa de água (HLPW¹, 2016; SAVENIJE, 2002). A fim de gerenciar a escassez de água e aumentar a disponibilidade hídrica, as primeiras medidas tomadas visavam à gestão da oferta: construir redes de abastecimento cada vez maiores e buscar mananciais em locais mais distantes (SILVA; TAMAKI; GONÇALVES, 2006). Entretanto, quando esgotadas boa parte das possibilidades e à medida que as áreas urbanas aumentavam de tamanho e as precipitações diminuía, tornou-se cada vez mais difícil manter o equilíbrio entre a procura e a oferta de água (ALI; SHAFIEE; BERGLUND, 2017).

Em 2010, a demanda mundial por água era em torno de 4.600km³/ano e calcula-se que esta irá aumentar de 20 a 30% até 2050, atingindo um volume entre 5.500 e 6.000km³/ano (BUREK *et al.*, 2016). No caso específico do Brasil, o país vem enfrentando uma crise hídrica, de acordo com a Agência Nacional de Águas (2018). O número de pessoas afetadas por estiagem passou de 9,6 milhões em 2015 para 37,9 milhões em 2017 e um dos impactos dessa seca foi o racionamento de água em muitas regiões do país (ANA, 2018).

Nesse contexto, a gestão da demanda de água surge como um paradigma promissor para a administração sustentável dos recursos hídricos. Ela se concentra em reduzir as demandas de água por meio de campanhas educativas, aumento da tarifa, incentivos e descontos para tecnologias que auxiliem na conservação de água, além de regulamentações específicas que visem diminuir o desperdício nas edificações (ALI; SHAFIEE; BERGLUND, 2017).

Sendo assim, a economia alcançada pelos programas de gestão da demanda de água pode ter um impacto significativo sobre o sistema de abastecimento. No entanto, esse benefício contínuo não vem sendo frequentemente considerado, mas é substancial, por exemplo, ao reduzir a demanda de água potável e, consequentemente, a carga de efluentes no sistema (WILLIS *et al.*, 2013).

Conforme Makki *et al.* (2015), o consumo residencial de água representa um componente significativo na demanda total de água no contexto urbano. Em razão

¹ HLPW – High Level Panel on Water.

disso, uma compreensão aprofundada do consumo doméstico de água e seus fatores determinantes permite a criação e implementação de estratégias, programas e políticas melhor orientados e mais eficazes na conservação desse recurso vital.

1.1 JUSTIFICATIVA

O desenvolvimento de pesquisas na área de gestão hídrica tem sido amplamente impulsionado pelo crescimento populacional, desenvolvimento econômico e mudanças nos padrões de consumo. Fatores como esses vêm aumentando a demanda de água a uma taxa de aproximadamente 1% ao ano e, conseqüentemente, tornaram a sua gestão e uso racional assuntos de relevância na criação de políticas públicas (WWAP², 2018).

No contexto do Brasil, o abastecimento urbano é responsável por 23,8% da quantidade total de água retirada, ficando atrás apenas do setor agrícola (ANA, 2018). Essa situação demonstra a necessidade de um gerenciamento efetivo para a conservação e uso racional de água em edificações urbanas. Por isso é de fundamental importância estudar e compreender as variáveis que afetam o consumo residencial de água a fim de fornecer uma base concreta de dados para auxiliar na elaboração de políticas públicas.

Estudos revelam que o consumo residencial de água é influenciado significativamente por aspectos socioeconômicos dos moradores. Alguns fatores apontados nas pesquisas estão relacionados à renda familiar, número total de moradores e nível de escolaridade (FAN *et al.*, 2017; DIAS; KALBUSCH; HENNING, 2018; SANT'ANA; MAZZEGA, 2018). O envelhecimento da população e a quantidade de mulheres na família são parâmetros sociais que também podem interferir diretamente no uso de água (SCHLEICH; HILLENBRAND, 2009; HUSSIEN; MEMON; SAVIC, 2016).

Quanto às características construtivas das edificações, a área construída, quantidade de banheiros e idade do edifício são aspectos que afetam o consumo de água (MATOS *et al.*, 2013; MATOS *et al.*, 2014; CRUZ *et al.*, 2017). O estudo de Silva (2010) comprova a hipótese de que a presença de sistema de medição individualizada de água reduz significativamente o consumo em edifícios residenciais.

² WWAP – World Water Assessment Programme

O uso de equipamentos economizadores de água é uma boa estratégia para reduzir o consumo de água em edificações, pois além de gerar economias substanciais de água e esgoto, ainda leva à redução do impacto ambiental devido à menor energia incorporada ao longo do ciclo de vida desses dispositivos (MARINOSKI; RUPP; GHISI, 2018).

Kontokosta e Jain (2015) constataram uma relação estatisticamente significativa entre o uso de água e energia, mostrando que um aumento no consumo de energia elevou o consumo de água. Da mesma forma, o custo mais baixo pela água consumida pode elevar o consumo (ROMANO; SALVATI; GUERRINI, 2014, 2016; CRUZ *et al.*, 2017). Ainda sobre tarifa de água, quanto maior o status socioeconômico dos usuários, menor é a influência dessa variável sobre o consumo (FAN *et al.*, 2017; SANT'ANA; MAZZEGA, 2018). Além disso, Willis *et al.* (2011) mostraram a importância de atitudes ambientais e de conservação de água no comportamento de consumo. Inclusive, para Manouseli, Anderson e Nagarajan (2018), é necessário incluir práticas de uso da água nos modelos de previsão da demanda desse recurso. Os autores salientam ainda a disponibilidade limitada na literatura atual de evidências sobre os diversos fatores que afetam o consumo doméstico de água, tanto em condições normais como de estiagem.

Dessa forma, o diferencial da presente pesquisa em relação às demais citadas nesta seção reside no fato de que no Brasil ainda há poucos estudos que investiguem os parâmetros determinantes no consumo de água em edificações unifamiliares. Além disso, a inclusão dos hábitos de consumo dos usuários nos modelos de regressão desse estudo é outro tópico pouco explorado na literatura.

1.2 OBJETIVOS

1.2.1 Objetivo Geral

O objetivo geral deste estudo é determinar os fatores que influenciam o consumo de água em residências unifamiliares na cidade de Joinville/SC por meio de modelagem estatística incluindo variáveis construtivas, socioeconômicas e hábitos de consumo dos usuários.

1.2.2 Objetivos Específicos

Os objetivos específicos deste trabalho são:

- Fornecer uma estimativa do consumo médio *per capita* de água para o tipo de edificação estudada;
- Investigar a presença de sazonalidade determinística no consumo de água das habitações avaliadas, considerando os meses, estações e épocas do ano;
- Propor modelos de regressão que relacionem o consumo de água com variáveis construtivas das edificações, aspectos socioeconômicos e hábitos de consumo dos usuários;
- Verificar o desempenho dos modelos pelo poder de assertividade na predição do consumo de água;
- Analisar espacialmente os padrões de consumo hídrico de acordo com as regiões da cidade.

1.3 SELEÇÃO DA ÁREA DE ESTUDO

O município de Joinville foi escolhido como área de estudo por ser a maior cidade do estado de Santa Catarina (SEPUD³, 2018). Também, a disponibilidade da concessionária de abastecimento de água Companhia Águas de Joinville e da Prefeitura Municipal em fornecer dados para a realização da pesquisa foi outro fator relevante na seleção do objeto de estudo.

³ SEPUD – Secretaria de Planejamento Urbano e Desenvolvimento Sustentável

2 REVISÃO DE LITERATURA

Este capítulo tem o propósito de apresentar uma revisão da literatura relativa ao tema da pesquisa. Inicialmente, são descritos estudos que avaliaram as variáveis envolvidas no consumo de água em edificações residenciais, com uma breve explanação do método adotado e resultados obtidos. Os fatores citados são abordados na proposta da pesquisa e envolvem: aspectos socioeconômicos e características construtivas, tarifa pelo uso de água, fontes alternativas de abastecimento, equipamentos economizadores de água, hábitos de consumo dos usuários e suas atitudes para conservação da água. Também são apresentadas pesquisas que salientam a importância de atributos geográficos e interações espaciais na análise do consumo de água. Por fim, são descritas as principais análises estatísticas da presente pesquisa, bem como a forma como são utilizadas em outros estudos.

2.1 FATORES QUE INFLUENCIAM O CONSUMO DOMÉSTICO DE ÁGUA

2.1.1 Aspectos demográficos e características construtivas

Segundo Gomes (2009), o consumo de água de uma população varia conforme a região, a cidade, e no caso de uma mesma cidade, varia também segundo o setor de distribuição. De acordo com o autor, os principais fatores que influenciam a quantidade de água a ser distribuída por pessoa e por dia são: clima, padrão de vida, hábitos da população, sistema de fornecimento (medição individualizada de água), qualidade da água fornecida, custo da água, pressão na rede de abastecimento, extensão do serviço de esgotos, extensão de áreas ajardinadas e perdas no sistema.

Estudos sobre as variáveis influentes no consumo doméstico de água vêm sendo desenvolvidos em diversos países. Schleich e Hillenbrand (2009) analisaram o impacto de fatores econômicos, ambientais e sociais sobre a demanda *per capita* de água em 592 áreas de abastecimento de água na Alemanha. Os autores utilizaram dados das regiões de fornecimento de água para elaborar um modelo de regressão log-log, em que todas as variáveis da equação estavam na forma logarítmica. Foi definida como variável resposta do modelo o consumo médio de água, em litros por pessoa por dia, de residências unifamiliares e edifícios multifamiliares no ano de 2003.

Além do preço da água potável e esgoto, renda líquida *per capita*, número de membros da família, as variáveis explicativas avaliaram os efeitos da idade da população, presença de poço nos domicílios, percentual de casas unifamiliares, número de dias com chuva superior a 1mm e temperatura média nos meses de verão. Os autores concluíram que a demanda de água na Alemanha é bastante inelástica, ou seja, responderia menos aos aumentos nos preços da água do que na maioria dos outros países da *Organization for Economic Cooperation and Development* – OECD. A renda familiar mais baixa aliada aos preços mais altos da água nos novos estados federais explicam por que o consumo *per capita* de água nesses estados é 30% menor do que nos antigos estados federais. Além disso, um aumento no tamanho do agregado familiar ou no percentual de poços nos domicílios diminuiria a demanda de água *per capita* do sistema público. Outra descoberta importante é que o envelhecimento da população influencia diretamente o uso da água, ou seja, o aumento na idade dos habitantes estava associado ao crescimento no consumo de água. Por fim, os resultados do estudo mostraram que, no que concerne ao impacto das variáveis climáticas, há algumas evidências de que padrões de chuva afetam o consumo de água e a temperatura não exerce influência. Portanto, análises baseadas em séries temporais poderiam fornecer visões adicionais sobre os efeitos das mudanças climáticas no consumo de água.

Sobre os aspectos construtivos da edificação, Silva (2010) avaliou os resultados alcançados com a implantação do sistema de medição individualizada de água em prédios populares de Salvador que anteriormente possuíam medição coletiva. Quinze prédios fizeram parte do estudo, totalizando 180 apartamentos. Dados do histórico de consumo de água referentes a 24 meses que antecederam e 24 meses que sucederam a individualização da medição foram disponibilizados pela concessionária local. O consumo médio mensal no período passou de 18m³ por apartamento para 14m³ por apartamento com a implantação da medição individualizada de água. Dessa forma, houve uma redução de 22% no consumo dos apartamentos avaliados, para um intervalo de confiança de 95%. Portanto, o resultado obtido reforça a hipótese de que o sistema de medição individualizada de água induz ao menor consumo em edifícios residenciais que adotam esse tipo de medição.

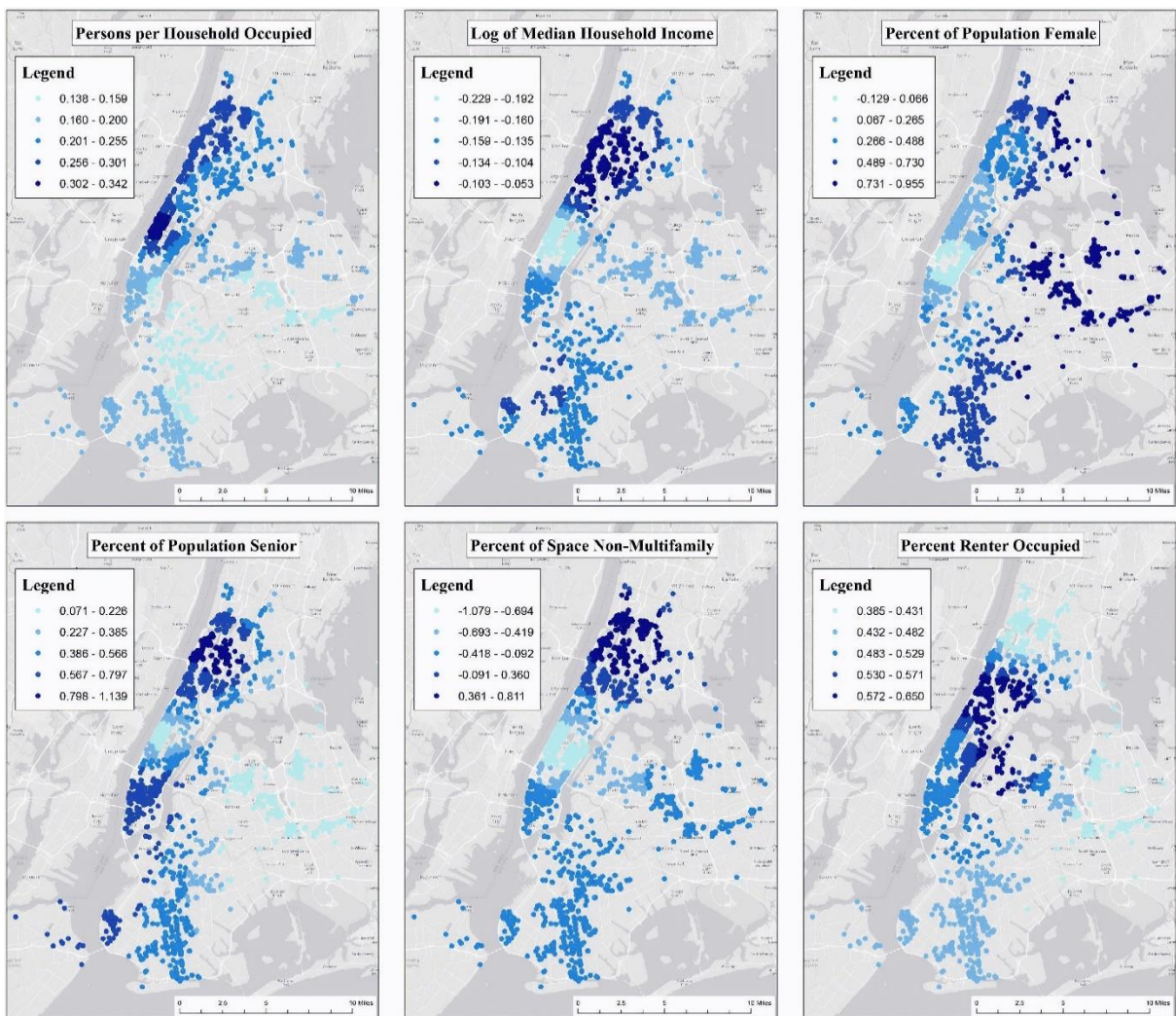
Uma pesquisa em três regiões no norte de Portugal (Vila Real, Valpaços e Porto) estabeleceu padrões no uso final de água por equipamento hidrossanitário em 36 domicílios, com 109 participantes no total. O método de amostragem adotado foi a

probabilística estratificada. Foi solicitado aos usuários que anotassem a frequência e tempo de utilização das torneiras e bacias sanitárias em um período de 7 dias, em março de 2011. No caso do lavatório e pia da cozinha, havia um registro detalhado do tipo de utilização, como escovar os dentes, lavar as mãos, etc. O estudo também analisou possíveis relações desses padrões com características sociodemográficas (número de residentes, presença ou ausência de idosos e crianças, nível de renda e nível educacional). Inicialmente, uma análise estatística descritiva das variáveis foi efetuada com a finalidade de averiguar as respectivas médias e frequências. Por fim, foi efetuada uma análise de correlação com o teste de Spearman com o propósito de identificar a influência de características sociodemográficas nos usos finais de água. Os resultados obtidos refletiram diferenças nos estilos de vida urbano e rural uma vez que Valpaços foi a cidade que registrou o menor número de utilizações nos dispositivos monitorados. Correlações significativas foram encontradas nos seguintes casos: entre a área da residência e o número de utilizações do lavatório e bacia sanitária; entre a presença de crianças no domicílio e o uso da máquina de lavar louça; entre o nível de renda e o número de utilizações da bacia sanitária, banheira, máquina de lavar roupa e máquina de lavar louça (MATOS *et al.*, 2013; MATOS *et al.*, 2014).

Na cidade de Nova York, um estudo sobre os fatores influentes e padrões espaciais no consumo de água foi aplicado em 2300 edifícios multifamiliares. Foram coletados dados referentes ao consumo de água e energia do banco de dados *Local Law 84 – LL84*, bem como informações geográficas dos lotes e sobre o uso do solo do banco de dados *Primary Land Use Tax Output – PLUTO*. A construção do conjunto de dados utilizado nessa análise consistiu em combinar as informações do LL84 e PLUTO com dados demográficos e socioeconômicos dos bairros do Censo de 2010 e da *American Community Survey – ACS*, de 2006 a 2010. Na estimativa dos parâmetros determinantes no uso de água foram utilizadas duas abordagens de modelagem. A primeira envolveu um modelo de regressão linear multivariada com erros padrão robustos, a qual possibilitou uma compreensão da magnitude e significância de fatores individuais de construção, ocupantes e fatores dos bairros que influenciam no consumo de água. A segunda abordagem aplicou um modelo de regressão geograficamente ponderada a fim de estudar a variação espacial na intensidade do uso da água e os efeitos geográficos das variáveis explicativas. Os resultados da análise de regressão robusta mostraram que a ocupação, o tamanho e a idade do edifício, tipo de morador (proprietário ou inquilino), variáveis demográficas

e renda familiar têm impacto significativo no consumo de água de edifícios multifamiliares na cidade de Nova York. Além disso, foi observada uma relação estatisticamente significativa entre a intensidade do uso de energia e a intensidade do uso de água. Como pode ser observado na Figura 1, a regressão geograficamente ponderada apontou que o número de pessoas por domicílio, renda familiar, população feminina e idosa, percentual de espaços não-familiares e percentual de inquilinos foram as variáveis com alta variabilidade espacial (KONTOKOSTA; JAIN, 2015).

Figura 1 – Resultados da regressão geograficamente ponderada



Fonte: Kontokosta e Jain (2015).

Hussien, Memon e Savic (2016) investigaram a influência das características dos usuários e dos domicílios no consumo *per capita* de água na cidade de Duhok, localizada no noroeste do Curdistão Iraquiano. Foi elaborado um questionário detalhado, o qual incluiu mais de 40 perguntas relativas às características

demográficas dos moradores como: número de crianças, idosos adultos do sexo masculino e feminino. Além disso, o estudo identificou características das habitações e da família, incluindo: tipologia, área construída total, área de jardim, número de quartos, número de pavimentos e renda mensal. Foram investigadas ainda questões sobre a frequência, duração de uso e vazão de aparelhos como chuveiro, lavatório, bacia sanitária, torneira e equipamentos para lavagem de roupa, louça, veículo, limpeza da casa, preparação de alimentos e rega de jardim. O questionário foi respondido por moradores de 407 domicílios selecionados aleatoriamente em fevereiro de 2015, sendo 92% residências e 8% apartamentos. A amostra foi dividida em dois subconjuntos: desenvolvimento e validação do modelo. Então 70% dos dados foram usados para ajustar o modelo enquanto os 30% restantes foram reservados para fins de teste. Utilizando o conjunto de dados obtido, foram desenvolvidos 24 modelos estatísticos por meio de duas técnicas: regressão linear múltipla – *Stepwise* e Regressão Polinomial Evolutiva (*Evolutionary Polynomial Regression*) – RPE. Os modelos foram então comparados através do seu coeficiente de determinação múltipla e correlação entre variáveis. Os valores de R^2 em ambas as técnicas, *Stepwise* e RPE, apresentaram-se relativamente altos (acima de 0,8) na maioria dos casos. As duas abordagens sugerem uma forte influência das características demográficas no consumo *per capita* de água quando os dados foram divididos em grupos de renda (baixa, média e alta).

Posteriormente, a análise de sensibilidade identificou o efeito de cada característica das famílias na demanda de água estimada, concluindo dessa forma que a renda *per capita*, o número de crianças e a quantidade de adultos do sexo masculino foram as variáveis mais influentes nos modelos desenvolvidos. Foram utilizados os modelos de melhor ajuste para prever a demanda futura de água na cidade com base em quatro cenários. Com a análise dos dados, os autores concluíram que: o consumo *per capita* de água cresce com o aumento da renda e diminui com o aumento do número de pessoas na família; a frequência dos usos finais de água cresce com o aumento da renda *per capita*, exceto na bacia sanitária, em que sua frequência de utilização é maior nas famílias de baixa renda do que nos grupos de renda média e alta; o consumo *per capita* diminui com o aumento do número de homens adultos, idosos e crianças, porém é maior com o aumento da quantidade de mulheres nas famílias (HUSSIEN; MEMON; SAVIC, 2016).

Em Hermosillo, no México, buscou-se identificar as variáveis explicativas no consumo doméstico de água através de um modelo de regressão linear múltipla com estimação por mínimos quadrados ordinários (*ordinary least square* – OLS). A amostragem adotada foi do tipo probabilística estratificada, distribuída segundo as *basic geographical areas* – BGAs e o índice de marginalização (muito alto, alto, médio, baixo e muito baixo). Dessa forma, uma amostra de 403 domicílios, com intervalo de confiança de 95%, forneceu informações das características da habitação e dos moradores, custo pelo uso da água e hábitos de consumo dos usuários pela aplicação de um questionário com 65 questões. A análise estatística revelou que as variáveis estatisticamente significativas foram: baixo custo da água, número de banheiros na edificação, medição do serviço de abastecimento, uso de água purificada em galões e número de mulheres na família. O modelo obtido indicou que um aumento no número de banheiros e de mulheres eleva o consumo de água na residência. Da mesma forma, custo mais baixo pela água consumida pode elevar o consumo. Também, habitações cujos moradores consumiam água purificada engarrafada tiveram valores menores de consumo, indicando que o perfil de consumo desses usuários é diferente. Embora em pequena quantidade, sua presença foi estatisticamente significativa na regressão. O modelo obtido foi então testado em quatro residências de diferentes regiões da cidade. Comparando o consumo mensal de água estimado pelo modelo com o consumo medido diretamente nas residências, um grau de aproximação aceitável (78 a 90%) foi alcançado (CRUZ *et al.*, 2017).

Fan *et al.* (2017) investigaram os parâmetros relevantes no consumo doméstico de água em 286 cidades da China a fim de fornecer uma base sólida para a formulação de políticas públicas, destacam os autores. Os dados utilizados na análise foram obtidos de bases estatísticas do governo chinês e incluíram o consumo *per capita* de água (litros/pessoa/dia) do período de 2000 a 2015, como também 15 fatores determinantes no consumo. Essas variáveis compreendiam: fatores climáticos (temperatura e precipitação); preço da água; aspectos socioeconômicos (área urbana, população, produto interno bruto *per capita*, taxa de crescimento populacional, escolaridade e gênero); equipamentos hidráulicos (máquina de lavar roupa e sistema de aquecimento de água para banho) e características do abastecimento e conservação de água (capacidade de abastecimento, taxa de reúso de água e investimentos na conservação de água).

O método estatístico baseou-se em árvore de inferência condicional ou *Conditional Inference Tree* – CIT, que incorpora árvore estruturada de modelos de regressão para identificar a importância hierárquica e contribuição dos fatores. A técnica da floresta aleatória (*random forest*) e o coeficiente de correlação tau de Kendall foram utilizados para quantificar a influência de cada preditor no consumo doméstico de água. Os resultados indicaram que na China o consumo de água dos moradores urbanos foi significativamente afetado pelo clima, *status* socioeconômico, capacidade de abastecimento e atitudes de conservação de água. Os fatores com maior influência sobre o consumo de água também variaram de acordo com as regiões da China. Nas cidades de alto consumo, o consumo *per capita* de água foi altamente afetado pela precipitação, investimentos em conservação de água, presença de sistema aquecedor de água para banho e produto interno bruto *per capita*. Ou autores mencionam que residências com alta renda *per capita* possuem vários equipamentos que podem consumir volume considerável de água. Ademais, residentes com alto *status* socioeconômico geralmente são menos sensíveis ao preço da água (FAN *et al.*, 2017).

Nas cidades de médio e baixo consumo, variáveis como capacidade de abastecimento de água e aspectos socioeconômicos (incluindo produto interno bruto e educação recebida) foram as que mais influenciaram o consumo *per capita* de água. Essas disparidades sugerem que políticas diferentes, segundo o consumo da região, devem ser implementadas na China. Nas cidades de alto consumo, investimentos na conservação de água devem ser mantidos para garantir a redução no uso de água. Por outro lado, nas regiões de baixo consumo, o sistema de fornecimento de água deve ser aperfeiçoado com o intuito de garantir a disponibilidade de água para as necessidades básicas de higiene e saúde (FAN *et al.*, 2017).

A pesquisa de Sant'Ana e Mazzega (2018) também mostrou o impacto de variáveis socioeconômicas e tipologia da construção na demanda interna de água e usos finais. O estudo buscou compreender padrões de consumo doméstico de água de edifícios com diferentes tipologias construtivas, faixas de renda e número de moradores no Distrito Federal, Brasil. Segundo os autores, foi adotado um método simples e de baixo custo, capaz de medir o consumo de água de um tamanho amostral significativo. Durante o período de 1 ano, foram coletados dados dos aparelhos hidrossanitários de 118 moradias. A aquisição de dados envolveu: a) um levantamento completo dos equipamentos hidráulicos e outros dispositivos que consomem água; b)

fiscalização do consumo para estimar os usos finais de água; c) análise profunda do consumo de água com entrevista destinada a entender os hábitos de consumo dos ocupantes. Posteriormente, questionários face-a-face levantaram aspectos socioeconômicos e características construtivas de uma amostra aleatória estratificada de 481 moradias, proveniente de 7 regiões administrativas do Distrito Federal. Os resultados foram agrupados de acordo com a tipologia do edifício e faixa de renda (em salários mínimos – SM): renda alta ou *high income* – HI (acima de 20 SM); renda média a alta ou *mid-high income* - MHI (10-20 SM); renda média a baixa ou *mid-low income* – MLI (5-10 SM) e renda baixa ou *low income* – LI (1-5 SM), como ilustra a Figura 2.

Os modelos de regressão sugeriram que variáveis características da edificação (como área construída e área de jardim), bem como número de residentes e renda familiar afetaram o consumo doméstico de água. Por exemplo, edificações com renda alta possuíam acessórios extra no banheiro, além de mais dispositivos na cozinha e área externa (jardins maiores, varandas, presença de piscina e área de lazer). Quanto maior a gama de dispositivos hidráulicos disponíveis, mais oportunidades de uso de água os residentes têm. Além disso, os moradores com renda alta (HI e MHI) tenderam a usar água por um período maior (torneiras, chuveiros, filtros de água, etc.) e mais frequentemente (descarga da bacia sanitária e máquina de lavar roupa) do que residências de renda baixa (MLI e LI). Sendo assim, em geral, quanto mais alta a renda, maior foi o consumo de água das residências estudadas. Uma explicação dos autores para isso é a percepção de custo da água uma vez que famílias com renda alta podem considerar o custo da água baixo, elevando seu consumo. Conforme os autores, pesquisas futuras em direção a programas de conservação de água podem ser efetuadas considerando o potencial de conservação de água pelo uso de equipamentos e dispositivos hidráulicos eficientes assim como sistemas de água não potável (captação de água da chuva ou reúso de água). Dessa forma, diferentes soluções podem ser encontradas para diferentes tipologias em função da variação na renda familiar (SANT'ANA; MAZZEGA, 2018).

Figura 2 – Vista superior das edificações de cada grupo



Fonte: Adaptado de Sant'Ana e Mazzega (2018).

Também no Brasil, Dias, Kalbusch e Henning (2018) buscaram identificar os fatores influenciadores do consumo de água em edifícios da cidade de Joinville por meio de uma análise de variáveis socioeconômicas dos usuários e construtivas das edificações. A pesquisa foi aplicada em 89 edifícios com padrão construtivo baixo ou médio, abrangendo um total de 3171 unidades habitacionais e 9112 moradores. A análise foi realizada com base em medições mensais do consumo de água durante dois anos, de janeiro de 2015 a dezembro de 2016, e em dados referentes às características do edifício e dos moradores, os quais foram obtidos diretamente com os síndicos dos condomínios. Sendo assim, ao aplicar uma análise de regressão linear

múltipla por meio do método *Stepwise*, as autoras elaboraram dois modelos. O primeiro modelo descreveu 87,41% do consumo bruto de água em edifícios (em m³/mês), enquanto o segundo explicou 79,68% do consumo *per capita* de água em Joinville (em litros/pessoa por dia). Os resultados apontaram que fatores como a distância da propriedade ao centro (km), presença do sistema público de coleta de esgoto, número de apartamentos, número de pavimentos, idade do edifício (anos), número de residentes por apartamento, presença de piscina e valor venal do edifício (R\$) apresentaram efeito positivo no consumo de água, isto é, o consumo de água aumentou. Por outro lado, variáveis como porcentagem de inquilinos (%), presença de sistema alternativo de abastecimento e medição individualizada diminuíram o consumo de água (DIAS; KALBUSCH; HENNING, 2018).

2.1.2 Status socioeconômico dos moradores

Pesquisas como a de Fan *et al.* (2017) e Sant'ana e Mazzega (2018) mostraram que usuários com *status* socioeconômico alto são menos vulneráveis ao custo pelo uso da água e, por esse motivo, consomem mais água do que aqueles com padrão socioeconômico mais baixo.

Garcia (2011) investigou a influência de características socioeconômicas da família e do domicílio sobre o consumo de água. Para isso, foram entrevistados moradores de 147 residências localizadas em uma região de baixa renda em Salvador, Bahia. O questionário aplicado possuía 45 questões, que abordavam as características socioeconômicas das famílias, características do domicílio, informações sobre o consumo de água do domicílio e percepção dos moradores sobre o uso de água. As variáveis avaliadas foram então comparadas com o consumo domiciliar e *per capita*, verificando o grau de dependência entre elas pelos coeficientes de correlação de Pearson e Spearman. Para a população analisada, foi identificada associação significativa entre o consumo residencial de água e aspectos como: número de moradores, número de dormitórios, pontos internos de água e parcela da renda comprometida com o pagamento da conta de água. Além disso, não foi encontrada associação significativa entre o consumo residencial de água e a renda, possivelmente em razão da pequena variação de renda para a população avaliada. Identificou-se ainda que a cobrança de taxa fixa para consumo inferior a 10m³ associada ao desconhecimento por parte dos moradores do consumo praticado na

residência representam um desafio à implementação de ações que visam o uso racional de água em áreas de baixa renda. Aspectos socioeconômicos são importantes para essa parcela da população, sendo assim o custo pelo uso da água de acordo com o volume de água realmente consumido poderia estimular o uso eficiente desse recurso já que 55% das edificações avaliadas encontram-se na faixa de consumo em que é cobrada a tarifa mínima, ou seja, até 10m³/mês.

Na Itália, buscou-se analisar os fatores determinantes na demanda residencial de água com 515 observações de 103 cidades italianas em um período de 5 anos, de 2007 a 2011. Embora dados de nível doméstico sejam preferíveis na estimativa dos parâmetros determinantes no consumo de água, eles não estavam disponíveis para o estudo. Por isso, foram coletados dados a nível comunitário considerando as maiores cidades da Itália, provenientes do Instituto Italiano de Estatística, da organização de consumidores Cittadinanzattiva, do Ministério de Agricultura, Alimentação e Silvicultura da Itália e das empresas que gerenciam o serviço de água em cada província. As informações obtidas incluíam população servida, altitude, precipitação média anual, temperatura média anual, tarifa cobrada pelo consumo de 192 m³ de água, renda média *per capita* dos indivíduos, tipo de propriedade (pública ou não pública) das empresas que gerenciavam o serviço de água e localização das províncias (norte, centro ou sul da Itália). Os autores desenvolveram um modelo linear de efeitos mistos com o método da máxima verossimilhança restrita visto que permitiu obter correlações de efeitos aleatórios para a mesma cidade. Os resultados confirmaram que o aumento da tarifa cobrada aos clientes causa uma redução significativa no consumo de água, enquanto o aumento da renda *per capita* eleva o consumo. Entre as características geográficas e climáticas medidas, os dados demonstraram que apenas a altitude exerceu efeito negativo sobre o consumo, ou seja, houve redução na demanda com o aumento da altitude. Temperatura, precipitação e renda *per capita* não influenciaram a demanda de água. Ademais, a população servida tem um efeito positivo sobre o consumo, de modo que as cidades maiores mostraram demanda residencial de água maior. A propriedade das concessionárias de água em si não teve efeito significativo no consumo de água. Entretanto, os preços da água eram significativamente mais baixos e o consumo de água mais alto em províncias com concessionárias públicas de água do que com empresas de propriedade mista ou privada (ROMANO; SALVATI; GUERRINI, 2014, 2016).

2.1.3 Fontes alternativas de abastecimento

O aumento da demanda de água em muitos países é consequência da urbanização desenfreada e crescimento populacional exponencial e dá origem à escassez de água quando combinado com a insuficiência dos sistemas de abastecimento de água (LEONG *et al.*, 2018). Nesse sentido, a diversificação das fontes atuais de fornecimento de água mostra-se como um meio alternativo para atender a essa demanda. Exemplos disso são o aproveitamento de água pluvial e reúso de água cinza em usos finais que não necessitam de água potável (GHISI; FERREIRA, 2007).

Em Jaffna, um distrito na região nordeste de Sri Lanka, a água subterrânea é a única fonte para suprir as necessidades de consumo doméstico e agrícola. Sendo assim, o estudo de Rajeevan e Mishra (2020) procurou investigar os padrões de consumo doméstico de água e práticas de gerenciamento desse recurso, além de comparar a eficácia de diferentes medidas para reduzir o déficit de água do distrito. Os dados primários foram coletados por meio de questionários de 142 famílias pertencentes a áreas com abundância de água, áreas com escassez de água, bem como zona costeira e urbana. Com base nos resultados, não houve diferença estatisticamente significativa no consumo *per capita* entre as diferentes áreas de estudo. Os resultados também mostraram que uma pessoa de Jaffna consome, em média, 200 litros por dia. Na análise de correlação, os fatores que apresentaram correlação positiva com o consumo de água incluíam tamanho da edificação, fonte de água própria, número de crianças, nível de escolaridade do chefe da família, renda mensal, frequência de banho e acesso contínuo à água. A idade do chefe da família mostrou correlação negativa com o consumo de água do imóvel. Essas variáveis foram utilizadas na criação de um modelo de regressão múltipla, o qual identificou que o tamanho da edificação e fonte de água própria como os preditores mais importantes no consumo de água das famílias da região. O modelo é estatisticamente significativo e explica 30,4% da variação na água total consumida. Outro resultado importante é que mais da metade dos moradores de domicílios investigados declararam que não se dedicam à conservação de água atualmente por terem acesso contínuo à água de sua própria fonte. Uma análise mais aprofundada usando a modelagem do sistema de avaliação e planejamento da água revelou que, se os atuais padrões de consumo continuarem, haverá uma demanda não atendida de 51,7 milhões de metros cúbicos

de água no ano de 2030. Por outro lado, implementando medidas de intervenção como aumento da recarga dos aquíferos, redução do consumo *per capita*, uso de água de fontes alternativas e instalação de plantas de tratamento de água cinza, a demanda de água para o ano de 2030 poderá ser atendida.

Garcia-Cuerva, Berglund e Binder (2016) realizaram uma pesquisa nos Estados Unidos com o intuito de avaliar o potencial de aceitabilidade no reúso de águas residuárias. Um total de 2800 pessoas foram recrutadas pelo GfK *Knowledge Networks* para responder um questionário com 22 questões. Dois conjuntos de testes estatísticos foram conduzidos para avaliar e interpretar os resultados. O teste qui-quadrado de Pearson buscou identificar alguma relação entre as diferentes categorias de resultado e as variáveis demográficas, com uma probabilidade de significância $p < 0,05$. A regressão linear foi utilizada para determinar possíveis relações entre valores decrescentes de *Palmer Drought Severity Index* – PDSI associados a cada estado e a porcentagem de entrevistados com preocupações em relação à escassez de água, comportamentos específicos de conservação de água bem como o apoio ao reúso de águas residuárias. PSDI é um indicador das condições climáticas relevantes para tendências de escassez de água. Os resultados demonstraram que uma pequena porcentagem dos participantes (6,5%) está preocupada com a escassez de água, a maioria da população pratica alguma atitude de conservação de água e 43% da população apoia o reúso de água cinza. A pesquisa identificou que clima, educação e *status* econômico podem influenciar atitudes e comportamentos em relação à água (como conscientização, conservação e suporte ao reúso de águas residuárias). Gênero, idade, valor da última conta mensal de água e localização do morador (conforme as regiões do estudo) não mostraram efeito significativo na aceitação do reúso de água enquanto etnia, nível de educação e renda apresentaram efeito significativo. O aumento nos períodos de seca eleva o número de entrevistados que estão preocupados com o uso da água, que adotam mais atitudes relacionadas à conservação de água e que limitam o uso de água para rega de jardim e gramado. Além disso, incentivos financeiros elevam o interesse dos entrevistados em participar de programas de reúso de água, como redução no custo mensal da água.

Na Malásia, pesquisadores quantificaram o potencial de seis sistemas de captação de água da chuva, dois sistemas de reciclagem de água cinza e um sistema híbrido de águas pluviais e cinzas usando características locais como entradas de um modelo de balanço hídrico. Com a simulação computacional, obtiveram-se os

seguintes resultados: (i) os sistemas de água pluvial supriram mais de 90% da demanda de água não potável; (ii) os sistemas híbridos domésticos usavam principalmente a água da chuva para suprir até 52,5% do total da demanda de água não potável e a demanda restante poderia ser suprida por reúso de água cinza (LEONG *et al.*, 2018).

Em Joinville, uma pesquisa quantificou a redução do volume de água pluvial escoado superficialmente em função do seu aproveitamento em edificações residenciais. Foram dimensionados reservatórios para armazenamento de água pluvial para as residências da cidade, considerando as tipologias casa e prédio. Para a Bacia do Rio Cachoeira, os resultados da simulação computacional indicaram um potencial de economia de água potável de 40,8% quando há uso de água pluvial para descarga de bacias sanitárias e máquina de lavar roupa nas edificações avaliadas. Analisando-se o uso de água pluvial apenas para a descarga de bacias sanitárias, o percentual de economia de água potável foi estimado em 18,5% (CUSTÓDIO, 2017; CUSTÓDIO; GHISI, 2019).

Apesar da influência impactante de políticas públicas na adoção de estratégias para a economia de água em edificações, no Brasil ainda não há uma legislação nacional que regule o uso dessas estratégias. Apenas alguns estados e municípios do país criaram regulamentações locais a fim de tornar obrigatória a implementação de medidas para a conservação de água (PACHECO *et al.*, 2017). Comparando as leis de Florianópolis e Seattle, outro estudo revelou que a legislação de Seattle, ao contrário de Florianópolis, já incorporou diversos programas voltados à sustentabilidade na utilização dos recursos hídricos como aquecimento solar de água, aproveitamento de água da chuva, utilização de equipamentos mais eficientes na cozinha, banheiro e irrigação, planos de tratamento de esgoto, bem como reflorestamento ao redor de córregos urbanos (RIBEIRO *et al.*, 2018).

2.1.4 Equipamentos economizadores de água

Certas atitudes contribuem diretamente na melhoria da gestão dos recursos hídricos como a utilização de equipamentos economizadores de água e a conscientização dos usuários uma vez que geram grande impacto no consumo final de água (ALEXANDRE; KALBUSCH; HENNING, 2017). Como exemplo de aplicação, em seu estudo, Soares (2010) concluiu que a incorporação de ar aos fluxos de água

é eficaz como ação complementar a ser empregada para racionalização e economia de água. Os resultados apontaram uma economia efetiva em torno de 50% no consumo de água, entre a máxima aeração e a nula.

Marinoski, Rupp e Ghisi (2018) avaliaram o benefício ambiental da adoção de estratégias para economia de água em domicílios de baixa renda localizados na região Sul do Brasil. O potencial para economia de água, a energia incorporada e a redução de esgoto foram estimados tanto para as estratégias individualmente quanto em conjunto, considerando as seguintes medidas: (i) aproveitamento de água pluvial; (ii) reúso de água cinza e (iii) instalação de equipamentos economizadores de água (bacia sanitária com duplo fluxo e torneira cromada com arejador). Vinte residências unifamiliares foram selecionadas para análise dos usos finais de água com auxílio de questionários e entrevistas estruturadas. A análise do ganho ambiental dessas estratégias foi realizada utilizando um indicador que inclui energia incorporada, economia de água potável, redução da geração de esgoto e consumo de energia na concessionária de água, além da produção de esgoto durante o ciclo de vida do sistema. O potencial de economia de água potável variou de 21,0% (reúso de água cinza) a 42,9% pela combinação de aparelhos com eficiência hídrica, aproveitamento de água da chuva e reúso de água cinza. Considerando a redução do esgoto doméstico, a maior redução ocorreu no reúso de água cinza combinado com aparelhos economizadores de água (36,8%). A energia incorporada mínima foi de 641,0 MJ (equipamentos economizadores de água) e a máxima de 25.634,6 MJ (combinando aparelhos eficientes, aproveitamento de água pluvial e reúso de água cinza). Os resultados indicaram também que a estratégia com maior benefício ambiental foi o uso de equipamentos economizadores de água, que resultou em economias substanciais de água e redução de esgoto, causando baixo impacto ambiental devido à menor energia incorporada ao longo do ciclo de vida.

Pérez-Urdiales e García-Valiñas (2016) investigaram a relação entre economia de água e fatores como hábitos de conservação de água, uso de aparelhos elétricos eficientes (máquina de lavar louça e máquina de lavar roupa) e equipamentos não elétricos (dispositivos economizadores de água). O estudo abrangeu 1465 domicílios de Granada, na Espanha. As variáveis explicativas consideradas nos modelos estatísticos foram categorizadas em três grupos: características socioeconômicas, atitudes e comportamentos e aspectos políticos. As características socioeconômicas incluíam número de membros da residência, idade e gênero do chefe da família,

escolaridade do responsável pelo domicílio, renda familiar, residentes proprietários ou inquilinos, idade do imóvel e reformas efetuadas nos últimos 5 anos). As atitudes e comportamentos avaliados foram preocupação com o meio ambiente e conhecimento da existência de alguma campanha de redução do consumo de água nos últimos 5 anos. Por fim, considerou-se o preço médio da tarifa de água em 2010 a fim de identificar se o aumento no custo da água poderia encorajar os moradores a investir em tecnologias eficientes de consumo de água e mudar seus hábitos de consumo. A análise estatística mostrou que famílias com alto nível de preocupação ambiental costumam exibir melhores práticas de conservação de água em relação aos equipamentos não elétricos e investem mais em tecnologias elétricas para reduzir o consumo de água. Dessa forma, políticas públicas poderiam focar em campanhas de educação ambiental, elevando a preocupação ambiental principalmente das famílias com menor probabilidade de serem equipadas com dispositivos economizadores de água.

2.1.5 Hábitos de consumo e atitudes de conservação de água

Manouseli, Anderson e Nagarajan (2018) enfatizam a necessidade de incluir práticas de uso da água nos modelos de previsão da demanda desse recurso. Além disso, Lindsay e Supski (2017) demonstraram a importância de avaliar práticas de uso de água a fim de explorar o comprometimento e resistência quanto à economia de água como também investigar a possibilidade de adaptação às mudanças climáticas.

Willis *et al.* (2011) analisaram a influência das atitudes de conservação de água no consumo final de água. Foram selecionadas quatro regiões da cidade de *Gold Coast* na Austrália considerando a diferença socioeconômica existente entre elas. A pesquisa foi aplicada em 132 residências e consistiu em duas atividades principais: a primeira com uma coleta e análise dos usos finais de água utilizando *smart meters* e um *software* de análise de fluxo; e a segunda com a elaboração, aplicação e análise estatística de um questionário. Como resultado, a análise estatística permitiu dividir a amostra em dois grupos: o grupo *Very High Level of Concern* – VHC, com nível alto de conscientização sobre a conservação de água; e o *Moderate to High Level of Concern* – MHC, com nível moderado de conscientização sobre a conservação de água. Os autores constataram que as famílias com nível alto de conscientização (VHC) tiveram um consumo de água de 128,2 litros/pessoa por dia contra 169,0

litros/pessoa por dia daquelas famílias com nível moderado de conscientização (MHC). Os valores representaram uma diferença de 24,1% no consumo total de água, indicando que as famílias com nível alto de conscientização sobre a conservação de água consomem menos água do que aquelas com nível moderado.

Outra pesquisa realizada na Austrália, mais precisamente no sudeste de *Queensland*, investigou a diferença entre o real consumo de água e o percebido pelos usuários. Inicialmente, o estudo identificou os usos finais de 252 residências por meio de medidores inteligentes de alta resolução. Além disso, os residentes responderam a um questionário com 27 perguntas, que buscaram identificar atitudes e comportamentos em relação à conservação da água no domicílio. Os entrevistados também responderam à questão chave do estudo: “você se considera um usuário com perfil de consumo de água baixo, médio ou alto?” (grifo nosso), com base no consumo médio diário presente na conta de água. Sendo assim, os usuários foram divididos em grupos que se autodenominaram com perfil de consumo baixo, médio ou alto. Uma série de análises de variância (ANOVA) foi aplicada para examinar um conjunto de variáveis que caracterizaram cada grupo (idade, renda, uso eficiente da água, tamanho e composição da família e atitudes de conservação de água). Com um intervalo de 95% de confiança, os resultados sugeriram que os usuários de água autodenominados com perfil de consumo alto usaram menos água (130 litros por pessoa/dia) do que os usuários autodenominados com perfil de consumo médio (156 litros por pessoa/dia) ou baixo (143 litros por pessoa/dia). Os usos finais associados ao aumento no uso da água foram predominantemente torneira, chuveiro e máquina de lavar roupas. Quanto às características sociodemográficas, havia uma tendência de subestimar a contribuição das crianças na demanda residencial de água pela falta de conhecimento de como e quando elas utilizavam a água. A renda familiar também foi associada com o consumo de água: usuários que se autodenominavam com consumo alto, porém que consumiam a menor quantidade de água, possuíam renda familiar significativamente menor do que os outros entrevistados. Isso comprova o fato de que a renda familiar está diretamente relacionada ao consumo de água, ou seja, famílias com renda mais alta tendem a consumir mais do que aquelas com menor renda (BEAL; STEWART; FIELDING, 2013).

Beal *et al.* (2018) buscaram compreender os padrões de consumo e identificar os parâmetros determinantes nos usos finais de água em três comunidades remotas da Austrália. O primeiro estágio da pesquisa envolveu 52 residências entre março de

2015 e junho de 2016. Foram instalados medidores inteligentes de alta resolução em cada residência para obter os usos finais de água. Outro componente fundamental do estudo foi a obtenção de dados qualitativos por meio da aplicação de entrevistas estruturadas com os participantes. O questionário consistiu em 43 questões de múltipla escolha e abrangeu temas como instalação de medição planejada, equipamentos hidrossanitários disponíveis, atitudes em relação à qualidade da água, autoidentificação como um cidadão preocupado com o abastecimento e alto consumo de água, entre outros. Como resultados, os autores verificaram que o uso externo de água passou de 1.500 para 5.300 litros por residência/dia, representando até 86% do uso total de água. Os principais impulsionadores do alto consumo externo estavam relacionados às atividades diárias e compreendiam: saúde (supressão de poeira, limpeza de superfícies externas); promoção de espaço verde com rega de plantas e jardins; preparação de alimentos (limpeza de peixe e carne); coleta de alimentos (limpeza de barcos e equipamentos de caça) e acesso às mangueiras durante o verão para brincadeiras das crianças. Um plano de gestão da demanda direcionada por esses dados será desenvolvido no segundo estágio do projeto. Segundo os autores, estratégias de gestão da demanda em torno do alto uso de água provavelmente envolverão o engajamento contínuo da comunidade, educação e consulta entre os residentes, autoridade local e outras partes interessadas. Outra abordagem a ser considerada na próxima fase do projeto será o papel do governo local em desenvolver práticas eficientes como uso de temporizadores nas torneiras e irrigação cautelosa de locais públicos.

Na Inglaterra, pesquisadores realizaram um estudo com o objetivo de identificar se havia relação entre o consumo de água e o bem-estar dos britânicos. Para isso, 187 pessoas foram selecionadas aleatoriamente nas ruas da região e responderam a um questionário com 34 questões relacionadas ao uso da água e ao bem-estar além de atitudes ambientais e atitudes de conservação e reúso da água. A população analisada era em média mais velha e com um nível socioeconômico ligeiramente mais alto do que a população do Reino Unido como um todo. O formulário contemplava questões sobre características socioeconômicas como a idade e ocupação do entrevistado, tipo de habitação (casa, apartamento ou bangalô), renda mensal e propriedade da habitação (alugada, hipotecada ou quitada). Também envolveu atitudes de conservação da água em perguntas sobre a utilização de acessórios economizadores de água, reúso da água do banho para bacia sanitária e jardim e

utilização da água da chuva para irrigação do jardim. Nesse sentido, no contexto do Reino Unido, os autores não encontraram correlação entre o uso de água e o bem-estar da população, sugerindo que níveis elevados de bem-estar são atingidos independentemente do baixo consumo de água, considerando o essencial para satisfazer as necessidades básicas (CHENOWETH *et al.*, 2016).

2.1.6 Modelagem espacial

Estudos recentes indicam que características geográficas e interações espaciais contribuem para a análise do consumo de água em edificações. Por isso, pesquisas nessa área estão sendo desenvolvidas com modelagem espacial pelo método de pesos (CHANG; PARANDVASH; SHANDAS, 2010; KONTOKOSTA; JAIN, 2015; CHANG *et al.*, 2017).

Ao analisar as variações espaciais no consumo de água e seus fatores determinantes, uma importante ferramenta de tomada de decisão é fornecida aos gestores públicos (SHI *et al.*, 2018). Sendo assim, é fundamental conhecer o processo dinâmico do uso da água no que tange à estrutura espacial urbana e examinar o efeito de vizinhança na escala de agregados espaciais (OUYANG *et al.*, 2014). A Tabela 1 apresenta alguns estudos que realizaram análises espaciais segundo setores censitários a fim de explorar os fatores influentes e padrões espaciais no uso residencial de água.

Tabela 1 – Estudos relativos à modelagem espacial e fatores influenciadores no consumo de água

Autor(es)	Região	Variáveis explicativas	Modelagem
GUHATHAKURTA; GOBER (2010); LEE; CHANG; GOBER (2015)	Phoenix	Efeito das ilhas de calor (+); temperatura (+); precipitação (-); presença de jardim e piscina (+)	Moran's I Global; regressão por mínimos quadrados ordinários (OLS); regressão espacial
CHANG; PARANDVASH; SHANDAS (2010); GOBER <i>et al.</i> (2012)	Portland	temperatura (+); precipitação (-); tamanho do edifício (+); idade do edifício (+).	Moran's I Global; OLS; regressão espacial.

Autor(es)	Região	Variáveis explicativas	Modelagem
CHANG <i>et al.</i> (2017)	Austin e Salt Lake	idade do edifício (-); valor venal (+); percentual de superfície impermeabilizada (-).	Moran's I Global; OLS; regressão espacial.
DAY; HOWE (2003)	Sydney	Jardim (+); uso do solo (+); comportamento no uso da água (+); demografia (+).	Moran's I Global; autocorrelação espacial
LI <i>et al.</i> (2017)	Lincoln, Grand Island e Sydney	Número de unidades habitacionais (+); tamanho médio da edificação (+); densidade populacional (+); áreas verdes (+).	OLS; Coeficiente de correlação de Pearson; Moran's I Global; regressão espacial.
JANMAAT (2013)	Kelowna	Lotes (+); tamanho do edifício (+); presença de piscina (+); uso de água nos banheiros (+); uso de água no paisagismo (+).	OLS; regressão espacialmente defasada

Fonte: Adaptado de Shi *et al.* (2018).

A modelagem espacial é considerada uma importante ferramenta na análise do consumo de água por determinar os fatores mais influentes no uso de água em cada região de análise. Dessa forma, podem ser formuladas políticas diferenciadas conforme a necessidade de cada região, por exemplo, melhoria do abastecimento e desenvolvimento de sistemas de economia de água (SHI *et al.*, 2018).

2.2 ANÁLISE ESTATÍSTICA

Nesta seção são apresentados os principais conceitos e estudos relativos às análises estatísticas da presente pesquisa, abrangendo os seguintes tópicos: análise exploratória dos dados, sazonalidade em séries temporais, diferença de médias de

duas populações, estudo de correlação entre variáveis, regressão linear múltipla, regressão robusta e regressão geograficamente ponderada.

2.2.1 Estatística descritiva dos dados

O processo de análise exploratória dos dados inicia com a identificação dos atributos do conjunto de dados (CUI *et al.*, 2019). É necessário também realizar uma análise univariada a fim de observar detalhes como: centralidade (média, mediana e moda) e dispersão (amplitude, variância, desvio padrão e assimetria). Outro passo importante é entender a relação entre variáveis por meio de análises estatísticas bivariadas e multivariadas, além de verificar interações entre os atributos (KRASKA, 2018).

Nas análises bivariadas, uma técnica útil para verificar a associação entre duas variáveis quantitativas é o gráfico de dispersão. Quando uma variável é quantitativa e a outra qualitativa, é comum analisar o que ocorre com a variável quantitativa dentro de cada categoria da variável qualitativa. Essa análise pode ser realizada com o auxílio de *boxplots* (MORETTIN; BUSSAB, 2013)

Além disso, dados atípicos ou *outliers* devem ser detectados uma vez que podem resultar em uma análise tendenciosa (enviesada) dos dados. Valores anormais ocorrem como resultado de erros de entrada ou de cálculo incorreto do usuário, enquanto *outliers* (dados espúrios) são observações que apresentam grande afastamento das demais ou que são inconsistentes (TUFFÉRY, 2011; GHOSH *et al.*, 2018).

Após obter uma visão detalhada do conjunto de dados, o processo de criação e transformação de variáveis pode ser necessário. Uma técnica bastante empregada nessa fase compreende a transformação de distribuições assimétricas em simétricas, transformando relações não-lineares complexas em relações lineares simplificadas (TUFFÉRY, 2011).

Dessa forma, a finalidade precípua da análise exploratória é descrever as principais características dos dados com o auxílio de tabelas, gráficos e medidas descritivas (BARBETTA; REIS; BORNIA, 2010). Para resumir e descrever características importantes dos dados coletados, são necessários métodos da estatística descritiva. Alguns deles são gráficos por natureza, por exemplo, construção de histogramas, *boxplots* e gráficos de dispersão. Outros envolvem o cálculo de

medidas numéricas: médias, desvios padrão e coeficientes de correlação (DEVORE, 2016).

Na análise exploratória também é importante avaliar o conjunto de observações ordenadas no tempo. Essa avaliação visa identificar possíveis padrões não aleatórios na série temporal da variável de interesse.

2.2.2 Sazonalidade determinística em séries temporais

Como é difícil modelar uma série temporal em sua totalidade, considerando suas diversas componentes, é útil dividi-la em categorias para melhor entendimento. Tradicionalmente, a série temporal é desmembrada em quatro componentes: tendência, ciclo, sazonalidade e flutuação irregular (NGUYEN; NOVÁK, 2019). A sazonalidade determinística é caracterizada por um padrão sazonal regular e estável no tempo, sendo assim é possível prever o comportamento sazonal perfeitamente a partir de dados anteriores (HYNDMAN; ATHANASOPOULOS, 2018).

Uma série apresenta comportamento periódico com período s quando similaridades na série ocorrem depois de s intervalos de tempo (BOX; JENKINS; REINSEL, 1994). Assim, um padrão sazonal ocorre quando uma série temporal é afetada por fatores sazonais como a época do ano ou dias da semana (HYNDMAN; ATHANASOPOULOS, 2018).

Os métodos de regressão são eficazes na análise de séries que podem apresentar sazonalidade determinística. O modelo de posto completo é dado pela Equação (1) (MORETTIN; TOLOI, 2006):

$$Z = \sum_{i=0}^m \beta_i t^i + \sum_{j=1}^{n_j} \alpha_j D_{jt} + \alpha_t \quad (1)$$

Em que:

Z_t = Variável resposta;

D_{jt} = Variáveis periódicas;

α_j e β_i = Coeficientes estimados da regressão.

Dessa forma, não existe sazonalidade determinística na série temporal se a hipótese nula de que todos os α_j são nulos não for rejeitada, ou seja, se $H_0: \alpha_1 = \dots =$

$\alpha_{n_j} = 0$ não for rejeitada (MORETTIN; TOLOI, 2006). Dado o modelo, a suposição de homocedasticidade da variância pode ser verificada pelos testes de Levene e Bartlett. O teste de Bartlett é o mais adequado nos casos em que a distribuição é normal. O teste de Levene é o melhor por ter baixa sensibilidade à não normalidade dos dados. Os testes avaliam a hipótese nula de que as variâncias são iguais $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$ contra a hipótese alternativa $H_1: \sigma_i^2 \neq \sigma_j^2$ para pelo menos um par (i, j) (TUFFÉRY, 2011).

Ademais, dois testes formais para verificar a existência de sazonalidade na série são: teste ANOVA rotineiro a uma análise de variância e o teste não-paramétrico de Kruskal-Wallis (MORETTIN; TOLOI, 2006; COELHO *et al.*, 2017). No teste não paramétrico de Kruskal-Wallis, assume-se que os dados são provenientes de k amostras independentes A_1, A_2, \dots, A_k ($k = 4$ se a série é trimestral; $k = 12$ se é mensal) de tamanho n_1, n_2, \dots, n_k respectivamente (LADIRAY; QUENNEVILLE, 2001). O teste é baseado na estatística da Equação (2):

$$W = \frac{12}{n(n+1)} \sum_{j=1}^k \frac{S_j^2}{n_j} - 3(n+1) \quad (2)$$

em que S_j representa a soma dos *ranks* das observações da amostra A_j dentre a amostra total de $n = \sum_{j=1}^k n_j$ observações. Essa estatística segue a distribuição qui-quadrado com $k - 1$ graus de liberdade.

2.2.3 Diferença de médias de duas populações independentes

Frequentemente, pesquisadores estão interessados em comparar duas condições diferentes para determinar se cada condição produz efeitos estatísticos distintos na resposta observada (MONTGOMERY; RUNGER, 2016).

A análise da diferença de médias de duas populações independentes pode ser realizada com o auxílio de testes de hipótese ou testes de significância. Sua finalidade é verificar se os dados fornecem evidência suficiente para a aceitação da hipótese em estudo (BARBETTA; REIS; BORNIA, 2010).

Para a situação em que as amostras possuem distribuição normal e as variâncias sejam desconhecidas, a estatística de teste é baseada na distribuição t de

Student. Por ser usada a estimativa combinada de variância, o procedimento é frequentemente chamado de teste t combinado (MONTGOMERY; RUNGER, 2016).

Quando o tamanho das amostras é grande, ou seja, excede 40, a hipótese de distribuição normal das populações é desnecessária. Nesse caso, o Teorema do Limite Central garante que a diferença de médias possui aproximadamente distribuição normal, independentemente das distribuições das populações subjacentes (DEVORE, 2016).

Nas situações em que as populações não sejam normais, porém possuam distribuições contínuas, com a mesma forma e dispersão, pode ser usado o teste de Wilcoxon Mann-Whitney (MONTGOMERY; RUNGER, 2016).

2.2.4 Verificação da hipótese de normalidade

Certamente, a distribuição normal é o modelo mais amplamente utilizado para descrever uma variável aleatória contínua (MONTGOMERY; RUNGER, 2016). Sua função de probabilidade é representada graficamente por uma curva em forma de sino e relaciona dois parâmetros: média, ou valor esperado, e desvio padrão. O modelo normal é caracterizado pela maior probabilidade de a variável aleatória assumir valores próximos ao centro, o que se torna evidente pela forma da distribuição (BARBETTA; REIS; BORNIA, 2010).

Devido as suposições de normalidade adotadas nos modelos de regressão linear, a aderência do conjunto de dados à distribuição normal deve ser avaliada. Para tanto, podem ser utilizados métodos gráficos e testes de aderência (TUFFÉRY, 2011).

A forma mais usual de apresentar a distribuição de frequências de variáveis aleatórias contínuas é por meio do histograma. Trata-se de um gráfico de barras contíguas, com as bases proporcionais aos intervalos de classes e a área de cada retângulo proporcional à respectiva frequência. Caso sua representação se assemelhe a uma curva em forma de sino, com concentração de dados no centro, sugere-se que a distribuição normal seja um bom modelo para representar os dados (BARBETTA; REIS; BORNIA, 2010; MORETTIN; BUSSAB, 2013).

O histograma resume graficamente a distribuição do conjunto de dados, apresentando características como posição central, dispersão, assimetria, curtose e presença de *outliers* (dados espúrios). No entanto, a maneira como o histograma retrata a distribuição dos dados é de alguma forma arbitrária, dependendo sobretudo

da escolha dos intervalos de classe (ÖZTUNA; ELHAN; TÜCCAR, 2006). Por isso, os testes de normalidade são uma opção segura de avaliar a aderência dos dados à distribuição normal, já que se baseiam em estatísticas teste e não meramente em análise gráfica.

De acordo com Tufféry (2011), a suposição de normalidade de uma variável pode ser verificada pelos testes de Shapiro-Wilk (o melhor), de Kolmogorov-Smirnov (o mais geral), de Lilliefors ou teste de Anderson-Darling.

Caso a distribuição não seja normal, podem ser aplicadas transformações (Box-cox, exponencial e logarítmica) às variáveis antes de utilizá-las em análises seguintes (TUFFÉRY, 2011). Segundo Barbetta, Reis e Bornia (2010), a transformações logarítmica torna distribuições com assimetria positiva (cauda mais longa à direita) em distribuições aproximadamente simétricas. Isto é, aumenta as distâncias entre os valores pequenos e reduz as distâncias entre os valores grandes. Com isso, tem-se uma situação mais adequada para as análises de correlação e regressão.

2.2.5 Análise de correlação

A análise de correlação visa determinar (mensurar) o grau de relacionamento entre duas variáveis e é usualmente realizada em termos exploratórios na pesquisa (BARBETTA; REIS; BORNIA, 2010). Uma ferramenta gráfica bastante útil para verificar a associação entre duas variáveis quantitativas é o diagrama de dispersão, em que os dados são representados por pontos em um sistema cartesiano (MORETTIN; BUSSAB, 2013).

Contudo, é importante quantificar a força desta associação mediante o cálculo do coeficiente de correlação. No caso da correlação linear, o valor do coeficiente varia de -1 a 1. Valores próximos de um indicam uma forte correlação linear entre X e Y, ao passo que valores próximos a zero indicam pouca ou nenhuma associação. Nessa situação, pode existir uma relação não linear entre as variáveis, a qual não é captada por esse método (WALPOLE *et al.*, 2009).

Para descrever a interdependência entre variáveis quantitativas, o coeficiente de correlação de Pearson é um bom método. O coeficiente de correlação ponto bisserial é indicado quando uma das variáveis é dicotômica (dummy) e a outra, contínua. No caso de variáveis categóricas mensuradas em nível ordinal, o método

de cálculo mais conhecido é o coeficiente de correlação de Spearman (LEWIS-BECK; BRYMAN; LIAO, 2004; LIRA; CHAVES NETO, 2006; SALKIND, 2007).

Em outras palavras, o coeficiente de correlação de Pearson foi desenvolvido para trabalhar com variáveis medidas em uma escala contínua. Caso a variável seja mensurada em nível ordinal, então recomenda-se a utilização do coeficiente de Spearman (PUTH; NEUHÄUSER; RUXTON, 2014).

Além do coeficiente de Spearman, outra medida de correlação não-paramétrica é o coeficiente de Kendall. A correlação de Kendall é mais robusta e um pouco mais eficiente do que a correlação de Spearman, tornando seu uso mais apropriado na presença de *outliers* (CROUX; DEHON, 2010).

Uma população que tenha duas variáveis não correlacionadas pode produzir uma amostra com coeficiente de correlação diferente de zero. Por esse motivo, é necessário também testar a significância dos coeficientes de correlação da amostra a fim de confirmar se as variáveis X e Y realmente são correlacionadas (MORETTIN; BUSSAB, 2013).

Contudo, a correlação refere-se a uma associação numérica entre duas variáveis, não implicando, necessariamente, relação de causa e efeito. A análise de correlação é feita em termos exploratórios e pode servir como elemento auxiliar na escolha de variáveis para a elaboração de um modelo de regressão linear múltipla (BARBETTA; REIS; BORNIA, 2010).

2.2.6 Regressão linear múltipla

Situações envolvendo variáveis que estão relacionadas de maneira não determinística ocorrem frequentemente em diversos ramos da engenharia e da ciência. Sendo assim, a análise de regressão é a parte da estatística que investiga a relação entre duas ou mais variáveis que obedecem a essa condição (MONTGOMERY; RUNGER, 2010; DEVORE, 2016).

De acordo com Montgomery e Runger (2016), no método de regressão linear múltipla, a variável dependente ou de resposta (Y) pode estar relacionada com k variáveis independentes, conforme a Equação (3).

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon \quad (3)$$

Em que:

Y = Variável dependente ou resposta;

x_j = Variáveis independentes (preditoras ou regressoras);

β_j = Coeficientes de regressão;

ϵ = Termo de erro aleatório.

O modelo de regressão (3) descreve um hiperplano no espaço k-dimensional das variáveis regressoras $\{x_j\}$. O parâmetro β_j , $j = 0, 1, \dots, k$, representa a variação esperada na resposta Y por unidade de variação unitária em x_j , quando todos os outros regressores forem mantidos constantes. Ademais, tal modelo é frequentemente usado como uma função de aproximação. Isto é, a verdadeira relação funcional entre Y e x_1, x_2, \dots, x_k é desconhecida; no entanto, sob faixas das variáveis independentes, o modelo de regressão linear é uma aproximação adequada (MONTGOMERY; RUNGER, 2016).

Para a construção do modelo descrito na Equação (3), precisa-se obter estimativas para os coeficientes de regressão a partir de um conjunto de observações. Ou seja, é necessário encontrar a reta que passe o mais próximo possível dos pontos observados. O método mais usual utiliza o princípio dos mínimos quadrados, que consiste em fazer com que a soma dos erros quadráticos seja a menor possível (BARBETTA; REIS; BORNIA, 2010).

Os estimadores de mínimos quadrados estão sujeitos a certas suposições sobre os termos do erro aleatório no modelo de regressão. Considera-se que os erros são estatisticamente independentes, com média zero e variância constante. Sob essas suposições, os estimadores de mínimos quadrados são estimadores não tendenciosos dos coeficientes de regressão (MONTGOMERY; RUNGER, 2016). Todavia, a reta dos mínimos quadrados não deve ser usada em previsões cujos valores das variáveis regressoras estejam muito além do intervalo dos dados (DEVORE, 2016).

Para Montgomery e Runger (2016), nos problemas de regressão linear múltipla, testes de hipótese relativos aos parâmetros do modelo são úteis na medida da sua adequação. Segundo os autores, a determinação do valor potencial de cada um dos regressores auxilia na decisão de inclusão ou exclusão de variáveis no modelo. A hipótese para testar se um coeficiente individual de regressão é significativo

considera: $H_0: \beta_j = 0$ e $H_1: \beta_j \neq 0$. A não rejeição de H_0 indica que o regressor x_j pode ser retirado do modelo. A hipótese nula $H_0: \beta_j = 0$ será rejeitada se $|t_0| > t_{\alpha/2, n-p}$. O teste é chamado de parcial ou marginal porque o coeficiente de regressão $\hat{\beta}_j$ depende de todos os outros regressores que estão no modelo. A Equação (4) indica a estatística do teste para essa hipótese.

$$T_0 = \frac{\hat{\beta}_j}{\sqrt{\sigma^2 C_{jj}}} \quad (4)$$

Em que:

T_0 = Estatística do teste parcial ou marginal;

$\hat{\beta}_j$ = Coeficientes de regressão estimados;

C_{jj} = Elemento da diagonal $(X'X)^{-1}$ correspondente a $\hat{\beta}_j$;

$\sqrt{\sigma^2 C_{jj}}$ = Erro padrão do coeficiente $\hat{\beta}_j$.

Um método de análise de variância também poderá ser aplicado para testar a significância da regressão. O teste determina se existe relação linear entre a variável resposta Y e um subconjunto de regressores x_1, x_2, \dots, x_k . As hipóteses apropriadas são: $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ e $H_1: \beta_j \neq 0$ para no mínimo um j (MONTGOMERY; RUNGER, 2016).

Esse teste é baseado na razão F , resultante da divisão do quadrado médio da regressão pelo quadrado médio do erro. O quadrado médio (ou variância) é calculado pela soma dos quadrados dividido pelo correspondente grau de liberdade (BARBETTA; REIS; BORNIA, 2010). Montgomery e Runger (2016) apresentam a estatística de teste conforme a Equação (5).

$$F_0 = \frac{SQ_R/k}{SQ_E/(n-p)} = \frac{MQ_R}{MQ_E} \quad (5)$$

Em que:

F_0 = Estatística do teste para a significância da regressão;

SQ_R = Soma de quadrados da regressão;

SQ_E = Soma de quadrados do erro ou resíduo;

$k, n - p$ = Graus de liberdade;

MQ_R = Média quadrática da regressão;

MQ_E = Média quadrática do erro ou resíduo.

Caso o valor calculado na Equação (5), f_0 , seja maior que $f_{\alpha,k,n-p}$, rejeita-se H_0 . A rejeição de H_0 implica que no mínimo uma das variáveis preditoras contribui significativamente para o modelo. O procedimento de análise de variância para testar a significância da regressão é sumarizado na Tabela 2 (MONTGOMERY; RUNGER, 2016).

Tabela 2 – Análise de variância (ANOVA)

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Média Quadrática	F_0
Regressão	SQ_R	k	MQ_R	MQ_R/MQ_E
Erro ou resíduo	SQ_E	$n - p$	MQ_E	
Total	SQ_T	$n - 1$		

Fonte: Montgomery e Runger (2016).

No caso dos testes de hipótese, as análises feitas por *softwares* estatísticos retornam a medida p-valor, o menor nível de significância que leva à rejeição da hipótese nula considerando os dados da amostra. Sendo assim, basta comparar o p-valor com o nível de significância adotado na pesquisa para decidir entre rejeitar ou aceitar uma hipótese (BARBETTA; REIS; BORNIA, 2010).

Outra estatística global para avaliar o ajuste do modelo de regressão múltipla é o coeficiente de determinação múltipla R^2 . Esse coeficiente é interpretado como a proporção da variação de Y que pode ser explicada pelo modelo de regressão (DEVORE, 2016). Todavia, pode levar a interpretações errôneas. Exemplificando, um alto coeficiente de determinação não indica necessariamente que o modelo de regressão teve um bom ajuste. Por outro lado, um modelo com coeficiente de determinação baixo pode ter variáveis independente e regressoras relacionadas (KUTNER; NACHTSHEIM; NETER, 2004). Também, a estatística R^2 é problemática como uma medida de qualidade do ajuste para o modelo, uma vez que ela sempre aumenta quando uma variável é adicionada ao modelo (MONTGOMERY; RUNGER, 2016).

Por consequência, a fim de analisar a qualidade do ajuste do modelo e comparar modelos competitivos de regressão, recomenda-se utilizar a estatística

$R^2_{ajustado}$. Essa estatística só aumenta se a variável adicionada ao modelo reduzir a média quadrática do erro. Ou seja, é possível avaliar a inclusão de regressores que não são realmente úteis ao modelo (MONTGOMERY; RUNGER, 2016).

Convém ainda verificar se os dados exibem multicolinearidade, isto é, se os preditores x_1, x_2, \dots, x_k estão altamente interdependentes na regressão. Isso afeta principalmente a estabilidade dos coeficientes. A multicolinearidade é sugerida por uma regressão em que R^2 é grande e cujo sinal dos coeficientes β_j é oposto ao que intuitivamente se espera. Se o teste F para significância da regressão for significativo, porém os testes para os coeficientes individuais de regressão não forem, então a multicolinearidade pode estar presente (DEVORE, 2016; MONTGOMERY; RUNGER, 2016).

Para diagnosticar a presença de multicolinearidade na regressão, utiliza-se comumente o método do *Variance Inflation Factor* – VIF. Esses fatores avaliam o quanto as variações dos coeficientes de regressão estimados são inflacionadas quando comparadas às variáveis preditoras não relacionadas linearmente (KUTNER; NACHTSHEIM; NETER, 2004).

Conforme Kutner, Nachtsheim e Neter (2004), a multicolinearidade pode estar influenciando indevidamente as estimativas dos mínimos quadrados quando os valores de VIF são superiores a 10. Em seu estudo envolvendo pegada hídrica, Fitzgerald e Auerbach (2016) analisaram o potencial de colinearidade dos termos regressores cujo $VIF > 8,0$. Em outra pesquisa similar, Cruz *et al.* (2017) adotaram o limite de 5 para o VIF como forma de explicar associações moderadas entre diversos pares de variáveis independentes. Connolly e Hagelman III (2015) adotaram $VIF < 7,5$ para indicar ausência de multicolinearidade do modelo. Na opinião de Gujarati e Porter (2011), a variável será altamente colinear com $VIF > 10$.

Outra etapa imprescindível na regressão é a análise dos resíduos. O resíduo $e_i = Y_i - \hat{Y}_i$ é definido como a diferença entre o valor observado Y_i e o valor ajustado pelo modelo \hat{Y}_i (KUTNER; NACHTSHEIM; NETER, 2004). Pela análise dos resíduos, podem ser detectadas: a presença de valores discrepantes; heterocedasticidade (variância não constante dos erros) e especificações errôneas do modelo (WALPOLE *et al.*, 2009). Valores discrepantes podem representar *outliers* e/ou observações influentes.

Um *outlier* é uma observação, escalar ou vetorial, que difere marcadamente dos outros dados da amostra. Coloquialmente, são pontos fora da curva, pois não partilham da mesma distribuição de probabilidades que a maioria das outras observações (BARROS, 2013). Um ponto influente, por sua vez, além de ser um *outlier*, pode exercer muita influência nas estimativas dos coeficientes de regressão, na magnitude da média quadrática dos erros e na determinação de R^2 (MONTGOMERY; RUNGER, 2016).

Para a validade dos intervalos de confiança e testes estatísticos, assume-se que os resíduos do modelo de regressão são independentes e normalmente distribuídos, com média zero e variância constante (BARBETTA; REIS; BORNIA, 2010). Walpole *et al.* (2009) e Devore (2016) afirmam que essas suposições podem ser avaliadas por meio de gráficos de diagnóstico e testes estatísticos. Os autores ressaltam ainda que diversos gráficos são frequentemente úteis para verificar a validade e utilidade do modelo, sendo preferível representar os resíduos '*studentizados*', por fornecerem uma padronização no seu comportamento.

Dessa forma, a suposição de normalidade pode ser verificada pelo gráfico de probabilidade normal e teste de Lilliefors. Quanto à homocedasticidade (homogeneidade da variância), o gráfico de resíduos padronizados *versus* valores ajustados é empregado em conjunto com o teste de Breush-Pagan. A autocorrelação e independência dos resíduos pode ser avaliada com auxílio do teste de Durbin-Watson. Por fim, a presença de possíveis padrões e *outliers* é visualmente investigada pelo gráfico de resíduos *versus* valores ajustados, bem como pela aplicação do teste de Bonferroni. Essas medidas de diagnóstico são sugeridas por Kutner, Nachsheim e Neter (2004).

Segundo esses autores, um *outlier* poderá ser influente se a sua exclusão causar alterações significativas na função de regressão ajustada. Os autores sugerem como medida amplamente utilizada na prática para mensurar tal influência a análise da Distância de Cook.

A fim de detectar possíveis observações influentes no modelo de regressão, pode ser utilizada a medida da distância de Cook, método desenvolvido por Dennis R. Cook. Trata-se de uma medida da distância ao quadrado entre a estimativa usual de mínimos quadrados de β , baseada em todas as n observações, e a estimativa obtida quando o i -ésimo ponto for removido (MONTGOMERY; RUNGER, 2016). De acordo

com Stevens (1984), distâncias de Cook > 1 indicam que os *outliers* serão influentes no conjunto de dados.

Nessas situações, ignorar ou deletar observações influentes pode levar a grandes desvios nas estimativas, por isso é um procedimento que deve ser realizado raramente e com precaução (ZHANG *et al.*, 2018). Ogu, Inyama e Achugamonu (2013) recomendam então o emprego de métodos de regressão robusta visto que são minimamente afetados por *outliers*.

2.2.7 Regressão robusta

Na presença de *outliers* ou quando os resíduos das observações não seguem a distribuição normal, as estimativas dos coeficientes de regressão pelo método dos mínimos quadrados podem não ser mais confiáveis. Nesses casos, a regressão robusta é amplamente utilizada por fornecer previsões menos sensíveis a *outliers* (TUFFÉRY, 2011).

Os métodos de regressão robusta podem melhorar consideravelmente a precisão das estimativas dos coeficientes, porém não devem ser implementados automaticamente antes dos métodos de regressão clássicos (ALMETWALLY; ALMONGY, 2018).

A utilização de métodos robustos de análise na presença de valores extremos é um aspecto importante na produção de resultados acurados e precisos. Isto porque a remoção de *outliers* com uma simples análise exploratória dos dados e consequente aplicação de procedimentos inferenciais clássicos aos dados restantes pode ser uma atitude falha em muitos aspectos. Primeiro, os *outliers* podem ser mascarados e muito difíceis de detectar. Segundo, é mais eficiente reduzir o peso em vez de descartar tais observações (FARCOMENI; VENTURA, 2010).

Dentre as diversas técnicas de estimação de parâmetros em uma regressão linear robusta, está a classe M-Estimadores (*M-Estimators*), introduzida por Huber em 1964. Seu objetivo é manter o valor das observações próximas ao centro da distribuição e minimizar o peso daquelas que se distanciem do centro (RIANI *et al.*, 2014). A classe de M-Estimadores generaliza o método da máxima verossimilhança e foi estendida para todas as distribuições de probabilidade, produzindo estimadores consistentes e assintoticamente normais (HERITIER, 2009). A classe MM estimadores combina os benefícios dos estimadores S e M ao manter a robustez e resistência do

estimador S e incorporar a eficiência do estimador M (SUSANTI *et al.*, 2014; BRENNER *et al.*, 2019).

A pesquisa de Almetwally e Almongy (2018) comparou seis métodos de estimativa em regressão robusta: M. Humpel, M. Bisquare, M. Huber, S, MM(S) e MM estimador. Os resultados indicaram que no caso de amostras com mais de 100 observações, o melhor método de estimativa é o MM. Após a estimativa dos coeficientes, é realizada a análise dos resíduos para verificar a adequação do modelo.

2.2.8 Regressão *stepwise*

Nas situações que envolvem muitas variáveis regressoras (30 ou mais), a seleção de variáveis é importante como forma de obter um modelo que desempenhe satisfatoriamente o uso pretendido (MONTGOMERY; RUNGER, 2016). Kutner, Nachtsheim e Neter (2004) recomendam um procedimento de busca automática que identifica o melhor subconjunto de preditores, auxiliando assim na escolha da melhor equação de regressão.

A regressão *stepwise* cria uma sequência de modelos de regressão, adicionando ou removendo uma variável x em cada etapa. O critério para adição ou remoção de uma variável pode ser definido em termos de redução da soma quadrática dos erros, coeficiente de correlação parcial, estatística t e estatística F (KUTNER; NACHTSHEIM; NETER, 2004).

O método de eliminação retroativa (ER), é o procedimento mais simples. Esse método começa com o modelo no qual são usados todos os regressores em análise. A eliminação progressiva de variáveis é definida pelas razões t . Se a razão t com menor valor absoluto for menor que uma constante t_{out} pré-especificada, então o preditor correspondente a essa razão menor é eliminado do modelo (DEVORE, 2016).

A seleção progressiva (SP), uma alternativa ao procedimento ER, ajusta o modelo sucessivamente considerando apenas x_1 , apenas x_2 e assim sucessivamente. A variável introduzida no modelo é aquela que gera a maior razão t , desde que essa razão ultrapasse a constante especificada t_{in} (DEVORE, 2016).

Conforme Devore (2016), existe ainda o procedimento retroativo que combina os métodos SP e ER, chamado de PR. Tal procedimento é iniciado pela seleção progressiva através da adição de variáveis ao modelo. Entretanto, após adicionar

cada variável ele analisa as variáveis anteriormente introduzidas, verificando se alguma delas precisa ser eliminada.

2.2.9 Regressão geograficamente ponderada

A técnica de regressão geograficamente ponderada, em inglês *Geographically Weighted Regression* – GWR, foi proposta por Brunsdon, Fotheringham e Charlton (1996) e possui como finalidade investigar relações espacialmente heterogêneas nos dados (LU *et al.*, 2014). Sua ideia básica é ajustar um modelo de regressão para cada ponto no conjunto de dados ponderado pelas observações mais próximas geograficamente (MEDINA, 2016).

O modelo GWR tem como base a regressão linear tradicional e também permite que os coeficientes sejam funções da localização espacial. Dessa forma, a técnica possibilita que o preditor varie para cada observação (i) no conjunto de dados com coordenadas geográficas (u_i, v_i) , conforme a Equação (6) (FOTHERINGHAM; BRUNSDON; CHARLTON, 2002).

$$y_i = \alpha(u_i, v_i) + \sum_{k=1}^m \beta_k(u_i, v_i) x_k(u_i, v_i) + \varepsilon_i \quad (6)$$

Em que:

y_i = Variável dependente na localização i ;

x_k = Variável independente k na localização i ;

m = Número de variáveis independentes;

α = Intercepto na localização i ;

β_k = Coeficiente local de regressão para a variável independente k na localização i ;

ε_i = Erro aleatório na localização i .

Segundo Fotheringham, Charlton e Brunsdon (1998), a regressão geograficamente ponderada possibilita que os coeficientes variem continuamente sobre a área de estudo. O método utiliza uma função de ponderação baseada no conceito de “colisão de influência”, de modo que, em cada ponto de regressão, observações mais próximas têm mais influência na estimativa do conjunto local de coeficientes do que observações mais distantes.

2.2.10 Validação do modelo

A validação é uma etapa útil e necessária no processo de construção de um modelo. A validade de um modelo refere-se à estabilidade e aceitabilidade dos coeficientes de regressão, à usabilidade da função de regressão e à habilidade de generalizar inferências extraídas da análise de regressão (KUTNER; NACHTSHEIM; NETER, 2004). A validade das premissas assumidas no desenvolvimento do modelo e estimação dos parâmetros, bem como das inferências futuras, tornam-se mais confiáveis se o modelo fornece previsões precisas com os novos dados (SNEE, 1977).

2.3 SÍNTESE DA REVISÃO E SUA INFLUÊNCIA SOBRE O MÉTODO ADOTADO

O consumo residencial de água foi apresentado como motivação para a pesquisa por constituir uma parcela considerável na demanda total de água no Brasil (ANA, 2018). Por isso, um entendimento claro dos fatores determinantes no consumo doméstico pode auxiliar no gerenciamento da água de forma sustentável (HUSSIEN; MEMON; SAVIC, 2016), além de contribuir no desenvolvimento de políticas voltadas à conservação desse recurso vital (MAKKI *et al.*, 2015).

A revisão abordou estudos mostrando que o consumo residencial de água é influenciado principalmente por aspectos socioeconômicos dos moradores como renda familiar, número total de moradores e nível de escolaridade do chefe da família (FAN *et al.*, 2017; DIAS; KALBUSCH; HENNING, 2018; SANT'ANA; MAZZEGA, 2018) e características construtivas das edificações, relacionadas à área construída, quantidade de banheiros e idade do edifício (MATOS *et al.*, 2013; MATOS *et al.*, 2014; CRUZ *et al.*, 2017; RAJEEVAN; MISHRA, 2020). Outros fatores apontados foram o uso de equipamentos economizadores de água (MARINOSKI; RUPP; GHISI, 2018) e fontes alternativas de abastecimento como reúso de água cinza e aproveitamento de água pluvial (GARCIA-CUERVA; BERGLUND; BINDER, 2016; LEONG *et al.*, 2018).

Dessa forma, a revisão apresentada foi fundamental para a definição de quais informações seriam abordadas no questionário, além da escolha das variáveis candidatas a preditores nos modelos de regressão. Optou-se também por incluir os hábitos de consumo dos usuários no presente estudo, pois Manouseli, Anderson e Nagarajan (2018) salientaram que a inclusão desses itens nos modelos de regressão é um tópico pouco explorado na literatura.

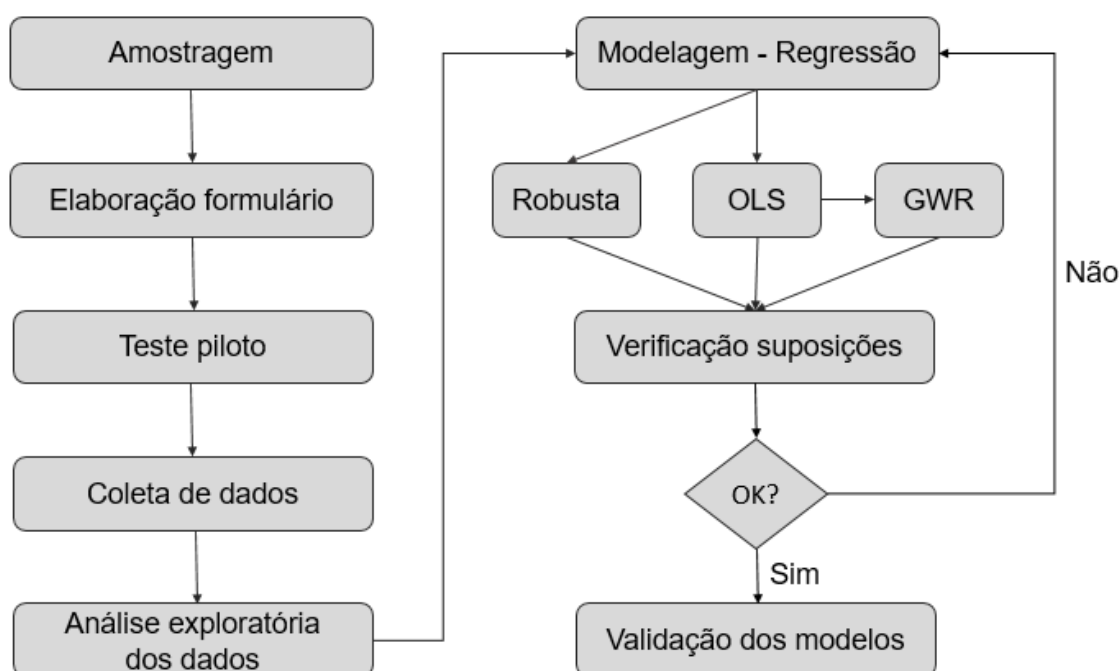
A revisão apresentou estudos realizados em algumas partes do mundo com propostas similares à desenvolvida neste trabalho de pesquisa. Percebeu-se que havia poucas pesquisas voltadas exclusivamente para residências unifamiliares, a maioria com esse enfoque era proveniente da Austrália (WILLIS *et al.*, 2011; Willis *et al.*, 2013; BEAL; STEWART; FIELDING, 2013). Consequentemente, o presente estudo foi aplicado somente em residências unifamiliares.

Por fim, na revisão foram abordados conceitos e definições das técnicas estatísticas que produziram os resultados da pesquisa: análise exploratória dos dados, sazonalidade, análise de correlação e modelos de regressão linear múltipla, robusta e geograficamente ponderada (KUTNER; NACHTSHEIM; NETER, 2004; BARBETTA; REIS; BORNIA, 2010; MONTGOMERY; RUNGER, 2010; TUFFÉRY, 2011; MORETTIN; BUSSAB, 2013; DEVORE, 2016). Essas técnicas foram escolhidas com base em algumas pesquisas apresentadas na revisão, principalmente as de Kontokosta e Jain (2015), Cruz *et al.* (2017) e Dias, Kalbusch e Henning (2018).

3 MÉTODO

Este capítulo apresenta os procedimentos metodológicos adotados na pesquisa, com o propósito de avaliar os fatores que influenciam o consumo de água em residências unifamiliares na cidade de Joinville/SC. A metodologia utilizada foi dividida em etapas, as quais estão sumarizadas no fluxograma da Figura 3.

Figura 3 – Fluxograma dos procedimentos metodológicos



No desenvolvimento deste estudo, primeiramente foi empregada uma pesquisa de caráter exploratório partindo de um levantamento bibliográfico. A revisão bibliográfica, fundamental na obtenção de uma base conceitual do assunto, contemplou temas como consumo residencial de água e seus fatores influentes, inferência estatística, modelos de regressão linear múltipla e regressão robusta e regressão geograficamente ponderada.

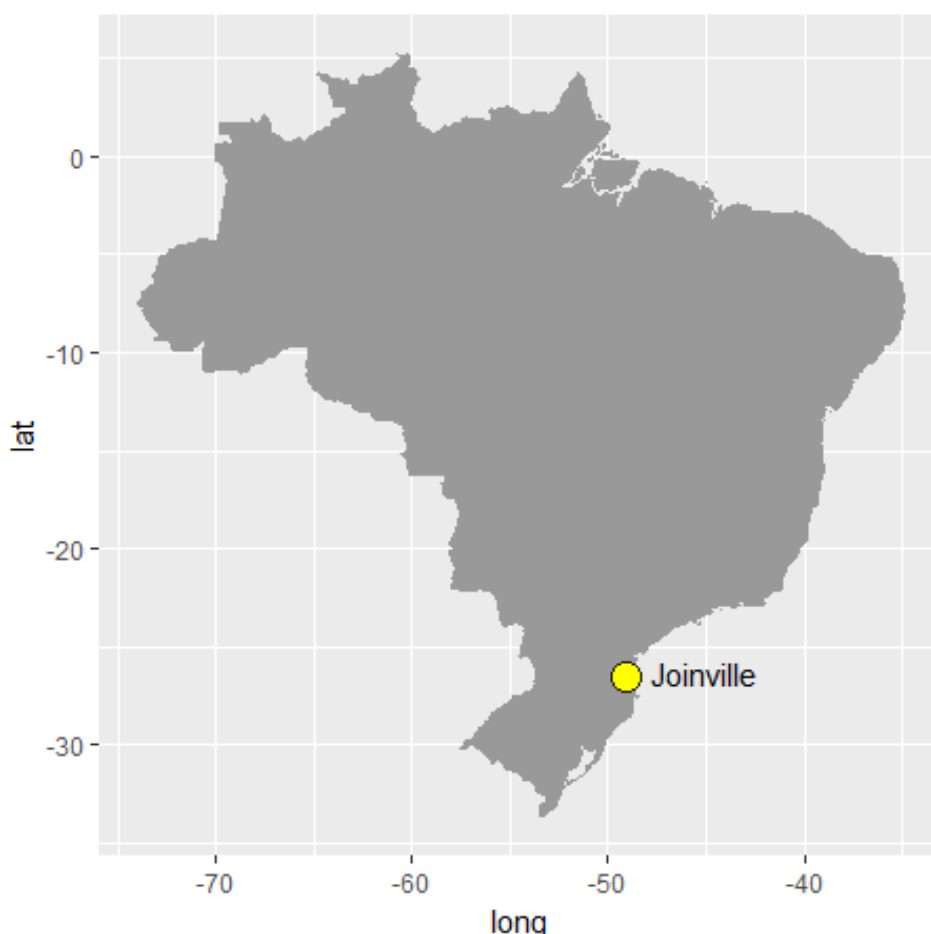
3.1 DESCRIÇÃO DO OBJETO DE ESTUDO

Outra etapa necessária ao estudo é a obtenção de informações sobre o município de Joinville, a fim de identificar características inerentes como localização,

número e gênero dos habitantes, previsão de crescimento populacional e serviço de abastecimento de água.

Joinville é um município do estado de Santa Catarina, localizado na região Sul do Brasil, como ilustra a Figura 4. Conforme o Instituto Brasileiro de Geografia e Estatística – IBGE (2019), possui uma área territorial de 1.127,946 km², altitude de 4,5m e está situado na latitude 26° 18' 05" sul e longitude 48° 50' 38" oeste, a 180 km da capital do estado, Florianópolis. É, também, um importante polo industrial da região Sul, com volume de receitas geradas inferior apenas ao das capitais Porto Alegre e Curitiba. A cidade concentra na indústria grande parte da atividade econômica, com destaque para os setores metalmeccânico, têxtil, plástico, metalúrgico, químico e farmacêutico (SEPUD, 2018).

Figura 4 – Localização geográfica do objeto de estudo



A população de Joinville em 2019 foi estimada pelo IBGE em 590.466 habitantes, sendo 49,77% homens e 50,23% mulheres. Desse total, 96,6% estão

distribuídos pela área urbana do município e somente 3,4% ocupam a área rural (SEPUD, 2018; IBGE, 2019). De acordo com o último censo demográfico, os idosos (≥ 65 anos) correspondiam a 8,79% enquanto as crianças e adolescentes entre 0 e 14 anos representavam 21,83% da população de Joinville no ano de 2010 (IBGE, 2010).

Na cidade de Joinville, os domicílios particulares permanentes do tipo casa representam um total de 134.199 unidades ao passo que residências de vila ou em condomínio somam 672 unidades. Em conjunto, essas habitações representam 83,95% do total de domicílios do município (IBGE, 2010). As habitações próprias quitadas ou financiadas representavam 89,15% das unidades domiciliares urbanas em Joinville enquanto 10,68% estavam alugadas ou cedidas e a situação “outros” abrangeu 0,16% dos domicílios (IBGE, 2010).

Sobre os aspectos socioeconômicos, a renda média mensal dos trabalhadores formais joinvilenses é de 2,9 salários mínimos (IBGE, 2017). Quanto à infraestrutura básica, o serviço de abastecimento de água abrange 98,2% da população enquanto apenas 31,8% dos habitantes são atendidos pela coleta de esgoto sanitário. O consumo *per capita* doméstico de água potável em Joinville foi 133,39 litros por pessoa por dia em 2018 (SNIS, 2018). Esse valor foi calculado com base nos dados apresentados no Sistema Nacional de Informações sobre Saneamento – SNIS (2018) considerando as variáveis volume micromedido nas economias residenciais ativas de água (em m^3) e população total atendida com abastecimento de água em 2018 (habitantes), conforme a Equação (7):

$$C_{per\ capita} = \frac{V_{micro} \times 1000}{P_{atendida} / 365} \quad (7)$$

Sendo:

$C_{per\ capita}$ = Consumo *per capita* doméstico de água de Joinville em 2018 (litros por pessoa por dia);

V_{micro} = Volume micromedido nas economias residenciais ativas de água (m^3 /ano);

$P_{atendida}$ = População total atendida com abastecimento de água (habitantes).

3.2 TAMANHO DA AMOSTRA

Uma fase preliminar da pesquisa de campo é a definição do tamanho da amostra, isto é, da quantidade de residências necessárias para coleta de informações. De modo geral, os estudos citados na revisão abrangem amostras relativamente grandes, superiores a 100 edificações. Willis *et al.* (2011) e Willis *et al.* (2013) analisaram, respectivamente, 132 e 151 residências de *Gold Coast*, na Austrália. O critério de seleção das amostras foi a diferença na composição socioeconômica e demográfica de cada região da cidade. Hussien, Memon e Savic (2016) distribuíram um questionário para 419 domicílios selecionados aleatoriamente na cidade de Duhok, no Curdistão iraquiano. Desse total, 407 famílias responderam o questionário, sendo 92% residências e 8% apartamentos. Beal, Stewart e Fielding (2013) selecionaram 252 casas de um grupo de 1750 participantes, ligados ao centro de pesquisa sobre uso residencial de água no sudeste de *Queensland*, na Austrália.

Neste trabalho, o tamanho da amostra n é calculado em função do erro amostral máximo tolerado (E_0) e do nível de confiança adotado no processo de estimação (γ), de acordo com a Equação (8) (BARBETTA; REIS; BORNIA, 2010).

$$n_0 = \frac{z\gamma^2}{4E_0^2} \quad (8)$$

Sendo:

n_0 = Tamanho da amostra;

$z\gamma^2$ = Valor crítico para o nível de confiança desejado;

E_0^2 = Erro amostral máximo tolerado.

No estudo, foi adotado o grau de confiança de 95% com um erro de amostragem α (nível de significância) de 5%, valor tradicionalmente utilizado pelos pesquisadores. Dessa forma, $z\gamma$ torna-se igual a 1,96 e a Equação (8) pode ser expressa, aproximadamente, pela Equação (9).

$$n_0 = \frac{1}{E_0^2} \quad (9)$$

Sendo:

n_0 = Tamanho da amostra;

E_0^2 = Erro amostral máximo tolerado.

Em seu estudo sobre avaliação da eficiência energética do envoltório de edificações, Carlo (2008) apresenta o número de amostras necessárias para representar uma população de edificações para erros de 5% e 10%, como base na Equação (8).

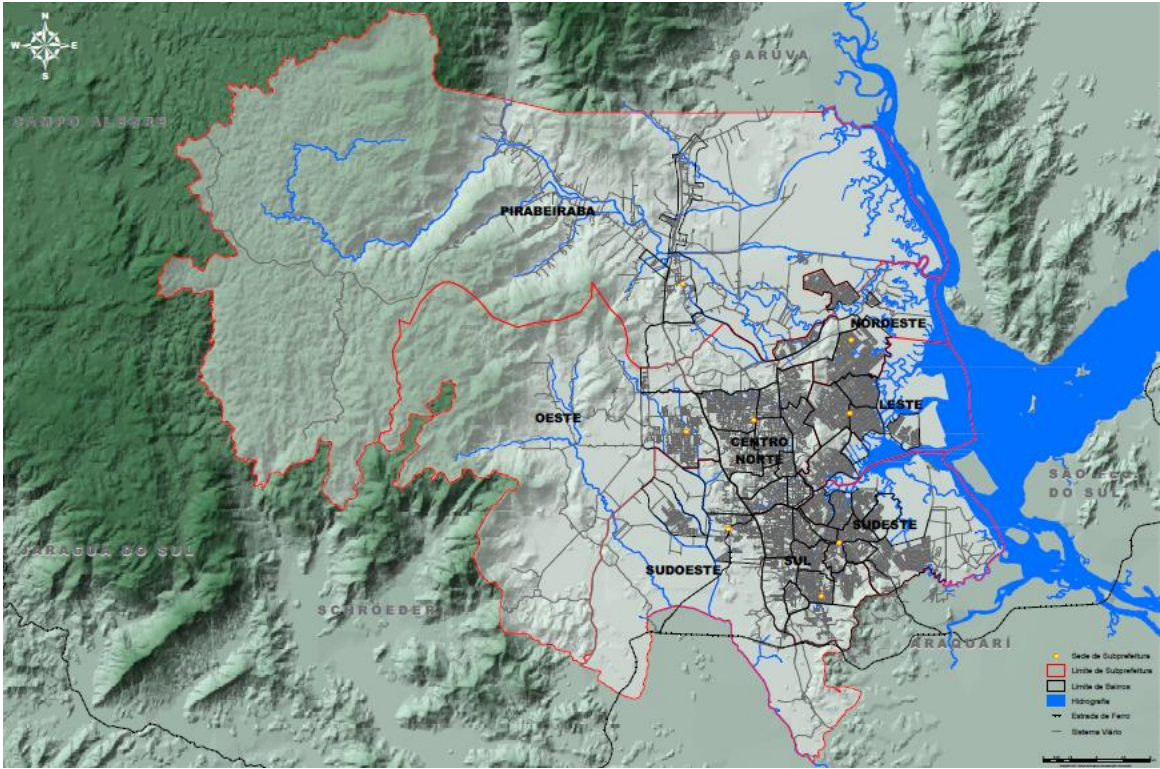
Na opinião de Kutner, Nachtsheim e Neter (2004), o conjunto de dados para a construção do modelo deve ser grande o suficiente para produzir um modelo confiável. Os autores reforçam que, para alcançar esse objetivo, o número de dados na amostra deve ser no mínimo 6 a 10 vezes a quantidade de variáveis candidatas a preditores.

3.3 AMOSTRAGEM

O método estatístico fundamenta-se em procedimentos apoiados na teoria da amostragem e é indispensável no estudo de certos aspectos da realidade em que se pretende obter o grau de correlação entre dois ou mais fenômenos (FACHIN, 2001). Amostragem não probabilística compreende a seleção intencional de indivíduos dentro da população a fim de estabelecer uma amostra representativa do universo populacional (KELLEY *et al.*, 2003; KOTHARI, 2004). Além disso, na amostragem por cotas, a população é dividida em vários subgrupos para compensar a falta de aleatoriedade, segundo características de controle como localidade, nível de escolaridade e faixa de renda. Seleciona-se então uma cota de cada subgrupo, proporcional ao seu tamanho (REIS, 2008).

Em razão disso, no presente estudo, foi empregado o procedimento de amostragem não probabilística por cotas, cuja característica de controle foi o número de unidades residenciais autônomas em cada região de abrangência das subprefeituras do município. A divisão geográfica de Joinville (Figura 5), segundo sua administração municipal, bem como o número de unidades residenciais autônomas em cada região estão presentes no documento Joinville Cidade em Dados, emitido anualmente pelo órgão de planejamento urbano e desenvolvimento sustentável do município (SEPUD 2017, 2019).

Figura 5 – Mapa das regiões de abrangência das subprefeituras de Joinville



Fonte: IPPUJ (2013)

Foi escolhida a amostragem não probabilística por cotas com o propósito de garantir a melhor representatividade da população. Como unidade de amostragem para a coleta de informações, adotou-se a edificação unifamiliar (casa e geminado) visto que representa 83,95% do total de domicílios do município (IBGE, 2010).

Sendo assim, conforme o método adotado, o número de questionários a serem aplicados em cada região obedece ao percentual (cota) das unidades residenciais daquela área geográfica de interesse (Tabela 3).

Tabela 3 – Distribuição da amostra por região e cota populacional

Região	Número de unidades residenciais (2019)	Unidades residenciais em cada região (%)
Centro-Norte	65.747	32,1
Distrital de Pirabeiraba	3.438	1,7
Leste	32.117	15,7
Nordeste	20.900	10,2

Oeste	10.064	4,9
Sudeste	30.151	14,7
Sudoeste	8.900	4,3
Sul	33.654	16,4
Total	204.971	100

O teste qui-quadrado destina-se a verificar se a frequência de determinado acontecimento observado se desvia significativamente ou não da frequência com que ele é esperado (MORETIN; BUSSAB, 2013). A estatística do teste é um tipo de distância entre as frequências observadas e as frequências esperadas (BARBETTA; REIS; BORNIA, 2010). Dessa forma, aplicou-se o teste qui-quadrado a fim de avaliar se havia diferenças significativas entre o número de questionários esperado e o realmente obtido em cada região da cidade.

3.4 ELABORAÇÃO DO QUESTIONÁRIO

A elaboração do questionário consistiu no primeiro passo da obtenção de dados para a pesquisa. Haja vista sua relevância, essa etapa demandou um período extenso de análise e revisão.

A definição de quais informações seriam levantadas foi baseada na revisão da literatura, em que pôde-se verificar as variáveis com maior possibilidade de exercerem influência sobre o consumo residencial de água. O questionário foi dividido em quatro sessões: (i) localização do imóvel e características do entorno; (ii) aspectos socioeconômicos e demográficos dos moradores; (iii) características construtivas e equipamentos instalados na edificação e (iv) hábitos de consumo dos usuários.

O questionário abrangeu 46 questões relacionadas à localização do imóvel, idade e gênero dos moradores, renda domiciliar mensal total e grau de escolaridade do responsável pela conta de água. Menores de 12 anos foram consideradas como crianças, entre 12 e 17 anos foram considerados adolescentes, entre 18 e 64 anos adultos e mais de 65 anos idosos (BRASIL; 1990, 2003). As categorias de renda domiciliar mensal total e grau de escolaridade foram extraídas dos questionários socioeconômicos aplicados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - INEP (2013).

Na investigação também foram levantados dados do projeto arquitetônico e características da edificação como área construída, área do terreno, idade do imóvel, quantidade de banheiros no domicílio, presença de piscina e de sistema alternativo de abastecimento de água (aproveitamento de água da chuva, captação de água subterrânea ou reúso de água).

Em relação aos equipamentos instalados, buscou-se identificar se a residência possuía máquina de lavar roupa, máquina de lavar louça, além de equipamentos economizadores de água como bacia sanitária com duplo acionamento, bacia sanitária com volume nominal de 6 litros/fluxo e arejador nas torneiras. Outros fatores investigados foram hábitos de consumo de água dos usuários, por exemplo, se compravam água mineral para beber, irrigavam o jardim ou lavavam o carro/áreas externas e com que frequência realizavam tais atividades.

As perguntas foram elaboradas em formatos alternativos, algumas com opções de resposta previamente determinadas (questões fechadas) e outras para inclusão de dados numéricos (questões abertas). O modelo de questionário está presente no Apêndice A.

3.5 ESTUDO PILOTO

De acordo com Greco, Walop e McCarthy (1987), para ter confiabilidade nos resultados de uma pesquisa, é importante assegurar que o questionário meça consistentemente aquilo para o qual foi proposto a medir. Em síntese, o questionário deve ser válido e confiável.

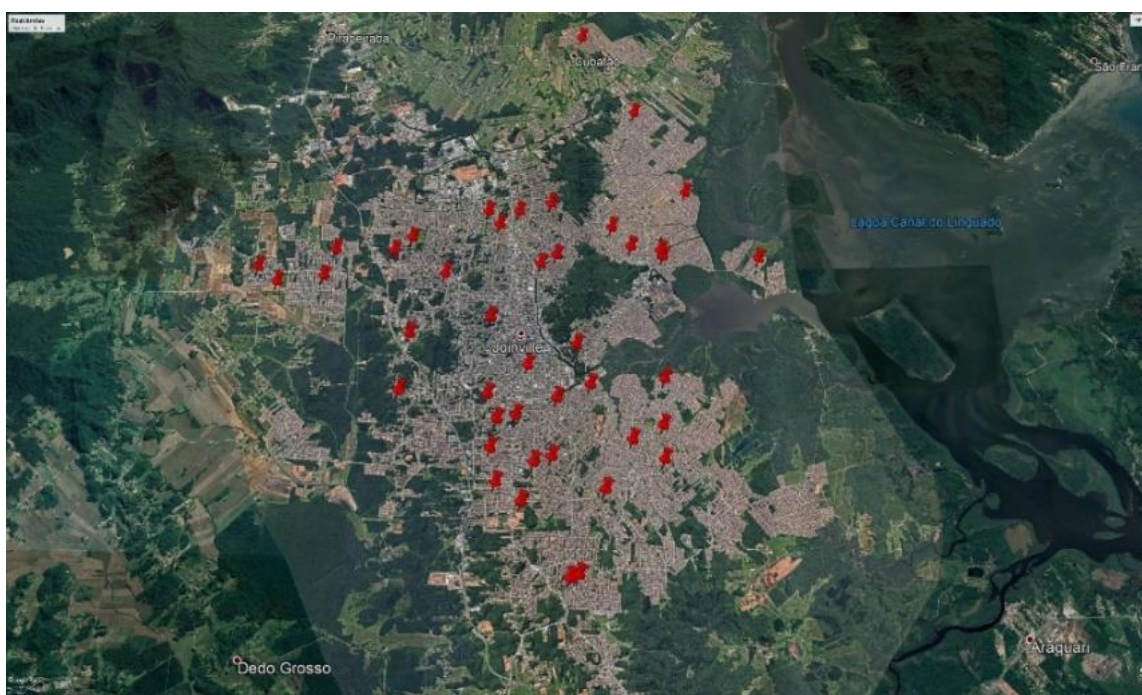
Os métodos mais adotados para esse fim são: validação de face e validação de conteúdo. A validade de face é determinada por uma amostra de respondentes, que dirão se o instrumento está adequado ao objetivo proposto (CALDANA; GABRIEL, 2017).

Inicialmente, um estudo piloto foi conduzido a fim de avaliar a validade de face e confiabilidade do instrumento de pesquisa (questionário), além de contabilizar o tempo requerido para sua aplicação. Um conjunto de indivíduos da amostra respondeu à versão pré-final do instrumento, analisando se a redação textual possibilitava uma compreensão clara das questões.

Nessa etapa, o método de amostragem utilizado foi a amostragem não probabilística por conveniência, em que os indivíduos mais acessíveis são

selecionados para constituir a amostra (KELLEY *et al.*, 2003). Os questionários foram respondidos dentro da Universidade por alunos, professores, técnicos administrativos e funcionários terceirizados. No total, 43 residências unifamiliares constituíram a amostra, distribuídas em 25 bairros da cidade, de acordo com a Figura 6. Foi solicitado aos respondentes que indicassem se realmente compreenderam cada item do questionário e, ainda, que descrevessem eventual dificuldade na interpretação das questões.

Figura 6 – Mapa de Joinville com a localização das residências do estudo piloto



Fonte: Google Earth (2018).

Ao final dessa etapa, percebeu-se que a redação do texto era clara, porém havia a necessidade de incluir algumas questões como o reúso de água da máquina de lavar para fins que não requerem água potável.

3.6 COLETA DE DADOS

Para a coleta de dados, as habitações unifamiliares foram selecionadas com base nas técnicas de amostragem citadas. A aplicação do questionário ocorreu presencialmente na Universidade e, na maioria dos casos, com o envio do questionário por correio eletrônico e redes sociais. No caso *online*, o questionário foi

desenvolvido pela plataforma Google Forms (Apêndice A). Os moradores levaram em torno de 10 minutos para responder o questionário.

A fim de alcançar o tamanho amostral almejado, os questionários também foram aplicados em locais com grande circulação de pessoas como terminais de ônibus e eventos públicos pela cidade. Outro procedimento adotado foi a aplicação presencial do questionário em residências de bairros mais afastados do centro da cidade.

Os valores referentes à área do terreno, área total das residências e idade dos imóveis foram obtidos no cadastro técnico da Prefeitura Municipal de Joinville, por meio do Sistema de Informações Municipais Georreferenciais – SIMGeo e espelho cadastral. Dados sobre a latitude e longitude da residência e tipologia da casa (térrea, sobrado ou geminado) foram extraídas do Google Maps. As leituras de consumo de água e informações sobre coleta de esgoto sanitário foram fornecidas pela Companhia Águas de Joinville – CAJ. Os valores de consumo de água foram lidos mensalmente pelos funcionários da companhia no hidrômetro de cada residência. A concessionária disponibilizou dados do consumo de água do período de dezembro de 2015 até novembro de 2019, totalizando 48 meses de medição.

3.7 MODELAGEM ESTATÍSTICA

Nesta etapa de modelagem estatística, os dados coletados foram avaliados por meio de inferência estatística utilizando os métodos de regressão linear múltipla, regressão robusta e regressão geograficamente ponderada para verificar a relação entre as variáveis estudadas.

As análises estatísticas foram realizadas com o auxílio do *software* R versão 3.5.1 (R CORE TEAM, 2018) e dos pacotes caret (KUHN, 2020), car (FOX; WEISBERG, 2019), lme4 (JEFFREY; HOTHORN, 2002), robustbase (MAECHLER *et al.*, 2020), ggplot2 (WICKHAM, 2016), ggmap (KAHLE; WICKHAM, 2013) e GWmodel (LU *et al.*, 2014; GOLLINI *et al.*, 2015). O nível de significância adotado na pesquisa foi $\alpha = 5\%$.

3.7.1 Variáveis dependentes e independentes

A variável dependente utilizada na análise estatística foi o consumo *per capita* de água (litros por pessoa por dia). Para obter esse indicador de consumo, o consumo bruto de água de cada mês foi dividido pelo número de pessoas que moravam na casa e pelo número de dias do respectivo mês.

Da mesma forma que o questionário, as variáveis independentes foram divididas em grupos: (i) infraestrutura do entorno; (ii) aspectos socioeconômicos e demográficos; (iii) características construtivas; (iv) equipamentos instalados na edificação e (v) hábitos de consumo dos usuários. Um total de 57 variáveis independentes foram levantadas. O Apêndice B apresenta um breve dicionário contendo o nome, tipo, descrição, categorização e fonte de obtenção dessas variáveis.

O modelo estatístico pode envolver tanto variáveis preditoras quantitativas como qualitativas. As variáveis quantitativas são medidas em uma escala numérica (MONTGOMERY; RUNGER, 2016). No estudo, tais variáveis abrangeram o número total de moradores, área do terreno, área total da edificação e idade do móvel. Em algumas situações, é necessário também incorporar variáveis qualitativas ou categóricas. Dessa forma, convém codificar as classes das variáveis qualitativas, podendo assumir os valores 0 ou 1, também denominadas de variáveis *dummy* (KUTNER; NACHTSHEIM; NETER, 2004). Exemplos destas variáveis nesta pesquisa são: tipologia da edificação (1-Térrea, 2-Sobrado, 3-Geminado térreo, 4-Geminado sobrado); presença de piscina (0-Não, 1-Sim) e presença de sistema alternativo de abastecimento de água (0-não possui, 1-aproveitamento de água da chuva, 2-captação de água subterrânea, 3-reúso de água).

3.7.2 Análise exploratória, sazonalidade e correlação

A análise exploratória dos dados foi realizada com o intuito de conhecer as principais características da amostra com o auxílio de tabelas, medidas resumo e gráficos (BARBETTA; REIS; BORNIA, 2010; CUI *et al.*, 2019). Foram adotados métodos da estatística descritiva como o cálculo de medidas numéricas: médias, medianas e desvios padrão além da construção de histogramas, *boxplots* e gráficos de dispersão (DEVORE, 2016). A fim de examinar a representatividade da amostra,

esses dados foram comparados com indicadores de órgãos governamentais para a cidade de Joinville.

O estudo da variável dependente consumo de água consistiu em dois métodos iniciais: a análise de sazonalidade e diferença de médias. A análise da componente sazonal em séries temporais já é implementada em muitos estudos relativos à água como previsão da demanda e qualidade da água (COMUNELLO; DAMINELLI; BORTOLOTTI, 2013; COELHO *et al.*, 2017; SABINO; LAGE; NORONHA, 2017; XU *et al.*, 2019). Sendo assim, essa etapa teve como objetivo verificar a existência de sazonalidade no consumo de água das residências avaliadas, isto é, averiguar se a série temporal é afetada por fatores sazonais como os meses e épocas do ano (HYNDMAN; ATHANASOPOULOS, 2018).

Na análise estatística, em cada mês, foram consideradas a média e mediana do consumo bruto de água, totalizando duas séries temporais. Isto porque pretendia-se avaliar qual das duas medidas de tendência central melhor representaria o conjunto de dados em análises futuras.

Foram então criados seis modelos de regressão considerando como variável dependente o consumo de água (y), sendo média (y_1) e mediana (y_2); e variáveis independentes qualitativas categóricas (x): meses do ano, de janeiro a dezembro (x_1); estações do ano, sendo primavera, verão, outono e inverno (x_2) e épocas semestrais do ano, sendo verão e inverno (x_3).

Dados os modelos, foram aplicados os testes Levene e Bartlett com a finalidade de averiguar a suposição de homocedasticidade da variância (TUFFÉRY, 2011). Por fim, a existência de sazonalidade foi examinada com os testes ANOVA e o teste não-paramétrico de Kruskal-Wallis (MORETTIN; TOLOI, 2006; COELHO *et al.*, 2017).

Dias, Kalbusch e Henning (2018) adotaram a média do consumo mensal de água de um período de dois anos. Por isso, buscou-se identificar se havia diferença significativa entre a média de 24 meses (outubro/2016 a setembro/2018) e 48 meses (outubro/2014 a setembro/2018) de medição mensal do consumo de água. Como a amostra excede 40 residências, não foi necessária a verificação da suposição de normalidade das populações. Nesse caso, o Teorema do Limite Central garante que a diferença de médias possui aproximadamente distribuição normal, independentemente das distribuições das populações subjacentes (DEVORE, 2016).

Para a comparação dos dois grupos, foram aplicados os testes t e seu equivalente não paramétrico Wilcoxon Mann-Whitney. O teste t é robusto em relação

à suposição de normalidade e variâncias iguais. Todavia, os testes não paramétricos, como o Wilcoxon Mann-Whitney, são mais adequados quando os dados provêm de distribuições assimétricas (KRZYWINSKI; ALTMAN, 2014). Além disso, o teste de Wilcoxon é aproximadamente 95% tão eficiente quanto o teste t em grandes amostras (MONTGOMERY; RUNGER, 2016). As amostras foram consideradas independentes e com variância desconhecida. Foi aplicado também o teste de Levene a fim de verificar se as variâncias das duas populações eram iguais.

Essa etapa de análise exploratória foi importante também para detectar a presença de dados atípicos ou *outliers*, já que eles podem resultar em uma interpretação tendenciosa dos dados (TUFFÉRY, 2011; GHOSH *et al.*, 2018). O processo de detecção de *outliers* iniciou pela análise gráfica do boxplot da variável consumo *per capita*. Foi utilizado o método univariado de Tukey (1977), calculando a distância interquartis, em inglês *Inter-Quartile Range* – IQR.

No estudo verificou-se ainda a hipótese de normalidade da variável dependente consumo *per capita* de água pela análise gráfica dos histogramas e aplicando-se o teste de Lilliefors (RAZALI; WAH, 2011). Esse teste compara a distribuição cumulativa dos dados com a distribuição normal esperada e os parâmetros são estimados com base na amostra (ÖZTUNA; ELHAN; TÜCCAR, 2006).

Caso a distribuição não seja normal, o processo de transformação de variáveis pode ser necessário. A técnica empregada foi a transformação logarítmica por ser amplamente utilizada em outros estudos envolvendo consumo de água (TORTELLA; TIRADO, 2011; CRUZ *et al.*, 2017; DIAS; KALBUSCH; HENNING, 2018). Segundo Barbetta, Reis e Bornia (2010), a transformação logarítmica torna distribuições com assimetria positiva (cauda mais longa à direita) em distribuições aproximadamente simétricas. Com isso, tem-se uma situação mais adequada para as análises de correlação e regressão.

Análises bivariadas de correlação foram realizadas a fim de identificar possíveis associações numéricas entre as 57 variáveis independentes e a variável dependente consumo *per capita* de água. No caso de distribuições não normais, os coeficientes mais adequados são tau de Kendall e rho de Spearman, pois calculam a correlação com base na posição (*rank*) dos dados e não a partir de seus valores reais (AKOGLU, 2018). Esses métodos são usados inclusive nas situações em que ao menos uma das variáveis é categórica (MIOT, 2018). Segundo Miot (2018), outra vantagem do uso dos

testes não paramétricos de Kendall e Spearman é que eles não se restringem a correlações lineares, desde que os dados apresentem comportamento monotônico.

Para determinar a correlação e significância de cada variável, adotou-se o coeficiente não paramétrico de Kendall por ser mais robusto e eficiente do que o coeficiente de correlação de Spearman, tornando seu uso mais apropriado na presença de *outliers* (CROUX; DEHON, 2010; MIOT, 2018). Fan *et al.* (2017) também adotaram o coeficiente tau de Kendall em sua pesquisa sobre consumo urbano de água na China.

Como a presente pesquisa abrangeu 57 variáveis, o estudo de correlação propiciou a seleção das variáveis estatisticamente significativas aos níveis de 5% e 10% para a elaboração de modelos de regressão. Dessa forma, foi utilizado o mesmo subconjunto de preditores no desenvolvimento dos modelos de regressão linear múltipla, regressão robusta e regressão geograficamente ponderada.

3.7.3 Modelos de regressão

Dado que a finalidade do estudo é propor um modelo com uma variável resposta quantitativa e mais de um regressor, foi utilizado o método de regressão linear múltipla. De acordo com Montgomery e Runger (2016), nesse método a variável dependente ou de resposta (Y) pode estar relacionada com k variáveis independentes ou regressoras, conforme a Equação (10).

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon \quad (10)$$

Em que:

Y = Variável dependente ou resposta;

x_j = Variáveis independentes (preditoras ou regressoras);

β_j = Coeficientes de regressão;

ϵ = Termo de erro aleatório.

Como o estudo envolve 57 variáveis regressoras, além da análise de correlação, aplicou-se também a regressão *stepwise*, a qual identifica o melhor subconjunto de preditores pela seleção de variáveis. Tal método inicia com o modelo completo e vai eliminando as variáveis, uma de cada vez, verificando em cada etapa

se a variável eliminada pode aprimorar o modelo (NACHTSHEIM; NETER, 2004; DEVORE, 2016).

Montgomery, Peck e Vining (2012) citam diversas técnicas úteis para medir a adequação de um modelo como a significância da regressão e de seus regressores, o coeficiente de determinação múltipla R^2 , a análise de resíduos e a detecção de observações influentes. De acordo com Montgomery e Runger (2016), nos problemas de regressão linear múltipla, testes de hipótese relativos aos parâmetros do modelo são úteis na medida da sua adequação. A análise de variância – ANOVA foi aplicada para testar a significância do modelo como um todo. No caso dos coeficientes da regressão, sua significância foi avaliada pelo teste t de Student. Outra estatística global para avaliar o ajuste da regressão é o coeficiente de determinação múltipla R^2 . Esse coeficiente é interpretado como a proporção da variação de Y que pode ser explicada pelo modelo de regressão (DEVORE, 2016).

Um método amplamente aceito para diagnosticar a presença de multicolinearidade na regressão é o *Variance Inflation Factor* – VIF. Valores de VIF > 10 são um indicativo de que a multicolinearidade pode estar influenciando indevidamente as estimativas dos mínimos quadrados (KUTNER; NACHTSHEIM; NETER, 2004). Na opinião de Gujarati e Porter (2011), a variável será altamente colinear com VIF > 10. Fitzgerald e Auerbach (2016), em seu estudo envolvendo pegada hídrica, analisaram o potencial de colinearidade dos termos regressores cujo VIF > 8,0. Em outra pesquisa similar, Cruz *et al.* (2017) obtiveram valores de VIF < 5, indicando ausência de multicolinearidade no modelo. Connolly e Hagelman III (2015), por sua vez, adotaram um limite de 7,5 para o VIF como forma de explicar associações moderadas entre diversos pares de variáveis independentes.

Nos modelos de regressão, assume-se que os resíduos são independentes e normalmente distribuídos, com média zero e variância constante (BARBETTA; REIS; BORNIA, 2010). Walpole *et al.* (2009) e Devore (2016) afirmam que essas suposições podem ser avaliadas por meio de gráficos de diagnóstico e testes estatísticos.

O estudo da distribuição da variável resposta é um passo importante, pois se os dados de consumo *per capita* seguirem a distribuição normal, os resíduos possivelmente terão distribuição normal. No entanto, a aplicação de teste de normalidade na variável resposta não considera o efeito das variáveis explicativas, o que torna a investigação incompleta. Ernst e Albers (2017) analisaram a aplicação de regressão linear em 172 artigos de 20 jornais de psicologia clínica. Os resultados

indicaram como equívoco mais comum a verificação indevida da normalidade das variáveis em vez da verificação da normalidade dos resíduos. Por isso, a análise dos resíduos é fundamental. Além do teste de normalidade, uma ferramenta muito informativa é a análise do gráfico quantil-quantil (q-q). A visualização do gráfico q-q permite identificar o motivo da fuga de normalidade: desvio de assimetria, curtose, mistura de distribuições e presença de *outliers* (ERNST; ALBERS, 2017).

Dessa forma, a suposição de normalidade foi verificada pelo gráfico de probabilidade normal e teste de Lilliefors. Quanto à homocedasticidade (homogeneidade da variância), o gráfico de resíduos padronizados *versus* valores ajustados foi empregado em conjunto com o teste de Breush-Pagan. A autocorrelação e independência dos resíduos foi avaliada com auxílio do teste de Durbin-Watson e a linearidade pelo teste de Rainbow. Por fim, a presença de observações influentes e possíveis *outliers* foi visualmente investigada pelo gráfico da distância de Cook, bem como pela aplicação do teste de Bonferroni. De acordo com Cook e Weisberg (1982), distâncias de Cook > 1 indicam que os *outliers* serão influentes no conjunto de dados. Essas medidas de diagnóstico são sugeridas por Kutner, Nachsheim e Neter (2004). Na presença de *outliers* ou quando os resíduos das observações não seguem a distribuição normal, as estimativas dos coeficientes de regressão pelo método dos mínimos quadrados podem não ser mais confiáveis. Sendo assim, além do método dos mínimos quadrados, na pesquisa foi adotada também a regressão robusta por fornecer previsões menos sensíveis a *outliers* e ser mais apropriada quando os resíduos das observações não seguem a distribuição normal (TUFFÉRY, 2011).

O estudo de Almetwally e Almongy (2018) comparou seis métodos de estimativa em regressão robusta: M. Humpel, M. Bisquare, M. Huber, S, MM(S) e MM estimador. Os resultados indicaram que no caso de amostras com mais de 100 observações, o melhor método de estimativa é o MM. Por isso, utilizou-se a classe de MM-Estimadores como técnica de estimativa dos parâmetros da regressão linear robusta.

A classe MM estimadores é uma alternativa ao método dos mínimos quadrados. Ela combina os benefícios dos estimadores S e M. O estimador M é robusto a *outliers* da variável resposta, porém não é resistente a *outliers* das variáveis explicativas (*leverage points*). O estimador S é altamente robusto na presença de *leverage points*, todavia sua aplicação é ineficiente. Dessa forma, o estimador MM mantém a robustez e resistência do estimador S enquanto ganha a eficiência do

estimador M (SUSANTI *et al.*, 2014; BRENNER *et al.*, 2019). O processo de validação é a etapa final da construção de um modelo e tem como objetivo investigar a sua capacidade de predição. Um procedimento usual é a divisão dos dados em dois conjuntos: o primeiro é usado para estimar os coeficientes do modelo e o segundo é usado para medir o desempenho preditivo do modelo construído (SNEE, 1977).

Montgomery, Peck e Vining (2012) recomendam que o conjunto de estimativa seja maior que o conjunto de predição, e esse último tenha pelo menos 15 observações a fim de obter uma avaliação satisfatória do desempenho do modelo. Sendo assim, a amostra inicial foi dividida aleatoriamente em 70% para ajuste e 30% para validação do modelo.

Por isso, a fim de validar os modelos de regressão, dados de 120 casas selecionadas aleatoriamente foram incluídos na equação para calcular a previsão do consumo de água e compará-la com o consumo medido pela Companhia Águas de Joinville. Dessa forma, pôde-se verificar o grau de aproximação entre os consumos observado e estimado e comparar o desempenho dos modelos pelo poder de assertividade. As métricas adotadas foram: coeficiente de determinação R^2 , erro absoluto médio percentual – MAPE, raiz quadrada do erro médio – RMSE e Critério de Informação de Akaike – AIC (TUFFÉRY, 2011). A análise permitiu ainda verificar em que situações os modelos não foram assertivos.

Estudos anteriores comprovaram a alta heterogeneidade espacial no uso de água em áreas urbanas, isto é, enfatizaram que o consumo de água pode variar sensivelmente em função da região de análise (CHANG; PARANDVASH; SHANDAS, 2010; KONTOKOSTA; JAIN, 2015; CHANG *et al.*, 2017). Dessa forma, além do modelo de regressão linear múltipla, um modelo de regressão geograficamente ponderada foi criado a fim de verificar as variações espaciais no consumo de água na cidade e compreender como os coeficientes de regressão variam em toda a região geográfica de análise.

O modelo de regressão geograficamente ponderada, em inglês *Geographically Weighted Regression* – GWR, calcula um conjunto exclusivo de parâmetros (intercepto α e coeficientes de regressão β_k) para cada observação (i), definida por coordenadas geográficas (u, v) , conforme a Equação (11) (FOTHERINGHAM; BRUNSDON; CHARLTON, 2002).

$$y_i = \alpha(u_i, v_i) + \sum_{k=1}^m \beta_k(u_i, v_i) x_k(u_i, v_i) + \varepsilon_i \quad (11)$$

Em que:

y_i = Variável dependente na localização i ;

x_k = Variável independente k na localização i ;

m = Número de variáveis independentes;

α = Intercepto na localização i ;

β_k = Coeficiente local de regressão para a variável independente k na localização i ;

ε_i = Erro aleatório na localização i .

Cada equação de regressão é calculada a partir dos valores das variáveis dependentes e independentes naquele ponto de localização (u, v) e ponderada pela pelas observações que se enquadram na largura de banda selecionada. Optou-se por utilizar a função de ponderação Bisquare de Kernel (*bisquare adaptative kernel function*) pois ela fornece pesos que diminuem à medida em que a distância entre o ponto principal e as observações vizinhas aumenta (GOLLINI *et al.*, 2013). A largura de banda foi selecionada automaticamente pelo método de validação cruzada (*leave-one-out cross validation score*), minimizando dessa forma o erro associado à regressão (BOWMAN, 1984).

4 RESULTADOS E DISCUSSÃO

O passo inicial da pesquisa consistiu em delimitar o público-alvo (residências unifamiliares de Joinville), e definir o tamanho amostral. Dessa forma, considerando um erro amostral máximo de 5% e a população de 134.871 casas em Joinville, determinou-se por meio da Equação (9) que a quantidade mínima de residências pesquisadas seria 400.

Kutner, Nachtsheim e Neter (2004) recomendam que o número de observações da amostra deve ser no mínimo 6 a 10 vezes a quantidade de variáveis candidatas a preditores. Dessa forma, como a quantidade de variáveis analisadas foi 57, o tamanho amostral estipulado de 400 residências está contido no intervalo recomendado (342 a 570 casas).

Com a aplicação de questionários via redes sociais, via *e-mail* e presencialmente na Universidade, em terminais de ônibus e em bairros mais afastados do Centro da cidade, obtiveram-se no total 449 questionários preenchidos.

Dessa amostra inicial, foram excluídas casas que possuíam mais de uma unidade consumidora no mesmo terreno ou sala comercial anexa, sem medidor de água individualizado. Também retiraram-se aquelas que tinham menos de 2 anos de idade ou cujo cadastro nos órgãos públicos (Prefeitura e Águas de Joinville) não foi encontrado. Por fim, foram excluídas as residências cujo consumo de água era suprido exclusivamente por água subterrânea (poço), resultando em 400 residências unifamiliares, distribuídas em 39 bairros da cidade de Joinville, de acordo com a Figura 7.

O método de amostragem proposto pressupõe a divisão da amostra por cotas, segundo o percentual de unidades residenciais em cada região de abrangência das subprefeituras de Joinville. Comparando o número de questionários obtidos em cada região com a quantidade de questionários esperada (Tabela 4), percebe-se que os valores estão próximos. Foi então aplicado o teste qui-quadrado para verificar se havia alguma diferença significativa entre esses valores. Como resultado do teste, a probabilidade de significância p-valor resultante (0,455) foi maior que o nível de significância adotado (5%), demonstrando que não havia diferenças significativas entre os valores observados e os esperados. Sendo assim, a amostra se aproxima do que era esperado.

Figura 7 – Mapa de Joinville com a localização das 400 residências unifamiliares

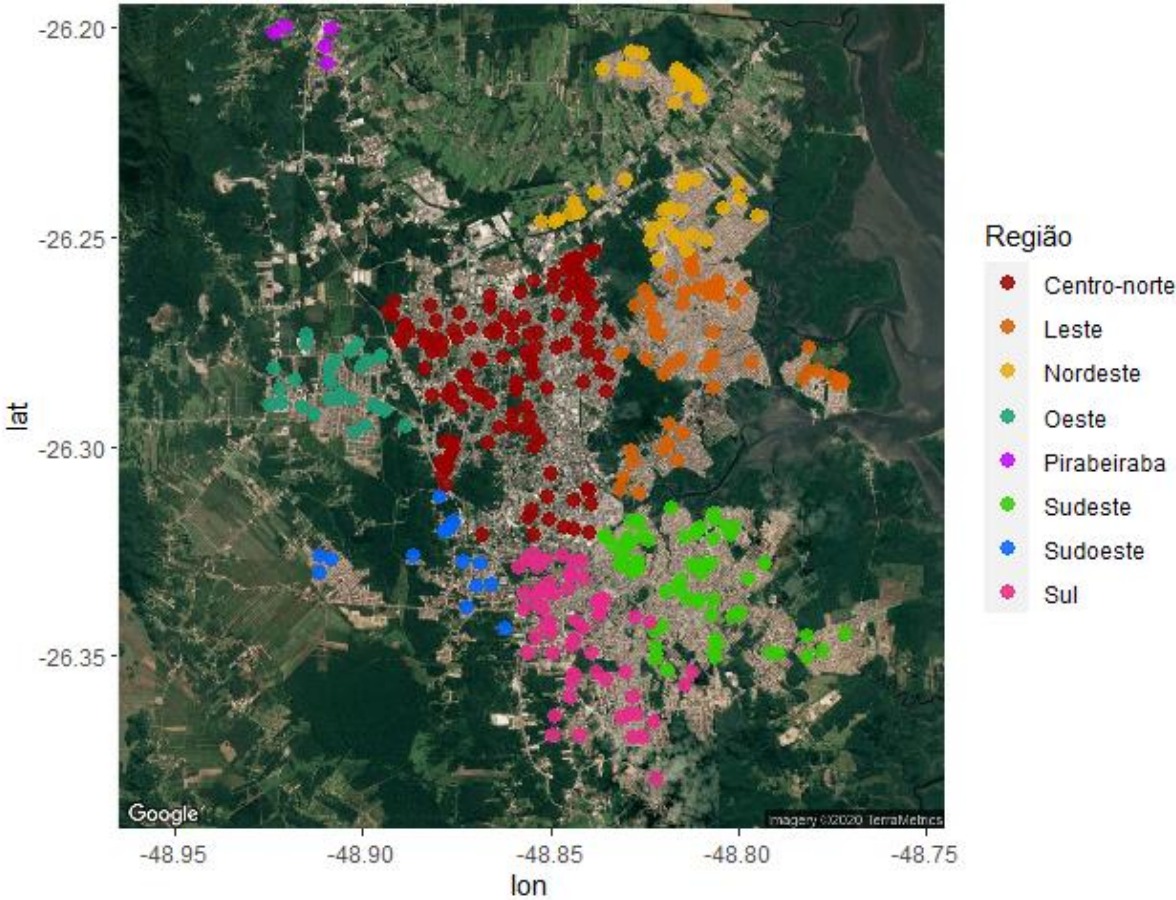


Tabela 4 – Distribuição da amostra por cotas

Região	Número de unidades residenciais (2019)	Unidades residenciais em cada região (%)	Nº questionários esperado em cada região	Nº questionários obtido na região
Centro-Norte	65.747	32,1	128	127
Distrital de Pirabeiraba	3.438	1,7	7	6
Leste	32.117	15,7	62	58
Nordeste	20.900	10,2	41	44
Oeste	10.064	4,9	20	29
Sudeste	30.151	14,7	59	64
Sudoeste	8.900	4,3	17	14
Sul	33.654	16,4	66	58
Total	204.971	100	400	400

A aplicação presencial dos questionários em terminais de ônibus e em bairros mais afastados da região central auxiliou a alcançar a cota estipulada em regiões com déficit inicial de dados como a distrital de Pirabeiraba, Nordeste e Sudoeste. Assim, o método de amostragem considerado juntamente com a estratégia de aplicação dos questionários assegurou uma melhor distribuição espacial da amostra e, consequentemente, melhor representatividade da população.

As características socioeconômicas, demográficas, hábitos de consumo e equipamentos instalados em cada residência foram obtidos diretamente pelo questionário. Os dados de pavimentação da rua, latitude, longitude e tipologia da casa foram extraídos do Google Maps. Por intermédio do sistema de cadastro técnico da Prefeitura Municipal de Joinville, obtiveram-se informações sobre a área do terreno, área total da casa e idade do imóvel. Informações sobre a coleta de esgoto e leituras de consumo bruto de água de 48 meses (dezembro de 2015 a novembro de 2019) foram disponibilizadas pela concessionária local de abastecimento de água, Companhia Águas de Joinville.

Nas seções seguintes serão abordados: as principais características das residências; o consumo de água; análises bivariadas pela correlação entre a variável dependente consumo *per capita* de água e as 57 variáveis independentes, além dos modelos de regressão linear múltipla, regressão robusta e regressão geograficamente ponderada.

4.1 CARACTERIZAÇÃO DA AMOSTRA

Nesta subseção é apresentada a caracterização da amostra. Inicialmente, são descritos os dados quantitativos como aspectos socioeconômicos e demográficos, além de características construtivas das edificações. Em seguida, são mostradas informações qualitativas relacionadas à infraestrutura do entorno, presença de equipamentos e dispositivos hidráulicos bem como os hábitos de consumo dos usuários.

Com o propósito de caracterizar a amostra, a análise das variáveis quantitativas das 400 residências é apresentada na Tabela 5. Observa-se que a ocupação média das residências foi 3,41 pessoas. Os dados do censo demográfico (IBGE, 2010) mostram que a população de Joinville em 2010 foi 515.288 habitantes, que existiam 134.871 casas e estas representavam 83,95% dos domicílios. Dessa forma,

calculando a ocupação média das residências em Joinville no ano de 2010, obtém-se o valor de 3,21 habitantes por casa.

Em termos de composição familiar, o número médio de adultos (18 a 64 anos) foi 2,47. O número médio de crianças e adolescentes (<17 anos) e idosos (≥65 anos) foi: 0,60 e 0,33, respectivamente. No último censo demográfico, os idosos (≥65 anos) correspondiam a 8,79% da população de Joinville (IBGE, 2010) e na amostra eles somam 9,68%, dessa forma foram bem representados pela amostra. As crianças e adolescentes entre 0 e 14 anos representavam 21,83% da população de Joinville no ano de 2010 (IBGE, 2010) e na amostra o percentual obtido foi 17,61%.

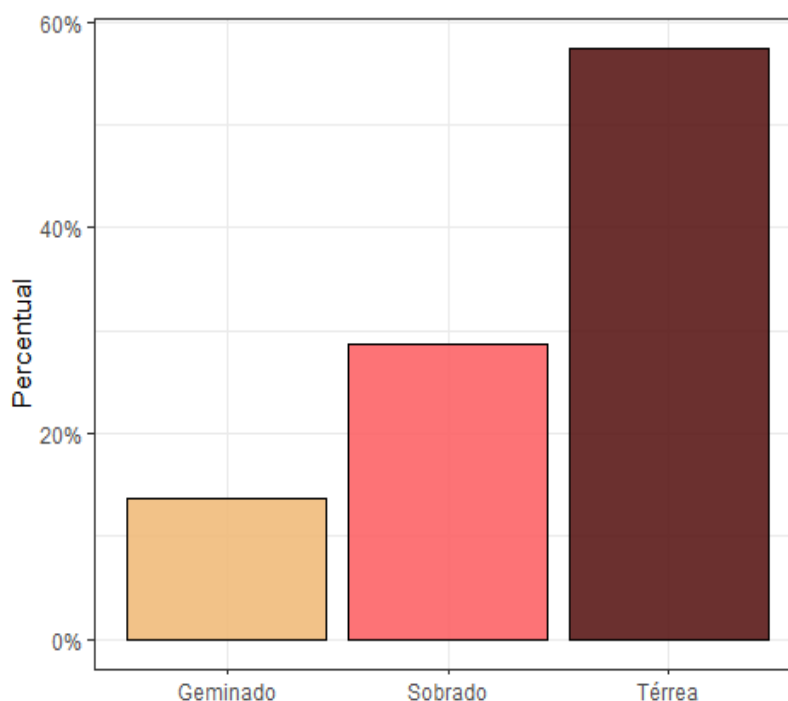
Conforme a Secretaria de Planejamento Urbano e Desenvolvimento Sustentável - SEPUD (2018), em 2017 as mulheres e os homens somaram 50,23% e 49,77% da população de Joinville, respectivamente. Na amostra analisada, as mulheres representaram 51,36% e os homens 48,64% dos 1.363 moradores das residências. Percebe-se então a proximidade entre os valores da amostra e os da população de Joinville.

Tabela 5 – Análise exploratória das 400 residências

Variáveis numéricas	Mínimo	Mediana	Média	Máximo	Desvio padrão
Número de moradores	1,00	3,00	3,41	8,00	1,14
Número de mulheres	0,00	2,00	1,75	5,00	0,84
Número de crianças (<12 anos)	0,00	0,00	0,39	3,00	0,70
Número de adolescentes e jovens (12-17 anos)	0,00	0,00	0,21	3,00	0,47
Número de adultos (18-64 anos)	1,00	2,00	2,47	6,00	1,07
Número de idosos (≥65 anos)	0,00	0,00	0,33	3,00	0,65
Área do terreno (m²)	57,26	376,50	457,36	7.597,00	519,48
Área construída (m²) (Inclui casa, garagem, edícula)	40,00	133,72	151,53	566,22	76,06
Área da casa (m²)	34,00	100,00	118,22	539,22	65,84
Número de quartos	1,00	3,00	3,01	7,00	0,84
Número de banheiros	1,00	2,00	2,50	8,00	1,13
Idade do imóvel (anos)	2,00	19,50	20,68	81,00	11,48

Quanto às características construtivas das edificações, a área do terreno variou de 57,26m² (em um geminado) até 7.597,00m², com média de 457,36m². As residências analisadas possuem área construída de, em média, 118,22m², 3,01 quartos, 2,50 banheiros e 20,68 anos. Das 400 residências, 57,5% eram construções térreas, 28,7% pertenciam à tipologia sobrado e 13,8% eram geminados (Figura 8).

Figura 8 – Tipologia das residências



Em 2010, 89,15% das unidades domiciliares urbanas em Joinville eram próprias quitadas ou financiadas, enquanto 10,68% estavam alugadas ou cedidas e a situação “outros” abrangeu 0,16% dos domicílios (IBGE, 2010). Na amostra, 90,5% das casas estavam financiadas ou quitadas, isto é, sua construção ou compra havia sido parcialmente ou completamente paga, enquanto 9,5% das casas eram alugadas ou cedidas.

Como mostra a Figura 9, a renda mensal total dos membros das famílias da amostra segue uma distribuição aproximadamente simétrica, com dados concentrados em torno da média. As famílias apresentaram renda média mensal de 3 a 6 salários mínimos – SM (R\$2.862,01 a R\$5.724,00), de acordo com o SM vigente em 2018. Segundo o IBGE (2010), em 2010 a renda média domiciliar per capita no município de Joinville era de 2,18 SM. A Figura 10 apresenta a distribuição de renda domiciliar, em número de SM, em Joinville no ano de 2010. São gráficos com

categorias e períodos distintos. Todavia, observa-se que em geral, a distribuição de renda da amostra apresentou comportamento análogo ao da população, com exceção da classe de renda superior a 15 SM.

Figura 9 – Renda familiar mensal da amostra

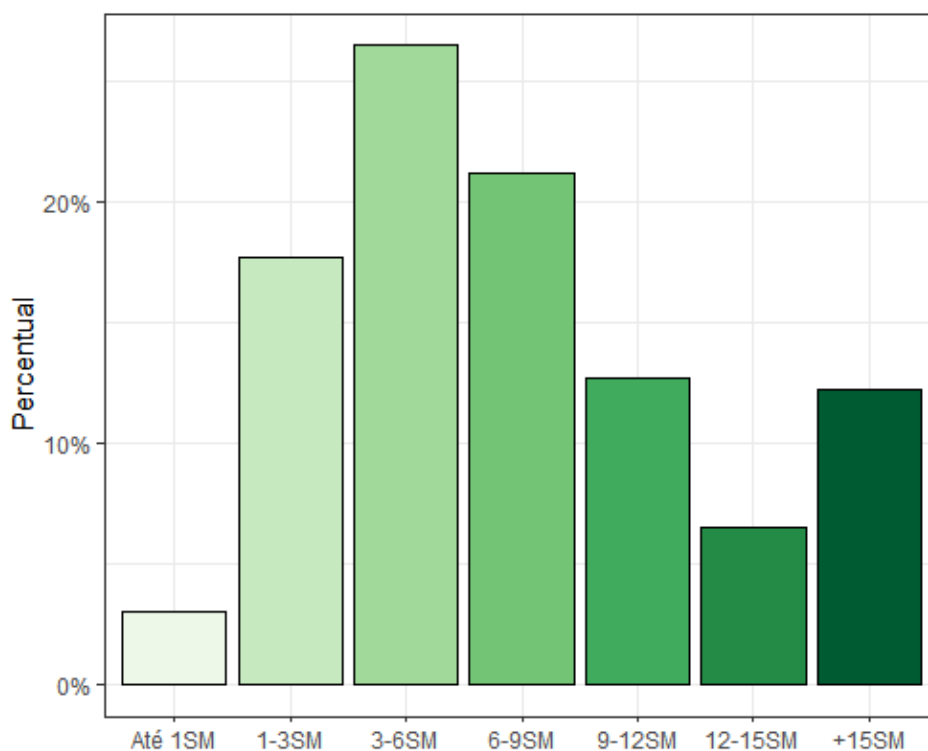
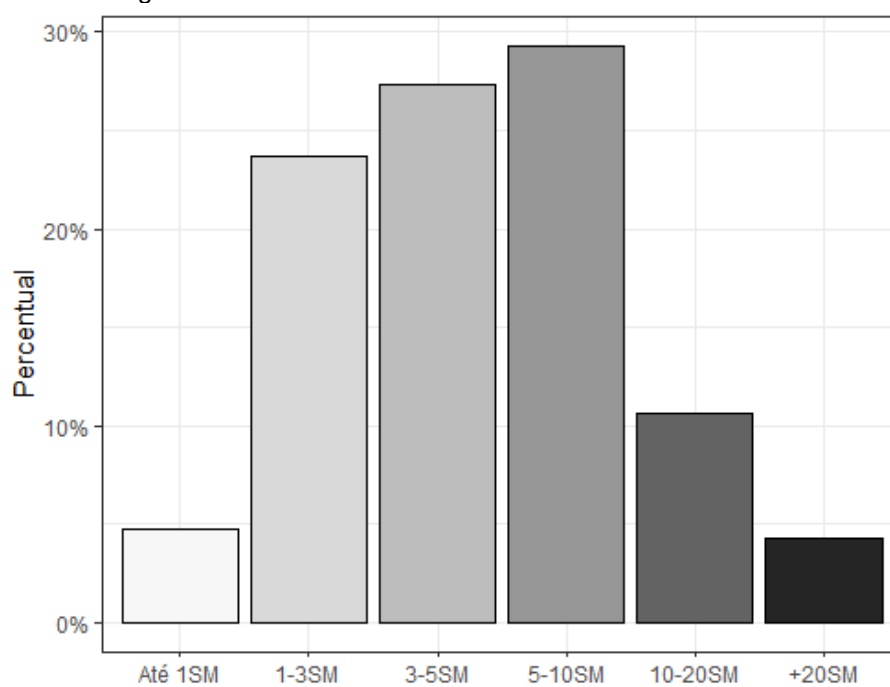


Figura 10 – Renda domiciliar em Joinville no ano de 2010



Fonte: Adaptado de IBGE (2010).

Definiu-se como critério de avaliação o grau de escolaridade do titular da fatura de água. Na amostra analisada, 19,5% dos responsáveis pela conta de água possuíam ensino fundamental (completo ou incompleto), 29,8% concluíram o ensino médio, 21,2% tinham alguma graduação e 29,5% eram titulares de pós-graduação (Figura 11).

A Figura 12 apresenta o grau de escolaridade da população de Joinville com 25 anos ou mais de idade no ano de 2010, conforme o último censo demográfico. Oficialmente não existem informações sobre pós-graduação para o município, o que impede a comparação direta dos dados. Porém, percebe-se que na época já havia um grande percentual de pessoas com no mínimo ensino fundamental completo.

A análise da representatividade da amostra quanto às variáveis socioeconômicas foi realizada com os dados oficiais mais recentes, sendo a maioria do último censo demográfico de 2010. No entanto, existe uma defasagem de 10 anos, que afetou principalmente a análise de renda e grau de escolaridade. Dessa forma, estudos futuros podem aprimorar essa avaliação com a utilização de dados do próximo censo demográfico de 2020.

Figura 11 – Grau de escolaridade do responsável pela conta de água da amostra

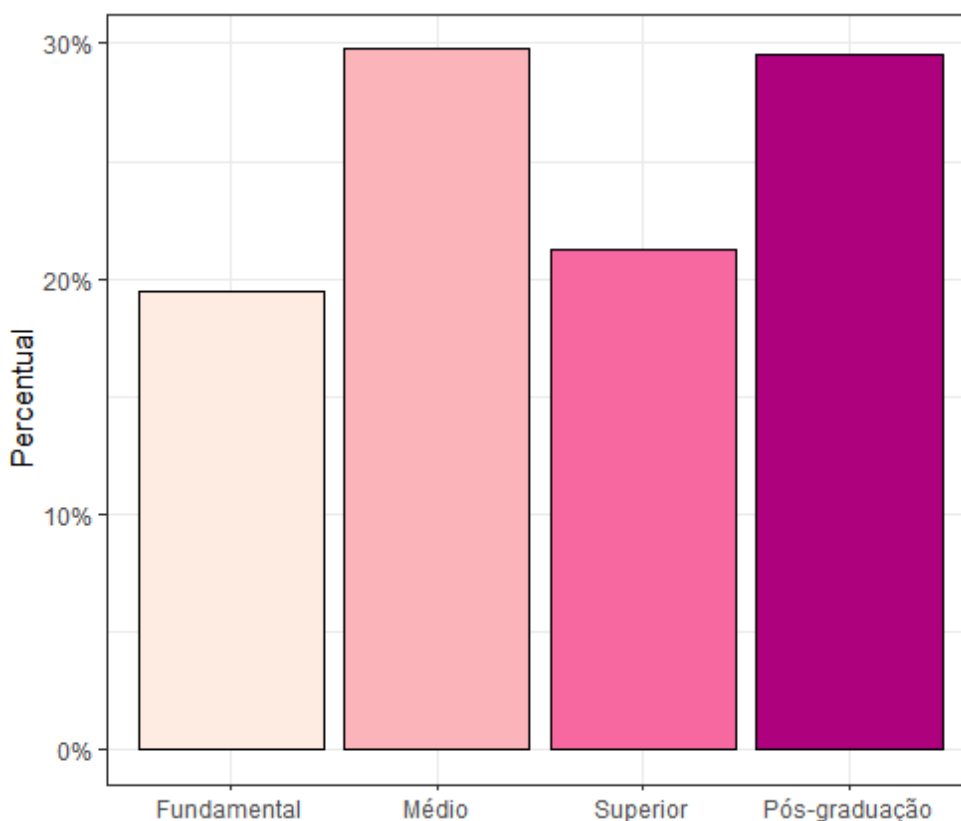
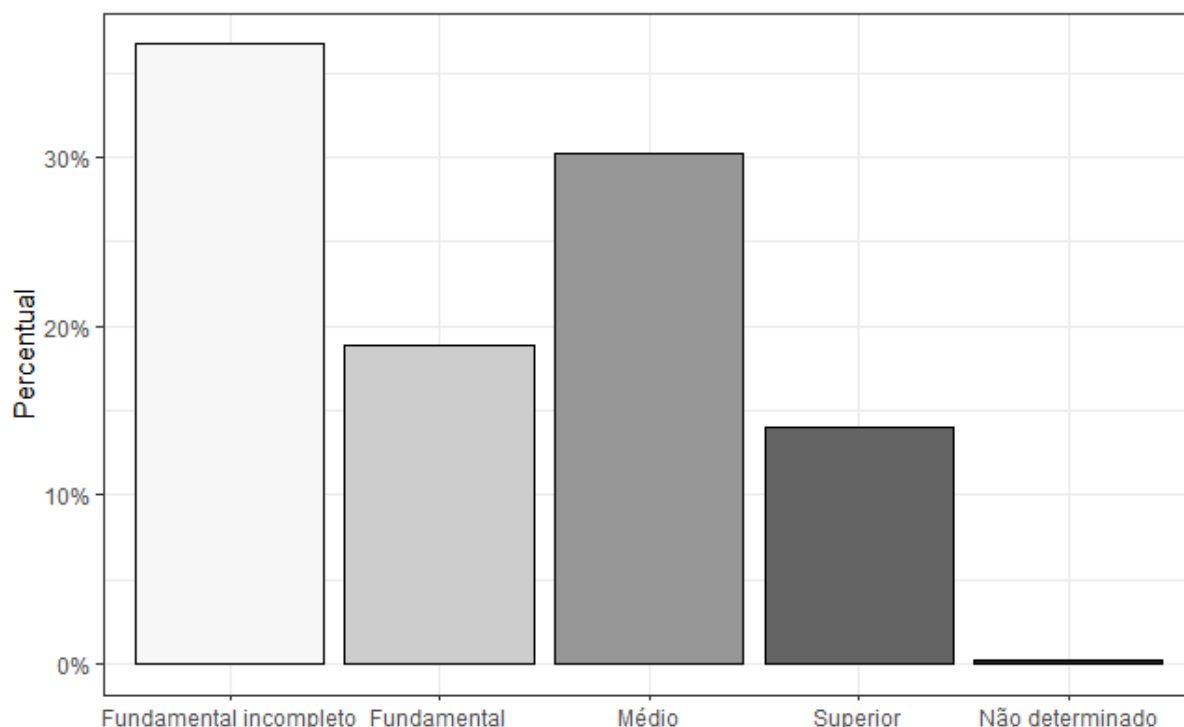


Figura 12 – Grau de escolaridade da população de Joinville com 25 anos ou mais de idade no ano de 2010



Fonte: Adaptado de IBGE (2010).

A Tabela 6 apresenta os resultados obtidos na análise das variáveis dicotômicas (*dummy*), codificadas como 0-Não e 1-Sim. Foi observado que 70,5% das residências da amostra possuíam pavimentação na via (asfalto, piso intertravado de concreto – *paver*, lajota sextavada de concreto, entre outros). Em 2010, Joinville apresentava 48% dos domicílios urbanos em vias públicas com presença de bueiro, calçada, pavimentação e meio-fio (IBGE, 2010). No ano de 2016, 58,9% das vias possuíam asfalto e calçamento (SEPUD, 2017).

Com relação ao esgoto sanitário, apenas 33,0% das casas da amostra eram atendidas pelo serviço municipal de coleta de esgoto. Os resultados são coerentes com o dado publicado pela Secretaria de Planejamento Urbano e Desenvolvimento Sustentável – SEPUD (2019), que aponta que 31,8% dos habitantes de Joinville são atendidos pela coleta e tratamento público de esgoto sanitário.

Tabela 6 – Análise das variáveis dicotômicas

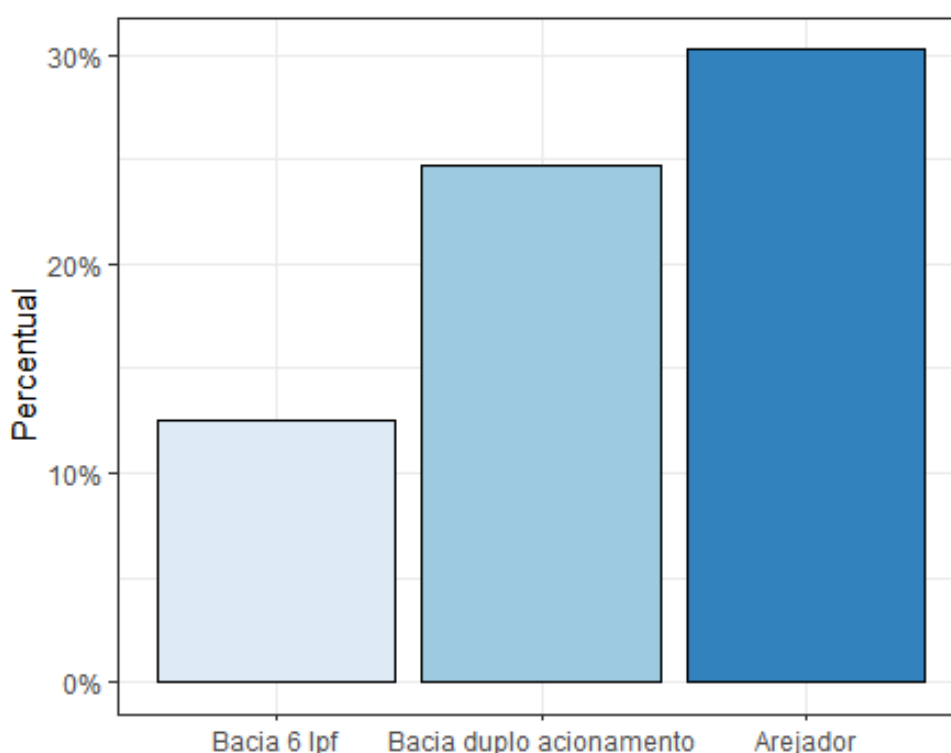
Variáveis dicotômicas (<i>dummy</i>)	Sim (%)	Não (%)
Pavimentação na via	70,5	29,5
Existência de rede coletora de esgoto na via	33,0	67,0
Presença de piscina	17,0	83,0
Existência de reservatório	93,0	7,0
Existência de pressurizador	13,8	86,2
Falta de água na residência	4,5	95,5
Presença de dependência de empregada	4,5	95,5
Presença de banheira	14,0	86,0
Presença de sistema alternativo abastecimento (aproveitamento água da chuva, captação de água subterrânea ou reúso de água)	6,8	93,2
Presença de tanquinho	35,5	64,5
Presença de máquina de lavar roupa	98,0	2,0
Presença de máquina de lavar louça	9,5	90,5
Presença de equipamento economizador de água (bacia sanitária com duplo acionamento, bacia com 6lp ou arejador na torneira)	43,5	56,5
Família costuma comprar água mineral para beber	27,5	72,5
Família reusa a água da máquina de lavar roupas	30,2	69,8
Família lava o carro em casa	58,8	41,2
Família costuma lavar a calçada/áreas externas	69,5	30,5
É feita irrigação do jardim da casa	25,0	75,0

Sobre os aspectos construtivos, 17,0% das residências possuem piscina e 13,8% utilizam pressurizador de água. Somente 7,0% das casas não possuem reservatório, possivelmente em decorrência da Lei Ordinária Municipal 2260/88, que torna obrigatória a existência de reservatório de água potável em residências unifamiliares construídas a partir de 1988 (PREFEITURA MUNICIPAL DE JOINVILLE, 1988). A NBR 5626 (ABNT, 1998) também obriga a instalação de reservatório para uso doméstico, sendo seu volume o necessário para 24h de consumo normal da edificação, com a recomendação de no mínimo 500L de reserva. Em 4,5% das residências houve falta de água. Apenas 6,8% possuem algum sistema alternativo de

abastecimento de água, seja aproveitamento de água da chuva, captação de água subterrânea (poço) ou reúso de água.

A investigação revelou que 43,5% das casas tinham algum equipamento economizador de água. De modo específico, 30,3% possuíam arejador na(s) torneira(s), 24,8% tinham bacia sanitária com duplo acionamento e a bacia sanitária com 6 litros por fluxo estava presente em 12,5% das residências, resultados apresentados na Figura 13.

Figura 13 – Equipamentos economizadores de água



A investigação dos hábitos de consumo dos usuários mostrou que 27,5% costumam comprar água mineral para beber, 58,8% lavam o carro em casa, 69,5% costumam lavar as áreas externas/calçada e apenas 25% irrigam o jardim. O equipamento mais utilizado na lavagem do carro foi o balde, seguido da mangueira e máquina de alta pressão. No caso da lavagem da calçada e áreas externas, os moradores utilizaram principalmente a máquina de alta pressão e o balde foi o menos usado. Além disso, 30,2% dos moradores informaram que reusam a água da lavadora de roupas em atividades como lavagem da calçada e áreas externas e descarga da bacia sanitária.

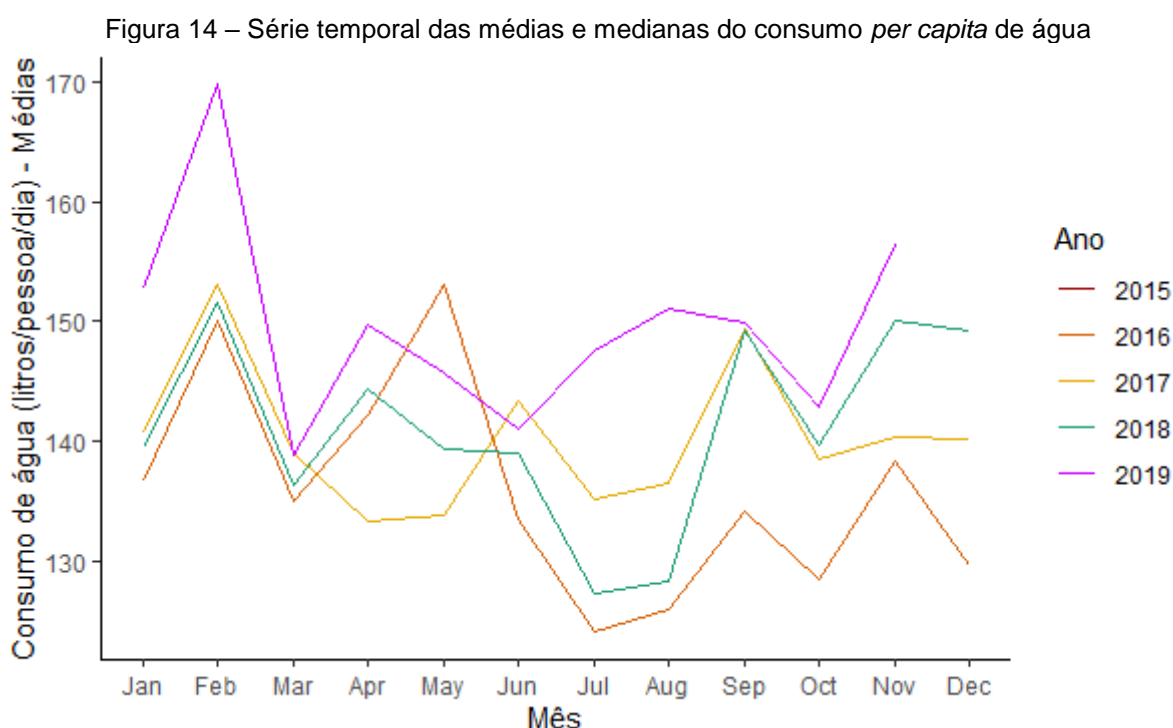
4.2 ANÁLISE DO CONSUMO DE ÁGUA

A análise do consumo *per capita* de água das residências avaliadas consistiu em: (i) verificar a presença de sazonalidade determinística; (ii) determinar se havia diferença significativa entre o consumo médio de água de 24 e 48 meses e (iii) descrever as principais características do indicador de consumo (em litros por pessoa por dia), além de examinar a aderência do conjunto de dados à distribuição normal.

4.2.1 Verificação da sazonalidade determinística

Nesta seção, são apresentados os resultados do estudo das séries temporais de consumo *per capita* de água, além da verificação da presença de sazonalidade. Primeiramente, foi observado o comportamento das séries de média e mediana do consumo *per capita* de água no período de dezembro de 2015 a novembro de 2019. Como ilustra a Figura 14, no mês de fevereiro houve um pico de consumo e nos meses de julho e agosto uma queda considerável, comportamento esse que se repete na maioria dos anos avaliados.

Em relação às características da amostra, a Tabela 7 apresenta a estatística descritiva das séries temporais de média e mediana do consumo *per capita* de água considerando os 48 meses de medição, em litros por pessoa por dia.



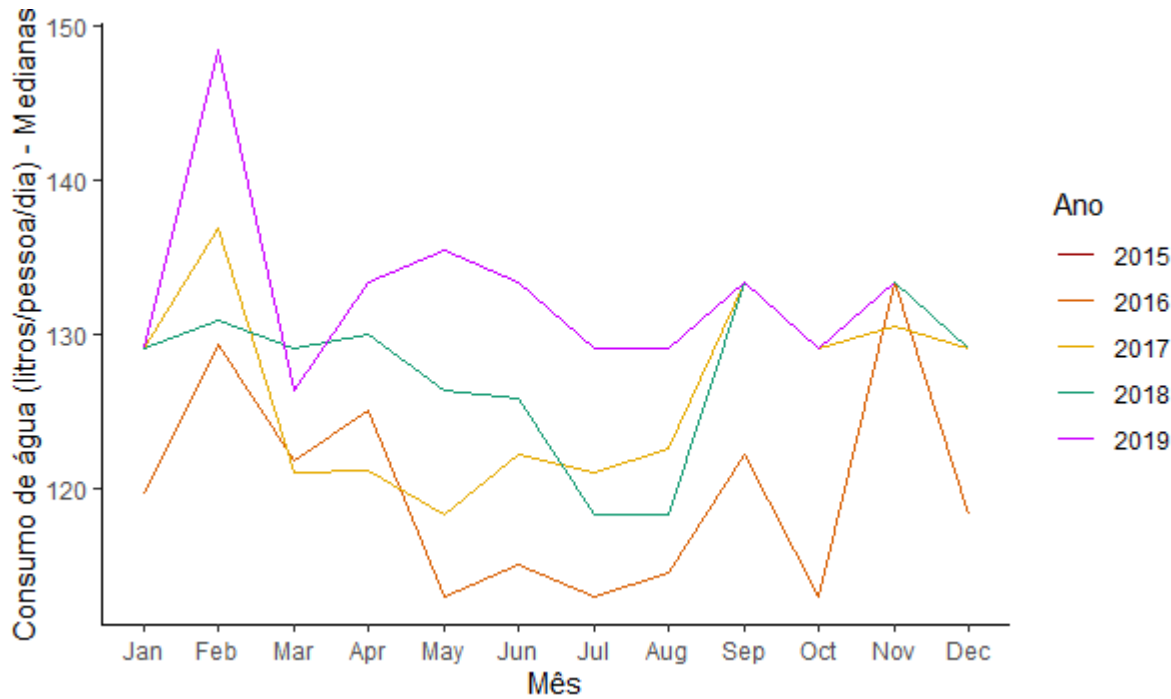
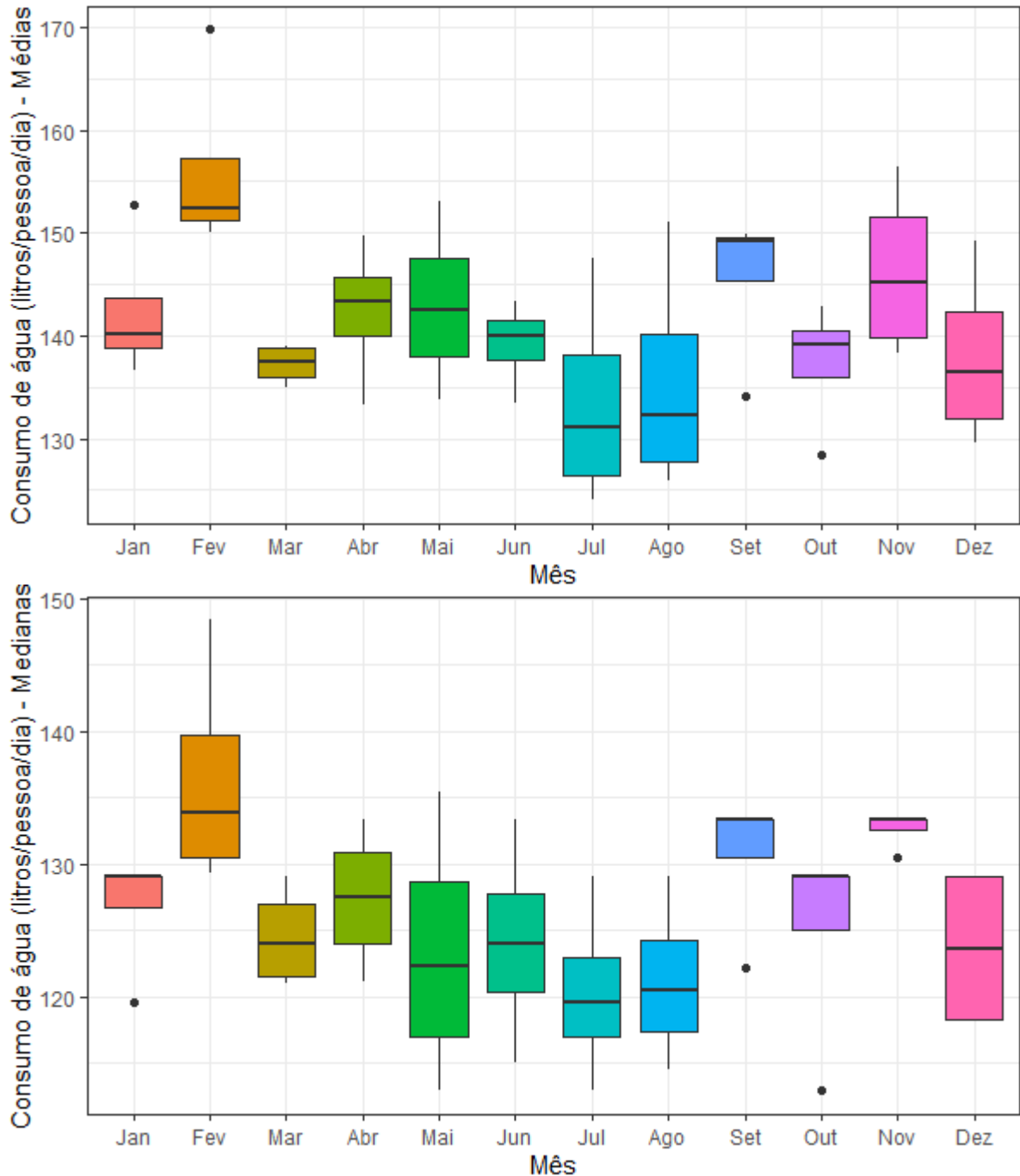


Tabela 7 – Estatística descritiva da variável consumo *per capita* de água

Variável	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Desvio padrão
Consumo <i>per capita</i> Médias (litros/pessoa/dia)	124,0	135,1	139,9	141,4	149,3	169,7	9,1295
Consumo <i>per capita</i> Medianas (litros/pessoa/dia)	112,9	121,0	129,0	126,3	130,7	148,4	7,3649

A análise da Figura 15 revela uma variabilidade no consumo *per capita* de água considerando os diferentes meses do ano. Por exemplo, em janeiro (2016 a 2019), o consumo médio de água estava próximo de 140 litros por pessoa por dia. Já nos meses de julho e agosto do período considerado, a média diminuiu para 130 litros por pessoa por dia e novamente o consumo voltou a ficar próximo de 140 litros por pessoa por dia em novembro e dezembro.

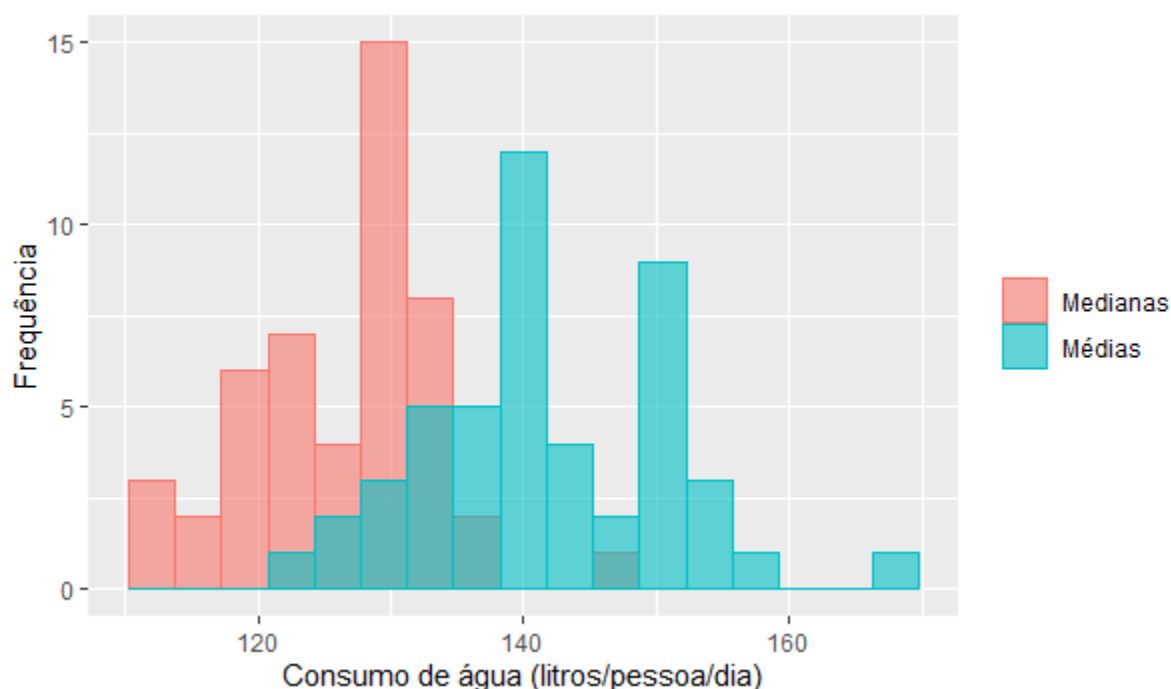
Figura 15 – Boxplots do consumo médio e mediano de água em relação aos meses do ano



Com o intuito de averiguar as suposições de normalidade, a Figura 16 apresenta o histograma da variável dependente consumo *per capita* de água, considerando a média e mediana das 400 residências. Ao se observar os histogramas (Figura 16), percebe-se que as médias do consumo de água apresentam indícios de serem superiores às medianas, possivelmente influenciadas por valores altos de consumo. A aplicação do teste de normalidade relevou que os dados de consumo médio de água atenderam à suposição de normalidade (p -valor = 0,2135). Todavia, ao

considerar as medianas do consumo de água, o p-valor obtido foi de 0,02533, abaixo do nível de significância adotado (0,05), demonstrando que a distribuição normal não era adequada para representar a amostra.

Figura 16 – Histograma das médias e medianas do consumo de água (litros/pessoa/dia)



A fim de avaliar a homocedasticidade das variâncias, foram aplicados o teste paramétrico de Bartlett para o consumo médio e seu equivalente não paramétrico de Levene no caso do consumo mediano de água. Como resultado, as probabilidades de significância foram bem superiores a 0,05, sendo assim as variâncias das populações são iguais e a suposição de homocedasticidade das variâncias foi atendida (Tabela 8).

Utilizou-se também a análise de variância (ANOVA) e o teste não paramétrico de Kruskal-Wallis tendo em vista os seis modelos e o atendimento à suposição de normalidade. Como os dados de consumo médio de água seguiam a distribuição normal, a sazonalidade foi avaliada pela ANOVA. Já no caso do consumo mediano, em virtude de os dados não serem normalmente distribuídos, utilizou-se o teste de Kruskal-Wallis na análise de sazonalidade.

De acordo com a Tabela 8, as probabilidades de significância (p-valor) resultantes foram inferiores a 0,05 na maioria das situações. A exceção foi o modelo de estações do ano (primavera, verão, outono e inverno) com variável resposta o

consumo médio de água, que apresentou p-valor de 0,0667. Mesmo assim, a probabilidade de significância resultante está bem próxima do nível de significância adotado (0,05).

Tabela 8 – Probabilidades de significância da ANOVA e dos testes Bartlett, Levene e Kruskal-Wallis

Variável resposta	Fator	Variância	Sazonalidade
		Valor p	Valor p
Consumo <i>per capita</i> em litros/pessoa/dia Médias (y1)	Meses do ano (x1)	0,6720 ^a	0,0254 ^c
	Estações do ano (x2)	0,3303 ^a	0,0667 ^c
	Épocas semestrais do ano (x3)	0,3511 ^a	0,0253 ^c
Consumo <i>per capita</i> em litros/pessoa/dia Medianas (y2)	Meses do ano (x1)	0,6742 ^b	0,0265 ^d
	Estações do ano (x2)	0,9191 ^b	0,0192 ^d
	Épocas semestrais do ano (x3)	0,5879 ^b	0,0048 ^d

^a – Bartlett ^b – Levene ^c – ANOVA ^d – Kruskal-Wallis

Logo, há evidências para rejeitar a hipótese nula de que tanto a série de consumo médio de água como a de medianas não possuem sazonalidade determinística. Como conclusão dos resultados obtidos, pode-se então inferir que existe sazonalidade determinística no consumo *per capita* de água das residências avaliadas, seja considerando os meses, estações ou semestres do ano.

4.2.2 Comparação entre consumo médio de água de 24 e 48 meses

Essa etapa objetivou identificar se havia diferença significativa entre a média de 24 meses (dezembro/2017 a novembro/2019) e 48 meses (dezembro/2015 a novembro/2019) de medição mensal do consumo *per capita* de água. Isto porque Dias, Kalbusch e Henning (2018) adotaram a média do consumo mensal de água de um período de dois anos.

A verificação da suposição de normalidade não foi necessária, pois a amostra excede 40 residências. Sendo assim, o Teorema do Limite Central garante que a diferença de médias possui aproximadamente distribuição normal, independentemente das distribuições das populações subjacentes (DEVORE, 2016).

Primeiramente, verificou-se a suposição de variâncias iguais pelo teste de Levene. Como a probabilidade de significância resultante (p-valor=0,7641) foi superior

a 0,05, constatou-se que as variâncias das duas populações eram iguais. A aplicação dos testes t e Wilcoxon Mann-Whitney resultou em probabilidade de significância de 0,4547 e 0,3166, respectivamente. Esses resultados corroboram a hipótese de que não há diferenças significativas entre o consumo médio de água dos últimos 48 e 24 meses ($p\text{-valor} > 0,05$). Por isso, optou-se por utilizar a média do consumo *per capita* de água considerando um período de 24 meses, de dezembro de 2017 a novembro de 2019.

Para evitar problemas de sazonalidade, o presente estudo limitou-se a utilizar o consumo médio de um período de 24 meses. Todavia, estudos futuros podem incorporar a sazonalidade por meio da elaboração de modelos de regressão que considerem como variável resposta o consumo de água no verão e no inverno (HUSSIEN; MEMON; SAVIC, 2018). Chang *et al.* (2017) adotaram essa ideia em sua pesquisa no Oeste dos EUA e criaram três modelos de regressão considerando: (i) média do consumo de água de todo o período; (ii) média do consumo de água no inverno (dezembro-fevereiro) e (iii) média do consumo de água no verão (junho-setembro).

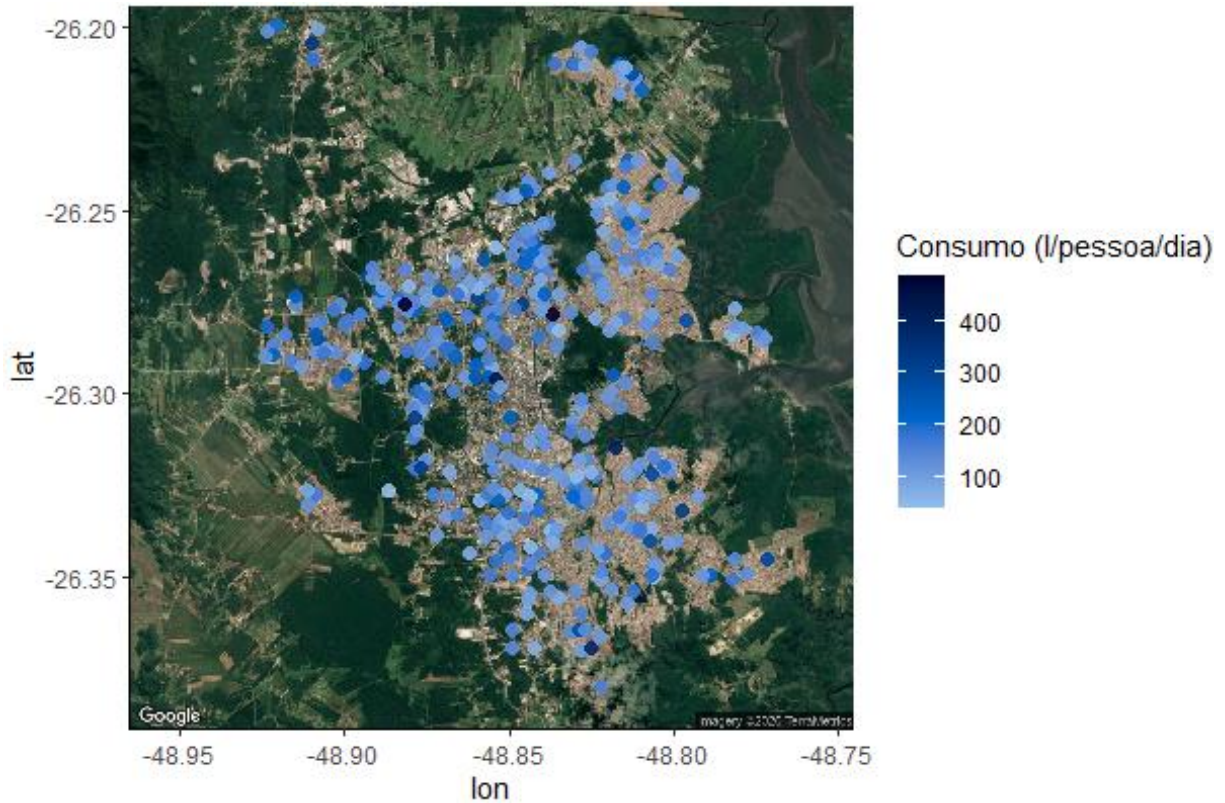
Outra opção é o uso de dados em painel, os quais permitem incluir a componente sazonal na estimativa do uso residencial de água, como no estudo de Makwiza *et al.* (2018). O modelo de regressão com dados em painel possui uma característica peculiar: combina séries temporais e dados de corte transversal, ou seja, a mesma unidade em corte transversal é pesquisada ao longo do tempo (GUJARATI; PORTER, 2011). Por exemplo, o consumo de água das residências unifamiliares de Joinville no período de 2019 a 2020, que combina 9600 observações (400 casas x 24 meses).

4.2.3 Variação espacial

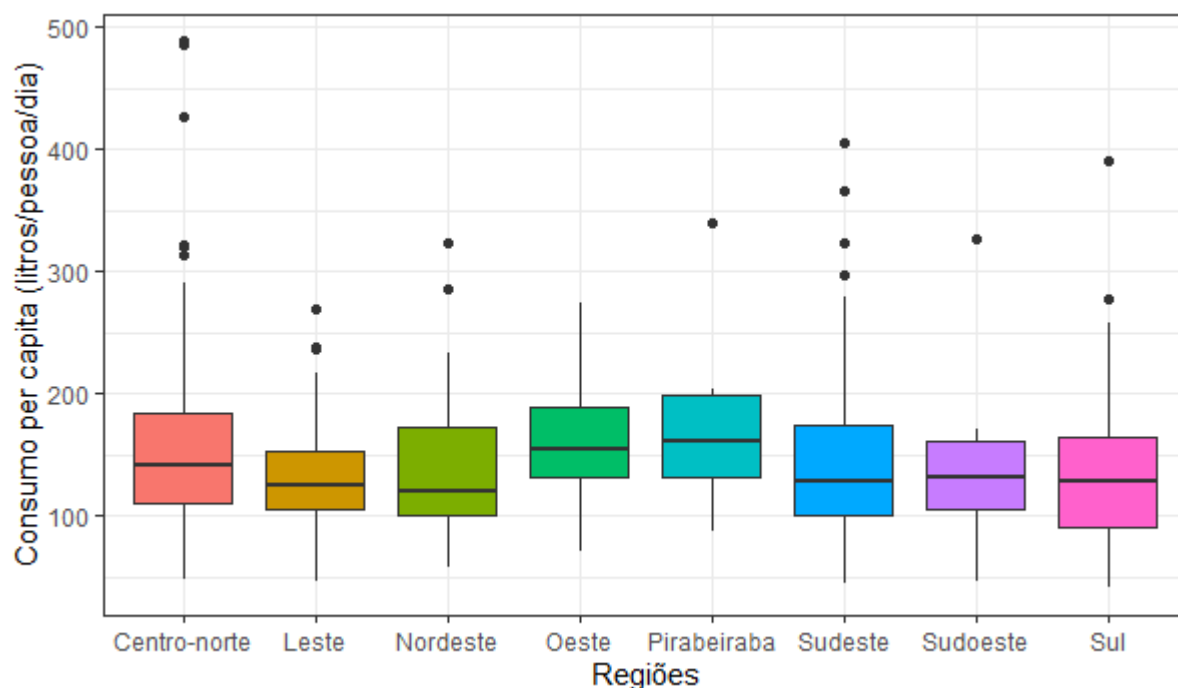
A Figura 17 ilustra a variação espacial no consumo *per capita* de água ao longo da área de estudo. A fim de comparar o consumo *per capita* das diferentes regiões de Joinville, foi aplicado o teste não paramétrico de Kruskal-Wallis. Como resultado, a probabilidade de significância ($p\text{-valor} = 0,0516$) ficou entre 5% e 10% e bem próxima ao limite de 5%, mostrando que a diferença é significativa ao nível de 10%. Esse resultado sugere a possível existência de diferenças significativas no consumo entre as regiões da cidade. Em função disso, optou-se por adotar a regressão

geograficamente ponderada na análise espacial do consumo de água. Estudos futuros podem avaliar a correlação espacial entre o consumo per capita de água e as regiões da cidade por meio do índice de Moran I (MORAN, 1950; LI *et al.*, 2017; CHANG *et al.*, 2017).

Figura 17 – Variação no consumo *per capita* de água das 400 residências



Percebe-se na Figura 18 que a região Centro-norte abriga as casas com maior consumo *per capita* de água. As regiões Oeste e Pirabeiraba possuem mediana de consumo superior às demais regiões.

Figura 18 – *Boxplot* do consumo *per capita* de água em relação às regiões de Joinville

4.2.4 Análise exploratória

Esta seção apresenta a caracterização da variável resposta consumo *per capita* de água mediante a construção de *boxplots*, histogramas e tabela descritiva dos dados. Ademais, foi verificada a suposição de normalidade dos dados.

Para obter o indicador de consumo, o consumo bruto de água de cada mês foi dividido pelo número de pessoas que moravam na casa e pelo número de dias do respectivo mês. O consumo *per capita* de água (litros por pessoa por dia) de cada residência foi calculado pela média de 24 meses de medição, no período de dezembro de 2017 a novembro de 2019. A Tabela 9 apresenta a estatística descritiva dessa variável dependente.

Tabela 9 – Estatística descritiva da variável consumo *per capita*

Variável	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Desvio padrão
Consumo <i>per capita</i> (litros/pessoa/dia)	40,83	105,10	134,02	146,40	172,30	487,92	64,598

Percebe-se pela estatística descritiva que a média do consumo *per capita* apresenta indícios de ser superior à mediana, possivelmente influenciada por valores

altos de consumo. Os resultados revelaram uma alta variabilidade dos dados em torno da média em virtude do percentual do coeficiente de variação, 44,12%.

O consumo *per capita* de água variou de 40,83 litros por pessoa por dia a 487,92 litros por pessoa por dia, com média de 146,40 litros por pessoa por dia. O valor mínimo de 40,83 litros por pessoa por dia justifica-se por se tratar de uma residência cujo consumo de água é complementado pela captação de água subterrânea (poço). De acordo com o Sistema Nacional de Informações sobre Saneamento – SNIS (2018), em 2018 Joinville registrou um consumo *per capita* doméstico de água potável de 133,39 litros por pessoa por dia. Outros estudos envolvendo edificações residenciais unifamiliares brasileiras obtiveram indicadores de consumo similares. Em Florianópolis, Ghisi e Oliveira (2007) investigaram duas residências e encontraram valores de 163,33 e 150,00 litros por pessoa por dia, respectivamente. Em Brasília, um estudo envolvendo 118 domicílios mostrou que as famílias com renda alta consumiram em média 321 litros por pessoa por dia. O consumo médio dos domicílios em que os moradores possuíam renda média a alta foi 205 litros por pessoa por dia; as casas cujos moradores possuíam renda média a baixa apresentaram consumo de, em média 146 litros por pessoa por dia e o consumo *per capita* das famílias com renda baixa foi 112 litros por pessoa por dia (SANT'ANA; MAZZEGA, 2018). Em Salvador, 58 residências de baixa renda tiveram consumo médio *per capita* e desvio padrão de 121,5 e 58,5 litros por pessoa por dia, respectivamente (GARCIA, 2011).

Pelo gráfico de *boxplot*, percebe-se a presença de possíveis *outliers*, representados por pontos pretos na Figura 19. Foi utilizado o método univariado de Tukey (1977) para detecção desses valores atípicos, calculando a distância interquartis e os limites superior (LS=273,10) e inferior (LI=4,30). Como resultado, constatou-se que a amostra possui 20 observações discrepantes, com valores acima do limite superior LS.

Além disso, a Figura 19 mostra uma assimetria no conjunto de dados visto que a mediana está mais próxima do quartil inferior. Embora com uma pequena diferença, a média (146,40) possui um valor superior à mediana (134,02), reforçando assim uma distribuição assimétrica. A análise gráfica do histograma dessa variável (Figura 20) confirma uma distribuição assimétrica positiva já que aproximadamente 85% dos dados possuem valor inferior a 200 litros por pessoa por dia, ou seja, a maioria dos dados estão concentrados à esquerda com uma cauda longa à direita.

Figura 19 – Boxplot da variável dependente consumo *per capita* (litros/pessoa/dia)

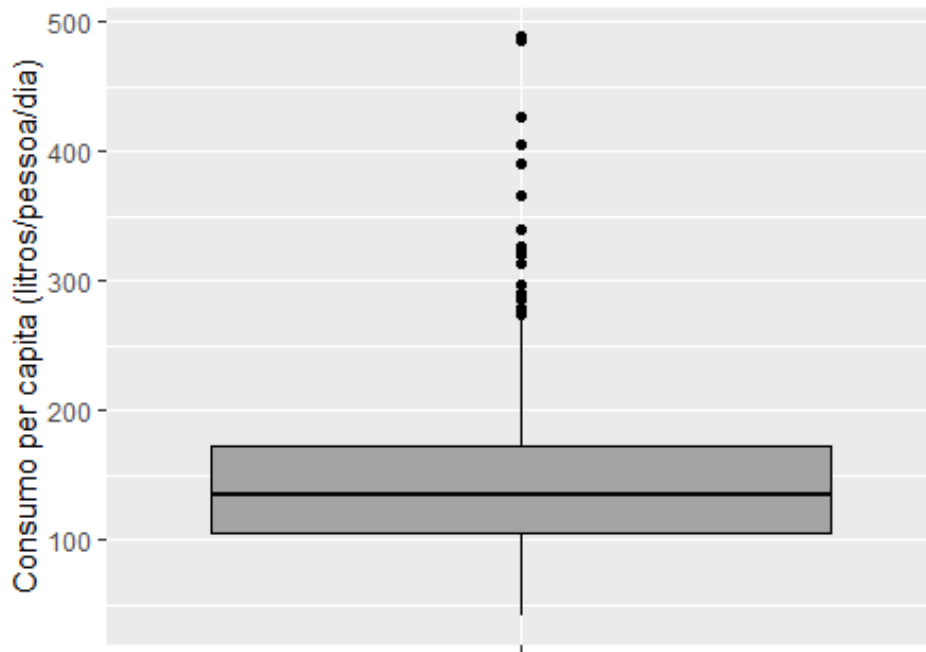
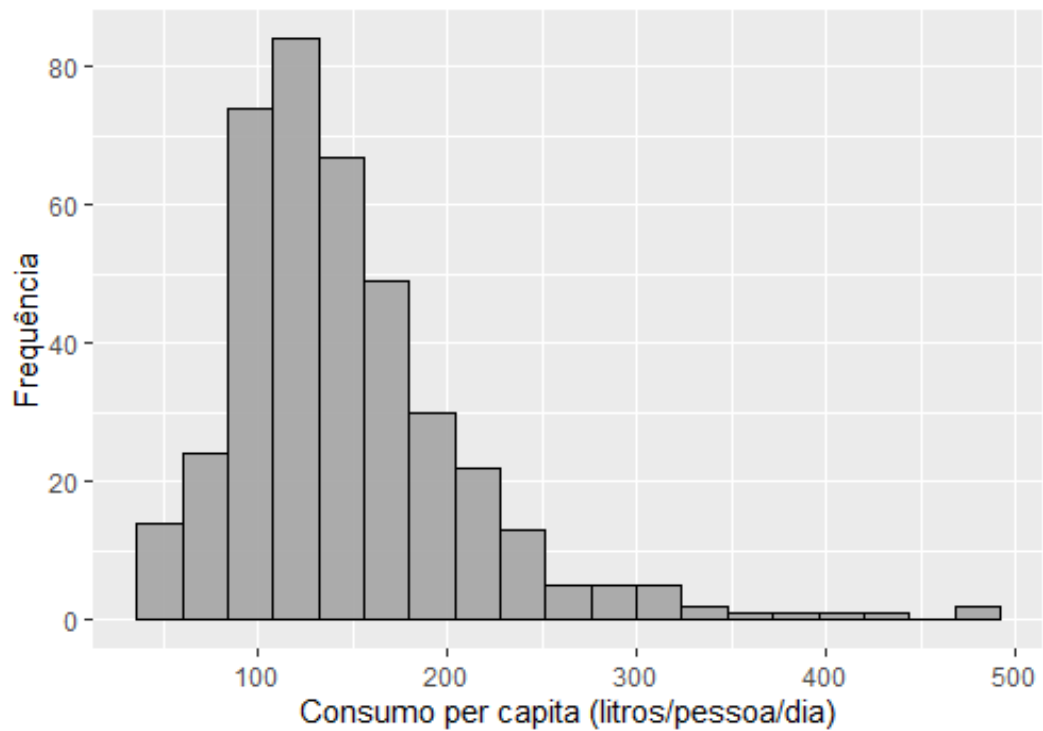


Figura 20 – Histograma da variável dependente consumo *per capita* (litros/pessoa/dia)

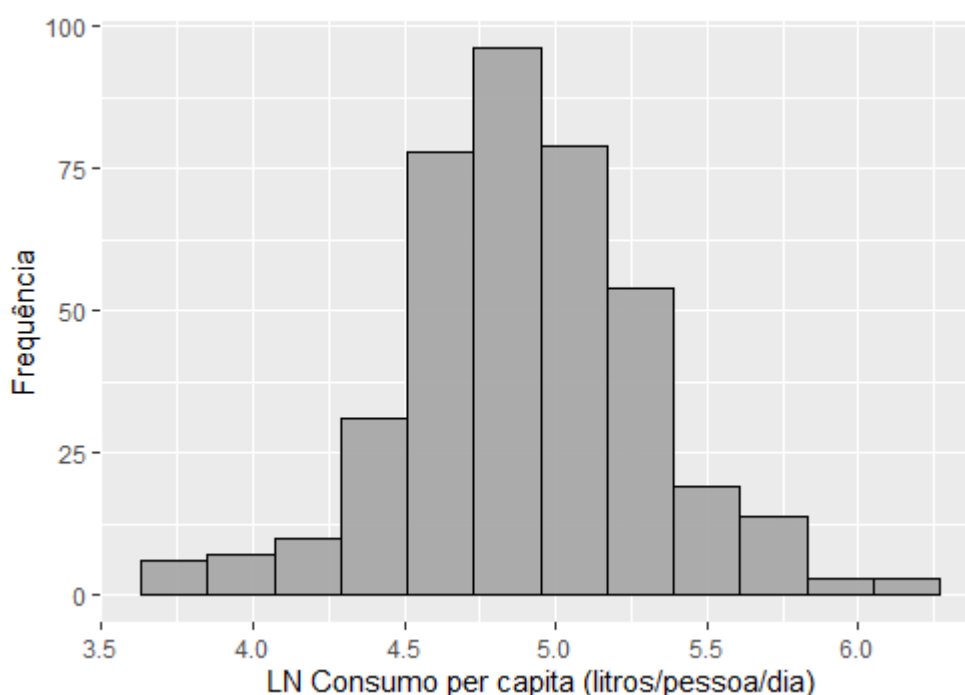


Com o intuito de averiguar as suposições de normalidade dos dados de consumo *per capita*, foi aplicado o teste de Lilliefors. A probabilidade de significância obtida foi $p\text{-valor}=4,07 \times 10^{-15}$, demonstrando que a distribuição normal não era adequada para representar a amostra.

Por esse motivo, foi empregada a transformação logarítmica nos dados de consumo *per capita*. Dias, Kalbusch e Henning (2018), Cruz *et al.* (2017) e Tortella e Tirado (2011) utilizaram o mesmo critério de transformação logarítmica da variável dependente consumo de água.

Como resultado, a probabilidade de significância foi maior que 0,05 (p-valor=0,15052). Sendo assim, a hipótese de normalidade da variável resposta foi atendida. O histograma (Figura 21) evidencia que, após a aplicação da transformação logarítmica, o comportamento da série de dados se aproximou de uma distribuição normal, sem apresentar aquela cauda à direita. Dessa forma, a transformação logarítmica ajudou a converter a distribuição assimétrica dos dados em uma distribuição simétrica (TUFFÉRY, 2011).

Figura 21 – Histograma da variável LN consumo *per capita*



Na Tabela 10, é possível observar que após a aplicação da transformação LN, os valores de média e mediana do consumo estão próximos. Isso sugere que a transformação logarítmica gera uma medida de posição central mais adequada, com distribuição tendendo à normalidade e dispersão razoavelmente simétrica dos dados.

Tabela 10 – Estatística descritiva da variável LN consumo *per capita*

Variável	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Desvio padrão
LN Consumo <i>per capita</i> (litros/pessoa/dia)	3,709	4,655	4,898	4,903	5,149	6,190	0,406

Além da transformação dos dados, outra etapa importante e realizada em termos exploratórios na pesquisa é a análise bivariada para verificar o grau de relacionamento entre duas variáveis.

4.3 FATORES RELEVANTES NO CONSUMO PER CAPITA DE ÁGUA

Esta seção apresenta os fatores relevantes no consumo *per capita* de água em residências unifamiliares de Joinville com base nos resultados da análise de correlação e modelos de regressão linear múltipla e regressão robusta.

4.3.1 Correlação entre variáveis

O processo de correlação objetivou identificar possíveis relações entre a variável resposta LN consumo *per capita* e as 57 variáveis independentes obtidas. As análises bivariadas foram realizadas por meio do cálculo do coeficiente de correlação de Kendall e teste de significância das variáveis. Considerando que a variável resposta segue a distribuição normal, optou-se pelo coeficiente tau de Kendall em vez do coeficiente rho de Spearman por ser mais robusto na presença de *outliers*, proporcionando maior capacidade de inferência populacional e menor erro na estimativa (CROUX; DEHON, 2010; AKOGLU, 2018; MIOT, 2018).

A Tabela 11 mostra as variáveis estatisticamente significativas aos níveis de 5% e 10% ($p\text{-valor} < 0,05$ e $p\text{-valor} < 0,10$), bem como seus coeficientes de correlação. Assim como no estudo de Cruz *et al.* (2017), embora a maioria dos coeficientes tivessem valores baixos, com módulos inferiores a 0,3 na escala de 0 a 1, as correlações foram estatisticamente significativas. Os resultados indicaram que 14 variáveis estavam associadas ao consumo *per capita* de água.

Tabela 11 – Variáveis independentes correlacionadas com o Consumo *Per Capita*

N	Nome da variável	Descrição	Coefficiente tau
1	COLETA_ESGOTO	Existência de rede coletora de esgoto na via	0,079 ^b
2	NUM_MORADORES	Número total de moradores	-0,175 ^a
3	NUM_MULHERES	Número de mulheres	-0,094 ^a
4	NUM_CRIANCAS	Número de crianças (idade inferior a 12 anos)	-0,117 ^a
5	NUM_ADULTOS	Número de adultos (idade entre 18 e 60 anos)	-0,089 ^a
6	PROPRIETARIO	Propriedade da residência (alugada ou financiada)	-0,106 ^a
7	NUM_BANHEIROS	Número de banheiros	0,088 ^a
8	IDADE_IMOVEL	Idade do imóvel (anos)	0,063 ^a
9	AREA_IMOVEL	Área construída (m ²) - inclui área da casa, garagem, edícula	0,122 ^a
10	AREA_TERRENO	Área do terreno (m ²)	0,105 ^a
11	PISCINA	Presença de piscina	0,147 ^a
12	BANHEIRA	Presença de banheira	0,097 ^a
13	BALDE_CALCADA	Uso de balde para lavar a calçada e áreas externas	-0,119 ^a
14	VAP_CALCADA	Uso de máquina de alta pressão para lavar a calçada e áreas externas	0,071 ^b

^a – A correlação é significativa ao nível de 0,05

^b – A correlação é significativa ao nível de 0,10

A partir desses resultados, foi possível selecionar as variáveis com maiores coeficientes de correlação, em módulo, e significativas estatisticamente para então ajustar modelos finais de regressão.

4.3.2 Modelos de regressão linear múltipla e robusta

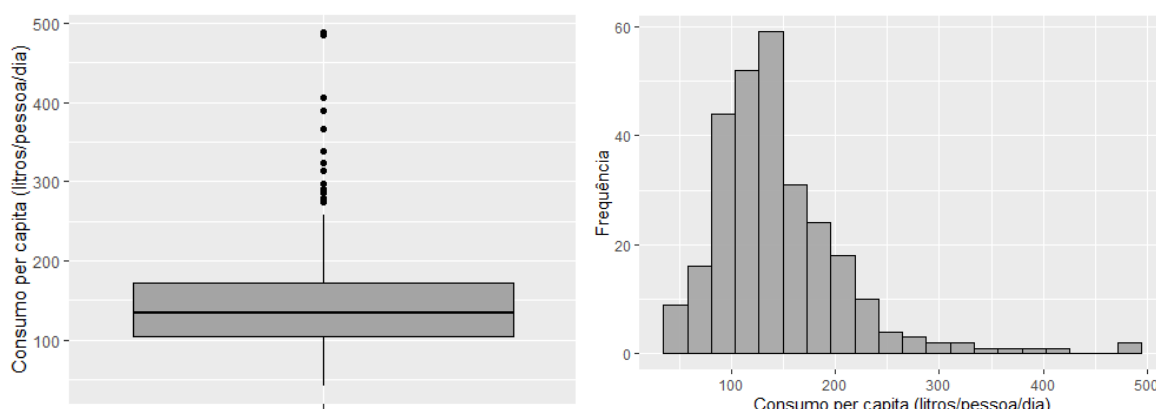
Neste subcapítulo são apresentados os resultados da análise de regressão linear múltipla e regressão robusta. O conjunto de treinamento possui 280 residências selecionadas de forma aleatória no processo de partição dos dados. Quanto às características da variável resposta consumo *per capita*, o valor mínimo observado foi 40,83 litros por pessoa por dia e o valor máximo foi 487,92 litros por pessoa por dia, com média de 145,55 litros por pessoa por dia (Tabela 12). O *boxplot* e histograma da variável (Figura 22) confirmam que esse conjunto de dados apresenta características

semelhantes à amostra original, a presença de *outliers* e uma distribuição assimétrica à direita.

Tabela 12 – Estatística descritiva da variável consumo per capita do conjunto de treinamento

Variável	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Desvio padrão
Consumo <i>per capita</i> (litros/pessoa/dia)	40,83	105,00	134,02	145,55	172,11	487,92	65,603

Figura 22 – *Boxplot* e histograma da variável consumo *per capita* do conjunto de treinamento



Com base no conjunto de treinamento, foi elaborado um modelo de regressão linear múltipla, considerando como variável resposta o consumo *per capita* de água (litros por pessoa por dia). Por meio da análise de correlação, das 57 variáveis iniciais, foram selecionadas para compor o modelo as 14 variáveis significativas aos níveis de 5% e 10%.

Considerando 14 variáveis na elaboração do modelo, o tamanho da amostra treino (280 residências) atende ao pressuposto da amostragem, de ter no mínimo 6 a 10 vezes a quantidade de variáveis candidatas a preditores. Portanto, garante-se que o tamanho do conjunto de dados é satisfatório para construir um modelo confiável (KUTNER; NACHTSHEIM; NETER, 2004).

As variáveis independentes selecionadas para a análise de regressão abrangiam: existência de rede coletora de esgoto; número total de moradores; número de mulheres; número de crianças (idade inferior a 12 anos); número de adultos (idade entre 18 e 60 anos); propriedade da residência (alugada ou financiada); número de banheiros; logaritmo da idade do imóvel (anos); logaritmo da área do terreno (m²); logaritmo da área construída (m²) – que inclui área da casa, garagem e

edícula; presença de piscina; presença de banheira; uso de balde e uso de máquina de alta pressão para lavar a calçada e áreas externas.

Por envolver muitas variáveis regressoras, a escolha do melhor modelo foi realizada mediante a aplicação da regressão *stepwise*. Os resultados da análise de regressão linear múltipla são mostrados na Tabela 13. Quatro variáveis apresentaram $p\text{-valor} < 0,05$, ou seja, seus coeficientes foram significativos: número de moradores, residência alugada ou financiada, área do imóvel e presença de piscina. A variável uso de balde para lavagem de calçada e áreas externas foi considerada significativa ao nível de 10%.

Tabela 13 – Estatísticas e coeficientes do modelo de regressão linear múltipla

	Estimativa	Erro padrão	t	p-valor	VIF
Intercepto	4,42309	0,24366	18,153	$< 2,20 \times 10^{-16}$ ***	-
Número total de moradores	-0,10266	0,01959	-5,241	$3,19 \times 10^{-7}$ ***	1,04845
Propriedade da residência (alugada ou financiada)	-0,14777	0,05252	-2,813	0,00526 **	1,04080
LN área construída (m ²)	0,17532	0,05006	3,502	0,00054 ***	1,16396
Presença de piscina	0,14735	0,06156	2,394	0,01736 *	1,08075
Uso de balde para lavagem de áreas externas	-0,10102	0,05185	-1,948	0,05238 .	1,00498

Nível de significância 0 (****) 0,001 (***) 0,01 (**) 0,05 (*)

Erro padrão residual = 0,3765

F = 12,26 p-valor = $9,744 \times 10^{-11}$

R² = 0,1828 R² ajustado = 0,1679.

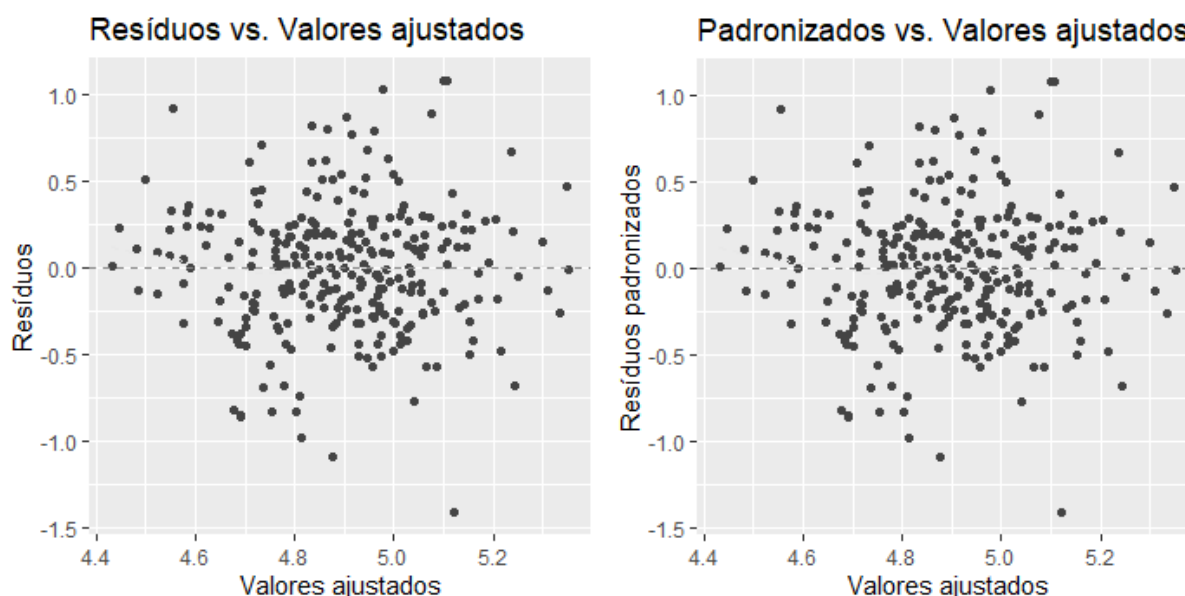
O modelo é estatisticamente significativo ($F = 12,26$, $p\text{-valor} < 0,05$) e explica 18,28% ($Adjusted R^2 = 0,1679$) da variação no consumo *per capita* de água das residências analisadas. Como mostra a Tabela 11, o resultado de *VIF – Variance Inflation Factor* para as variáveis independentes mostrou-se adequado uma vez que os valores foram inferiores a 5. Sendo assim, não há indícios de multicolinearidade no modelo.

As suposições de linearidade, homocedasticidade e normalidade dos resíduos, além da presença de *outliers*, foram verificadas pela avaliação de gráficos e aplicação de testes estatísticos (DEVORE, 2016), os quais são apresentados a seguir.

A Figura 23 mostra que os resíduos apresentam distribuição aparentemente aleatória, isto é, formam uma nuvem de pontos e não há qualquer tendência ou padrão reconhecível no gráfico. Os resultados do teste Breusch-Pagan levam à não rejeição

da hipótese nula ($p\text{-valor}=0,11993>0,05$). Dessa forma, a variância dos resíduos pode ser considerada constante.

Figura 23 – Resíduos *versus* valores ajustados da regressão linear múltipla



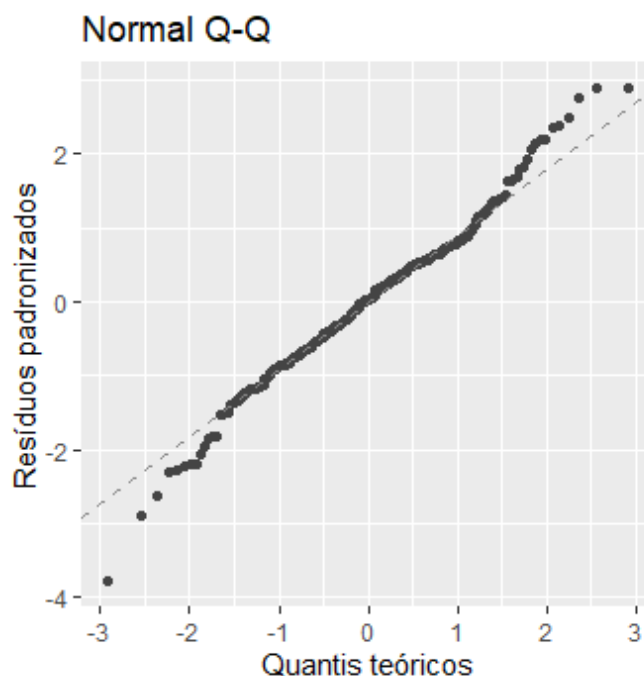
A aleatoriedade e dispersão dos resíduos permite inferir que a suposição de linearidade do modelo é verdadeira. Do contrário, a Figura 23 apresentaria um padrão característico, revelando possíveis problemas de não-linearidade. A probabilidade de significância do teste Rainbow ($p\text{-valor}=0,27996>0,05$) confirmou que a função linear era apropriada para o modelo em questão. Os resultados mostraram que os resíduos são independentes e não apresentam autocorrelação visto que $p\text{-valor}=0,48546>0,05$ no teste de Durbin-Watson.

Como ilustra a Figura 24, os pontos nas caudas do gráfico de probabilidade normal desviam da linha de tendência diagonal, sugerindo a não normalidade dos resíduos. A aplicação do teste Lilliefors comprovou que os resíduos do modelo de regressão linear múltipla não atendem ao requisito de normalidade já que $p\text{-valor} = 0,02619$, menor que $\alpha (0,05)$.

Conforme Schmidt e Finan (2018), em amostras de tamanho grande (número de observações por variável > 10), os modelos de regressão linear são robustos à violação da suposição de normalidade. Nessas situações, os pesquisadores devem se concentrar na detecção de especificações errôneas do modelo como valores extremos, *high leverage points*, heterocedasticidade, autocorrelação dos resíduos,

não linearidade e possíveis interações, que podem influenciar os resultados, independentemente do tamanho da amostra.

Figura 24 – Probabilidade normal dos resíduos da regressão linear múltipla



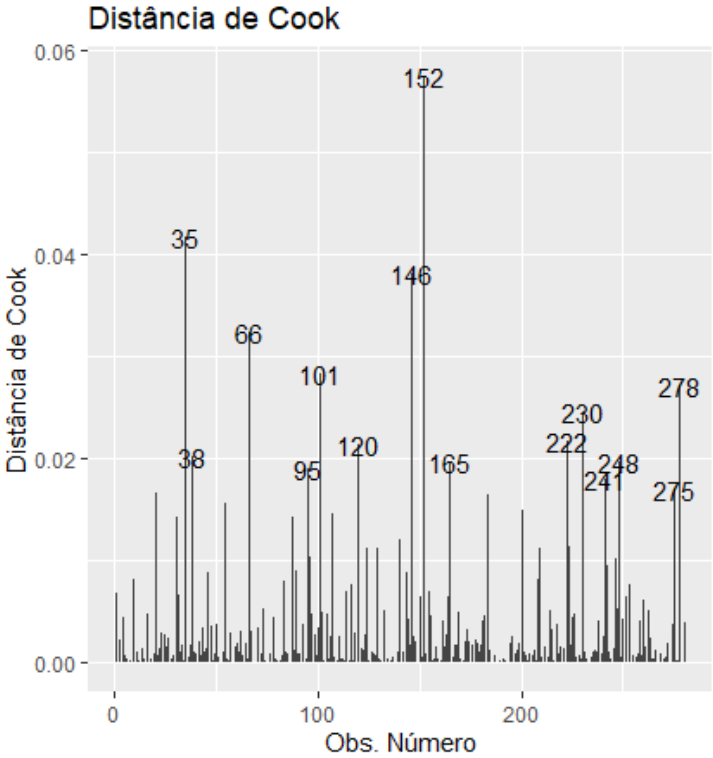
Por fim, investigou-se a presença de observações influentes e possíveis *outliers* pela distância de Cook bem como pela aplicação do teste de Bonferroni (KUTNER; NACHSHEIM; NETER, 2004). Observa-se na Figura 25 que a linha 152 se destaca por apresentar distância de Cook bem superior às demais. Entretanto não existem observações com distância de Cook > 1 , logo não há pontos na amostra que possam alterar significativamente a função de regressão ajustada (KUTNER; NACHSHEIM; NETER, 2004).

Baseado nos resultados do teste de Bonferroni ($p\text{-valor}=0,03476 < 0,05$) concluiu-se que a observação 152 possui o maior valor de resíduo e representa um *outlier* no modelo de regressão linear múltipla.

Conforme Zhang *et al.* (2018), nessas situações ignorar ou deletar observações influentes pode levar a grandes desvios nas estimativas. Sendo assim, a utilização de métodos robustos permite reduzir o peso de tais observações em vez de descartá-las, tornando-se uma alternativa eficiente (FARCOMENI; VENTURA, 2012). Portanto, adotou-se a regressão robusta por fornecer previsões menos sensíveis a *outliers* e pela não aderência dos resíduos à distribuição normal (TUFFÉRY, 2011). Os

parâmetros da regressão robusta foram estimados pela classe MM-Estimadores (SUSANTI *et al.*, 2014).

Figura 25 – Distância de Cook dos resíduos da regressão linear múltipla



A Tabela 14 exibe as estatísticas e coeficientes do modelo de regressão robusta. As variáveis número de moradores, área do imóvel, presença de piscina e uso de balde para lavagem de calçada e áreas externas foram significativas ao nível de 5%, isto é, apresentaram p-valor menor do que 0,05. A variável propriedade da residência (alugada ou financiada) mostrou-se significativa ao nível de 10%.

Tabela 14 – Estatísticas e coeficientes do modelo de regressão robusta

	Estimativa	Erro padrão	t	p-valor	VIF
Intercepto	4,24897	0,23034	18,447	$< 2,20 \times 10^{-16}$ ***	-
Número total de moradores	-0,08878	0,01832	-4,846	$2,11 \times 10^{-6}$ ***	1,05476
Propriedade da residência (alugada ou financiada)	-0,10584	0,05668	-1,867	0,06292 .	1,13163
LN área construída (m²)	0,19869	0,04513	4,403	$1,53 \times 10^{-5}$ ***	1,26194
Presença de piscina	0,16552	0,05785	2,862	0,00454 **	1,12034
Uso de balde para lavagem de áreas externas	-0,10886	0,04760	-2,287	0,02296 *	1,07633

Nível de significância 0 '***' 0,001 '**' 0,01 '*' 0,05 '.'
Erro padrão residual robusto = 0,32931
 $R^2 = 0,20166$ R^2 ajustado = 0,18710

Na elaboração do modelo, o peso das 280 residências seguiu a estatística da Tabela 15. No total, 18 observações tiveram peso aproximadamente igual a 1. As 262 observações restantes foram ponderadas conforme o seu resíduo, isto é, quanto maior o resíduo em módulo, menor a influência da observação na estimativa dos parâmetros da regressão.

Tabela 15 – Estatística descritiva do peso das observações no modelo de regressão robusta

	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Desvio padrão
Peso das observações	0,01639	0,89215	0,95949	0,89708	0,98839	0,99999	0,15998

As observações com menor peso na regressão robusta foram aquelas cujo consumo *per capita* representa os valores extremos, mínimo ou máximo, da amostra de treinamento (Tabela 16).

Tabela 16 – Observações com menor peso no modelo de regressão robusta

Observação	Consumo <i>per capita</i> (l/pessoa/dia)	Resíduo	Peso
152	40,83	-1,44070	0,01639
66	487,92	1,13756	0,20827
241	485,37	1,11465	0,22852
209	44,30	-1,08114	0,25903
38	405,23	1,07361	0,26601

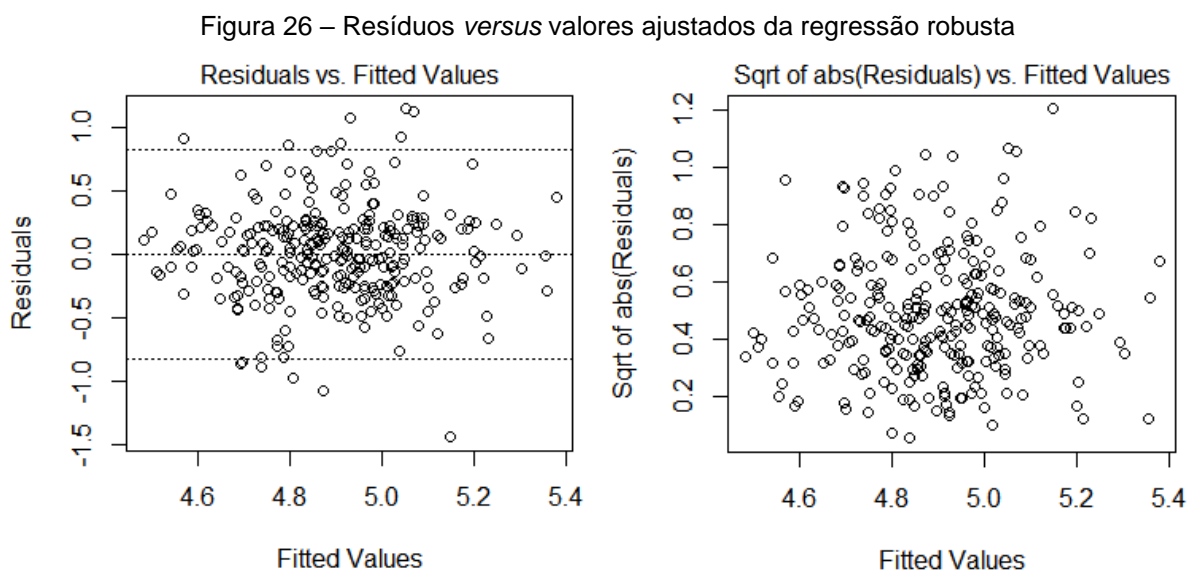
Percebe-se que a observação 152 foi a que teve o maior resíduo e, conseqüentemente, menor peso. Novamente, aparece como o ponto mais discrepante da amostra. Trata-se da casa com o menor consumo *per capita*, 40,83 litros por pessoa por dia, valor justificado pelo uso de outra fonte de abastecimento de água, o poço. A observação 209 também possui captação de água subterrânea, razão pela qual o consumo *per capita* foi 44,30 litros por pessoa por dia. Além disso, essas residências estão situadas nas regiões Sul e Sudeste de Joinville e possuem idade de 7 e 13 anos, respectivamente. Em contrapartida, as observações com maior consumo *per capita* de água (66 e 241) localizam-se na região centro-norte da cidade e são construções mais antigas, com 17 e 27 anos de idade.

O modelo de regressão robusta explica 20,17% ($Adjusted R^2 = 0,1871$) da variação no consumo *per capita* de água das residências analisadas, o que representa uma sutil melhora no poder de explicação em relação ao modelo de regressão linear múltipla.

O valor de *VIF* – *Variance Inflation Factor* para as variáveis independentes integrantes do modelo foi inferior a 5 (Tabela 11), indicando ausência de multicolinearidade (CONNOLLY; HAGELMAN III, 2015; FITZGERALD; AUERBACH, 2016; CRUZ *et al.*, 2017).

Segundo Kutner, Nachtsheim e Neter (2004), em um modelo de regressão adequado, os resíduos não exibem tendências sistemáticas positivas nem negativas e seus valores ficam dispersos em torno de zero. Na Figura 26, observa-se que os resíduos apresentam distribuição aparentemente aleatória, isto é, formam uma nuvem de pontos e não há qualquer tendência ou padrão reconhecível no gráfico. A aplicação do teste Breusch-Pagan resultou em $p\text{-valor} = 0,11993 > 0,05$, levando a não rejeição da hipótese nula de homocedasticidade dos resíduos. Portanto, a variância dos resíduos pode ser considerada constante.

Mediante análise gráfica, constatou-se também que os resíduos possuem, em módulo, valor menor do que os resíduos do modelo anterior. Esse resultado revela que a regressão robusta foi eficaz ao diminuir a magnitude dos resíduos.



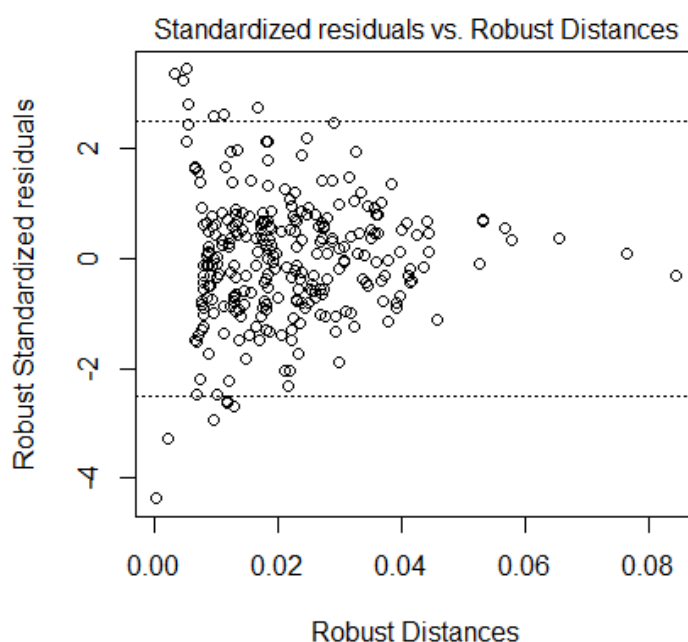
Como a Figura 26 não apresenta um padrão característico e os resíduos estão aleatoriamente dispersos, supõe-se que a hipótese de linearidade do modelo seja

verdadeira. O resultado do teste Rainbow ratifica o adequado emprego de função linear no modelo de regressão em análise ($p\text{-valor}=0,27996>0,05$). Quanto à suposição de independência dos resíduos, a probabilidade de significância resultante do teste de Durbin-Watson foi $p\text{-valor}=0,48546$, superior a 0,05, indicando que os resíduos são independentes e não apresentam autocorrelação.

Por fim, investigou-se a presença de observações influentes e possíveis *outliers* pelo gráfico de resíduos padronizados *versus* distâncias robustas (RD), introduzido por Rousseeuw e van Zomeren (1990). Os autores classificam os dados em quatro categorias: as observações regulares, com pequenas distâncias e pequenos resíduos; os *outliers* verticais, com pequenas distâncias e grandes resíduos; os *good leverage points*, com grandes distâncias e pequenos resíduos; e por fim os *bad leverage points*, com grandes distâncias e grandes resíduos.

Na Figura 27, observa-se a presença de pontos fora da linha de tolerância horizontal $[-2,5, 2,5]$ e com baixo valor de distância robusta, dessa forma são considerados *outliers* na regressão. Como não há observações com distância robusta superior a 0,975, conclui-se que não existem *leverage points* na regressão (ROUSSEEUW; VAN ZOMEREN, 1990).

Figura 27 – Probabilidade normal dos resíduos da regressão robusta



Dessa forma, a regressão robusta auxiliou a reforçar que os *outliers*, sinalizados na análise de resíduos da regressão múltipla, não são pontos influentes.

Na realidade são apenas *outliers* verticais, os quais aumentam o valor do erro associado à estimativa, porém, ao contrário dos pontos influentes, não modificam a inclinação da reta e equação de regressão.

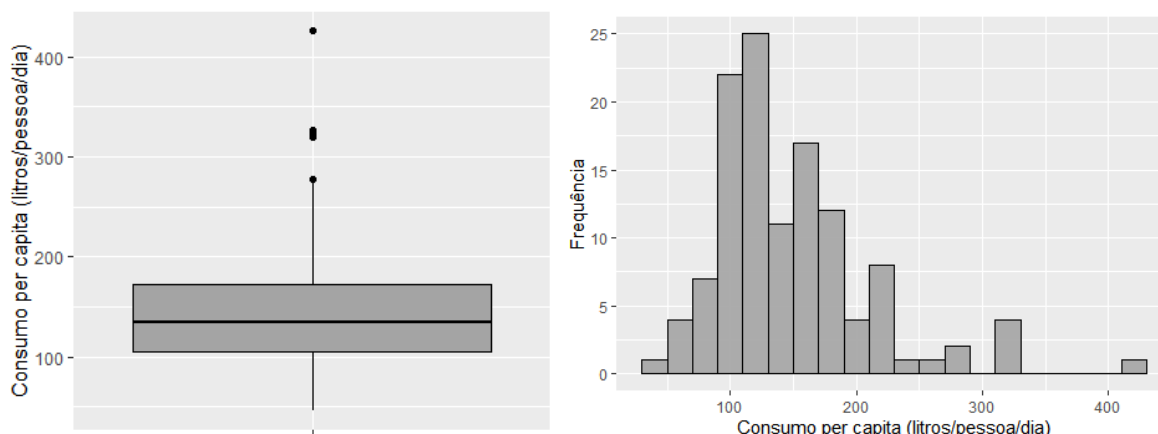
Durante a modelagem das equações de regressão, foi necessário avaliar a qualidade dos modelos, isto é, qual a sua precisão na previsão da resposta para novas observações. Sendo assim, a técnica empregada para checar os modelos de regressão e sua capacidade preditiva foi o uso de *holdout sample*. Da amostra inicial de 400 residências, 120 casas foram selecionadas de forma aleatória para constituir a amostra de validação.

Conforme a Tabela 17, o consumo *per capita* de água das residências variou de 45,65 litros por pessoa por dia a 425,92 litros por pessoa por dia, com média de 148,37 litros por pessoa por dia. O *boxplot* e histograma da variável (Figura 28) confirmam que esse conjunto de dados apresenta características semelhantes à amostra original, a presença de *outliers* e uma distribuição assimétrica com cauda à direita.

Tabela 17 – Estatística descritiva da variável consumo *per capita* do conjunto de validação

Variável	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Desvio padrão
Consumo <i>per capita</i> (litros/pessoa/dia)	45,65	105,10	134,11	148,37	172,85	425,92	62,414

Figura 28 – *Boxplot* e histograma da variável consumo *per capita* do conjunto de validação



A amostra de validação foi utilizada para testar a capacidade preditiva dos modelos, ou seja, comparar o consumo real e predito. Os resultados são mostrados no Apêndice C e na Tabela 18.

Tabela 18 – Indicadores da validação dos modelos de regressão linear múltipla e regressão robusta

Método	MAE	MAPE	RMSE	R² predição
Regressão linear múltipla	42,23846	0,29174	58,19039	0,17290
Regressão robusta	42,13574	0,29181	57,99137	0,17863

O Erro Absoluto Percentual representa a diferença entre o consumo *per capita* lido pela Companhia Águas de Joinville e o previsto pelos modelos. Os modelos de regressão linear múltipla e regressão robusta apresentaram erro absoluto médio de 29,17% e 29,18%, respectivamente. Outra métrica importante, o erro absoluto mediano foi de 23,42% para o modelo de regressão linear múltipla e 23,98% para o modelo de regressão robusta. Além disso, na regressão linear múltipla, o real consumo *per capita* de 4,17% das observações ficou fora do intervalo de predição enquanto no caso da regressão robusta, esse percentual subiu para 6,67%. Esses resultados sugerem que o modelo de regressão linear múltipla foi mais assertivo na previsão do consumo *per capita* do que o modelo de regressão robusta. Percebeu-se também que os dois modelos não foram assertivos justamente nas casas com consumos extremos, principalmente nas casas com consumo inferior a 60 litros por pessoa por dia. Uma sugestão para estudos futuros é a utilização de regressão quantílica para encontrar equações específicas de acordo com as faixas de consumo de água, minimizando dessa forma o erro associado à estimativa.

O estudo de Dias, Kalbusch e Henning (2018) envolveu 8 edifícios na validação do modelo de consumo *per capita* e obteve um erro absoluto médio de 18,14%. Cruz *et al.* (2017) utilizaram quatro residências no processo de validação e a aproximação variou de 78,0% a 90,4%. Os resultados da presente pesquisa foram um pouco superiores aos descritos acima, possivelmente por envolver 120 residências e pela presença de observações discrepantes. Mesmo assim, esses valores indicam uma boa aproximação com o real consumo das residências unifamiliares de Joinville e, consequentemente, capacidade satisfatória de predição dos modelos.

Os intervalos de predição (95%) calculados para cada observação estão no Apêndice C. Para o modelo de regressão linear múltipla, o valor de R² (predição) foi 0,1729 (17,29%). Portanto, espera-se que o modelo explique 17,29% da variabilidade na previsão de novas observações. Como esse valor é similar ao R² do modelo original (0,1828), pode-se concluir que o modelo possui uma adequada habilidade de predição. Em relação ao modelo de regressão robusta, os coeficientes de

determinação também foram semelhantes. O R^2 de predição obtido foi 0,1786 (17,86%), similar ao R^2 (0,2017).

Apesar dos coeficientes de determinação (R^2) terem sido baixos, com valores inferiores a 0,25 numa escala de 0 a 1, eles são aceitáveis para o tipo de fenômeno que está sendo investigado. Esse poder explicativo dos modelos é consistente com estudo similares envolvendo consumo de água. Cruz *et al.* (2017) escolheram um modelo que explicava 31,6% da variabilidade do consumo doméstico de água em Hermosillo, no México. Conforme os autores, o valor obtido ($R^2 = 0,316$) é aceitável para este tipo de estudo (CRUZ *et al.*, 2017). Em Jaffna, um distrito do Sri Lanka, Rajeevan e Mishra (2020), obtiveram um modelo estatisticamente significativo com coeficiente de determinação $R^2 = 0,304$. Kontokosta e Jain (2015) determinaram modelos de regressão que explicavam aproximadamente 23% ($R^2 = 0,234$) da variação no uso de água em edifícios de Nova York enquanto Chang *et al.* (2017) ajustaram um modelo de regressão com coeficiente de determinação $R^2 = 0,22$ para explicar o consumo mensal de água em residências unifamiliares da cidade de Phoenix, nos Estados Unidos da América.

Após o ajuste e validação dos modelos, é importante verificar a relevância das variáveis explicativas no contexto do fenômeno que está sendo analisado, o consumo *per capita* de água.

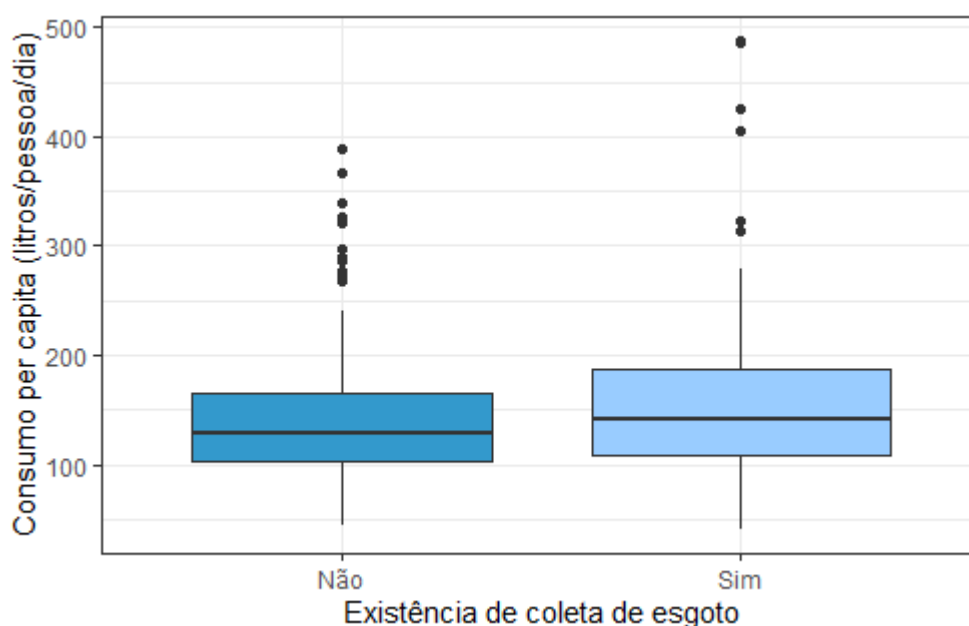
4.3.3 Discussão dos fatores

De uma lista de 57 variáveis independentes, foram identificados os fatores mais relevantes no consumo de água por meio de análise de correlação, regressão linear múltipla e regressão robusta. As variáveis abrangiam a infraestrutura do entorno, aspectos socioeconômicos e demográficos, características construtivas da edificação, além dos hábitos de consumo dos usuários.

Dessa forma, o consumo *per capita* das residências analisadas é afetado pela existência de rede coletora de esgoto; número total de moradores; número de mulheres; número de crianças (idade inferior a 12 anos); número de adultos (idade entre 18 e 60 anos); propriedade da residência (alugada ou financiada); número de banheiros; idade do imóvel (anos); área do terreno (m^2); área construída (m^2) – que inclui área da casa, garagem e edícula; presença de piscina; uso de balde e uso de máquina de alta pressão para lavar a calçada e áreas externas.

No estudo de correlação, a existência de rede coletora de esgoto na via apresentou associação numérica positiva com o consumo *per capita*, indicando que o consumo de água aumenta com a presença de sistema de coleta de esgoto (Figura 29). Em Joinville, é cobrada uma tarifa de 80% sobre o valor do consumo de água para custear a coleta de esgoto, portanto nesses casos a conta fica mais cara. Tradicionalmente, a tarifa aplicada tem um efeito negativo sobre o uso de água, ou seja, tende a diminuí-lo (ROMANO; SALVATI; GUERRINI, 2014). Portanto, era esperada uma redução no consumo das residências onde há coleta de esgoto. A razão para o aumento no consumo de água pode estar relacionada ao poder aquisitivo da população servida, como ocorreu no estudo de Dias, Kalbusch e Henning (2018). O teste qui-quadrado comprovou a relação entre existência de rede coletora de esgoto na via e renda familiar ($\chi^2=31,845$, $p=1,75 \times 10^{-5}$), isto é, o sistema de coleta de esgoto está presente nos bairros com renda mais alta, em que o consumo de água é mais elevado.

Figura 29 – Relação entre existência de coleta de esgoto e consumo *per capita* de água



A análise de correlação mostrou que as variáveis demográficas número de mulheres e número de crianças são significativas. Dessa forma, o aumento no número de mulheres ou crianças pode indicar uma redução no consumo *per capita* de água. Hussien, Memon e Savic (2016) observaram que o consumo *per capita* diminui com o

aumento de crianças, porém é maior com o aumento da quantidade de mulheres nas famílias.

Estudos anteriores demonstraram a influência de características socioeconômicas como renda e grau de escolaridade no consumo de água (SCHLEICH; HILLENBRAND, 2009; BEAL; STEWART; FIELDING, 2013; MATOS *et al.*, 2014; GARCIA-CUERVA; BERGLUND; BINDER, 2016; FAN *et al.* 2017; SANT'ANA; MAZZEGA, 2018; RAJEEVAN; MISHRA, 2020). Todavia, esses fatores não foram estatisticamente significativos no presente estudo. Uma explicação comprovada pelo coeficiente de correlação tau de Kendall é o fato de que a renda ($\tau=0,28505$, $p=3,98 \times 10^{-15}$) e o grau de escolaridade ($\tau=0,10151$, $p=0,00674$) dos moradores das residências analisadas estão correlacionados com o tamanho da edificação, tal como na pesquisa de Chang, Parandvash e Shandas (2010). A quantidade de banheiros da residência também é afetada pela renda domiciliar ($\tau=0,51513$, $p=2,22 \times 10^{-16}$) e grau de escolaridade ($\tau=0,31051$, $p=1,35 \times 10^{-13}$). Os resultados indicaram que a presença de piscina também exibiu correlação com a renda ($\tau=0,28789$, $p=8,61 \times 10^{-11}$) e grau de escolaridade dos moradores ($\tau=0,18603$, $p=4,88 \times 10^{-5}$). Logo, pode-se inferir que as variáveis socioeconômicas afetam diretamente as variáveis construtivas e, conseqüentemente, influenciam de forma indireta o consumo de água.

O consumo *per capita* exibiu correlação positiva com o número de banheiros (Figura 30). Ou seja, quanto maior a quantidade de banheiros da casa, maior foi o consumo de água. Esse resultado foi possivelmente influenciado pela renda familiar dos moradores visto que o número de banheiros apresentou-se correlacionado à renda. Em Madri, González, Rueda e Les (2008) obtiveram resultado semelhante, que o consumo era influenciado pelo número de banheiros, equipamentos sanitários e número de quartos. Cruz *et al.* (2017) também observaram que o aumento no número de banheiros elevou o consumo de água dos domicílios em um estudo no México.

A variável idade do imóvel mostrou-se correlacionada ao consumo *per capita* de água, isto é, quanto maior a idade do imóvel, maior o consumo de água (Figura 31). A idade e a condição dos sistemas hidráulicas podem ter sido fatores contribuintes nesse resultado, pois favorecem o surgimento de vazamentos e o conseqüente aumento do consumo. Outra explicação é que os aparelhos sanitários foram aperfeiçoados ao longo do tempo e tornaram-se mais eficientes. Nos edifícios multifamiliares de Nova York, a idade da edificação também mostrou efeito

significativo sobre a intensidade no uso de água (KONTOKOSTA; JAIN, 2015). A idade da construção (-) e valor venal da propriedade (+) foram os principais parâmetros na explicação da variação no uso de água residencial em quatro cidades do Oeste dos EUA: Austin, Phoenix, Portland e Salt Lake City (CHANG *et al.*, 2017).

Figura 30 – Relação entre número de banheiros e consumo *per capita* de água

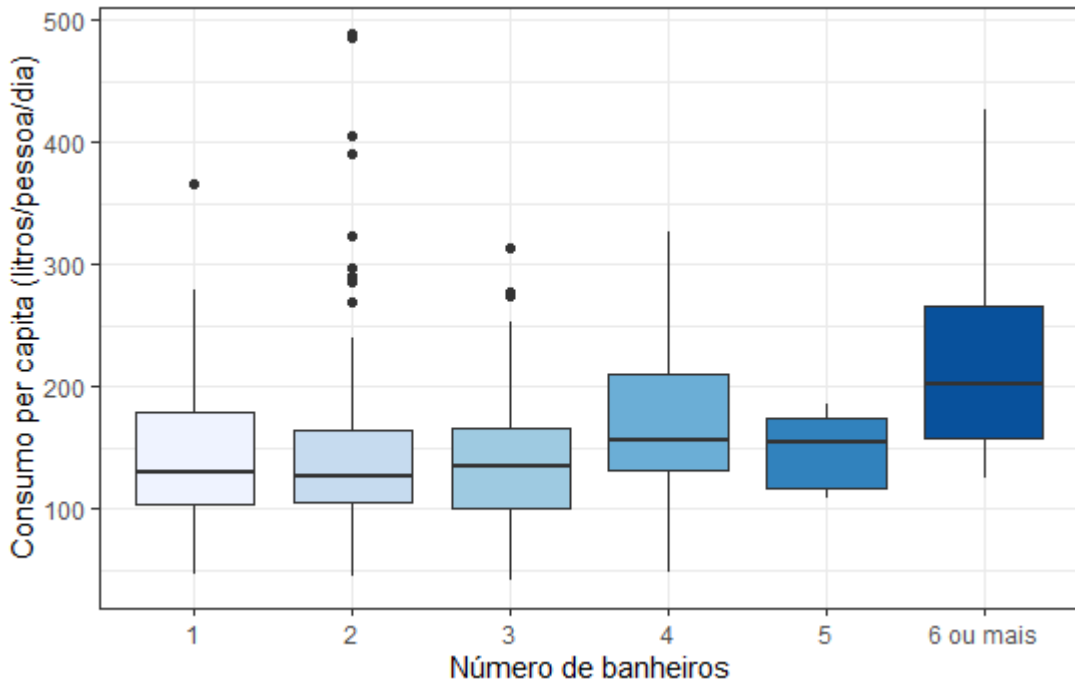
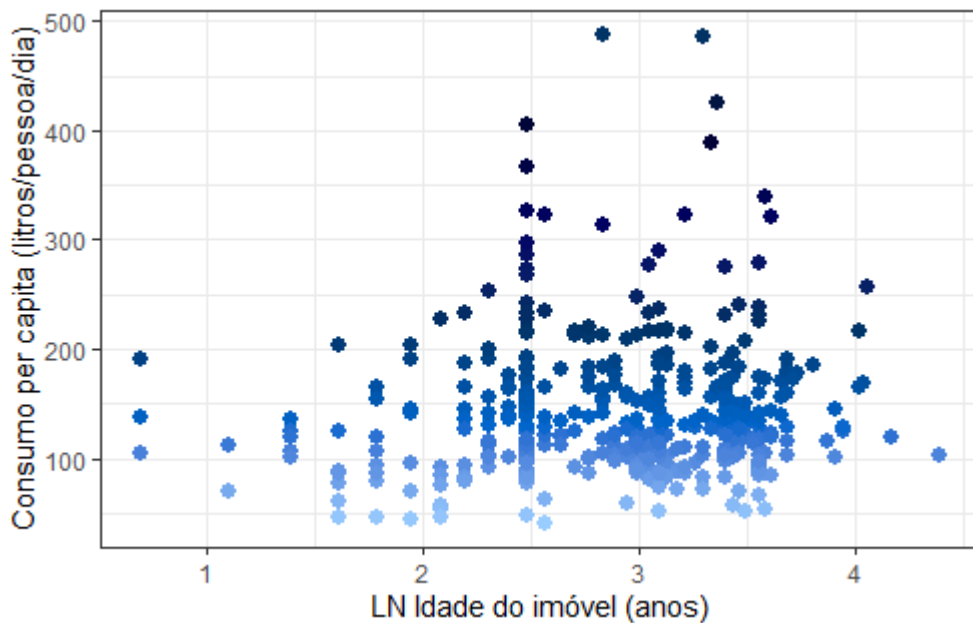


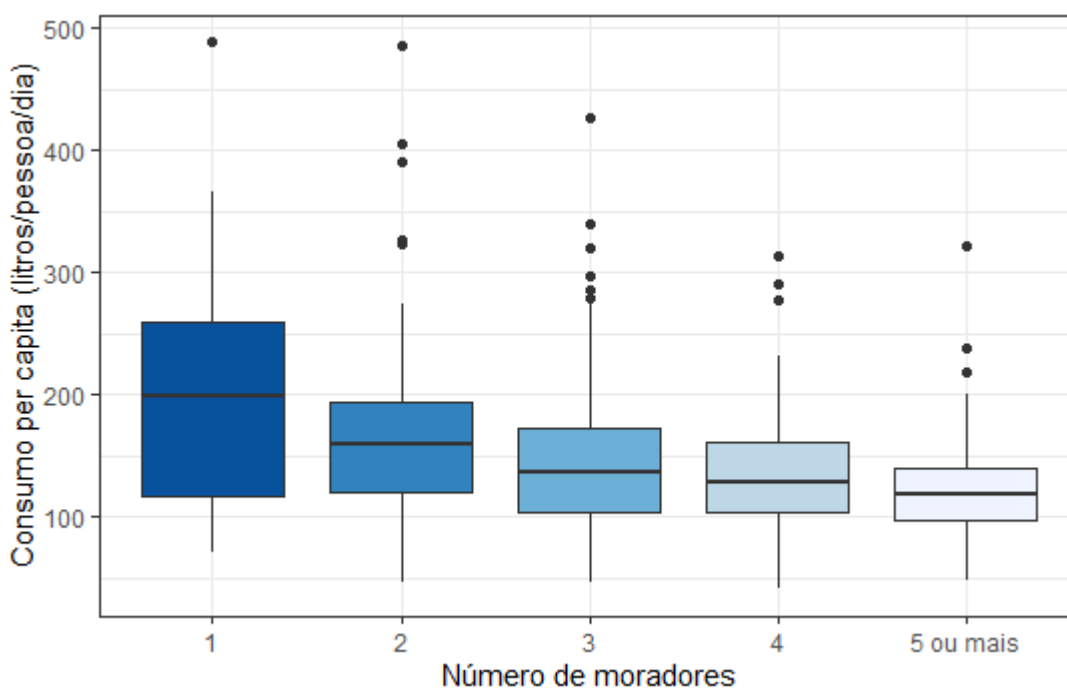
Figura 31 – Relação entre idade do imóvel e consumo *per capita* de água



Os fatores estatisticamente significativos apresentados até o momento foram avaliados individualmente por meio da análise de correlação. A seguir, são apresentadas as variáveis explicativas provenientes de uma análise integrada dos fatores, a modelagem de regressão linear múltipla e regressão robusta.

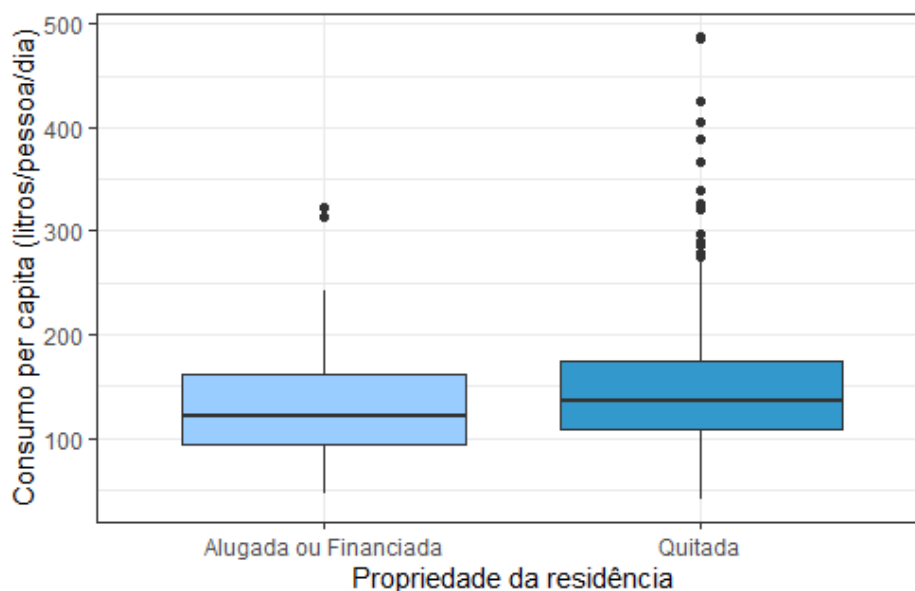
Como esperado, a análise de regressão mostrou que o número de moradores apresentou relação negativa e estatisticamente significativa com o consumo de água, portanto quanto maior o número de moradores na residência, menor o consumo *per capita*, conforme Figura 32. No Iraque, Hussien, Memon e Savic (2016) também identificaram que o consumo *per capita* diminuía com o aumento do número de moradores. Segundo Schleich e Hillenbrand (2009), um aumento no tamanho do agregado familiar diminuiria a demanda de água *per capita* do sistema público da Alemanha. Certamente, o tamanho da família exerce um efeito relevante sobre o consumo doméstico de água. Alguns usos de água são constantes e não dependem do número de pessoas por domicílio, como irrigação do jardim ou enchimento de piscina. Outros usos, como lavagem de roupas e louças, podem ser otimizados quando o número de usuários é maior. Por esse motivo, o consumo *per capita* pode ser maior em casas com famílias pequenas (DOMENE; SAURÍ, 2006).

Figura 32 – Relação entre número de moradores e consumo *per capita* de água



Os modelos de regressão apontaram que o consumo *per capita* tende a diminuir quando a residência é financiada ou alugada (Figura 33). Em Perth, na Austrália, Syme *et al.* (2004) encontraram uma correlação positiva entre o uso de água e a propriedade da residência, indicando que as casas ocupadas pelo proprietário tendem a usar mais água. Em um estudo em edificações residenciais multifamiliares em Joinville, Dias, Kalbusch e Henning (2018) concluíram que, quanto maior o percentual de inquilinos, menor era o consumo de água no edifício. Esses resultados podem indicar que quem possui despesas com aluguel e financiamento tende a economizar mais no uso de água, já que no Brasil despesas como conta de água e energia são de responsabilidade do inquilino e não do dono da residência.

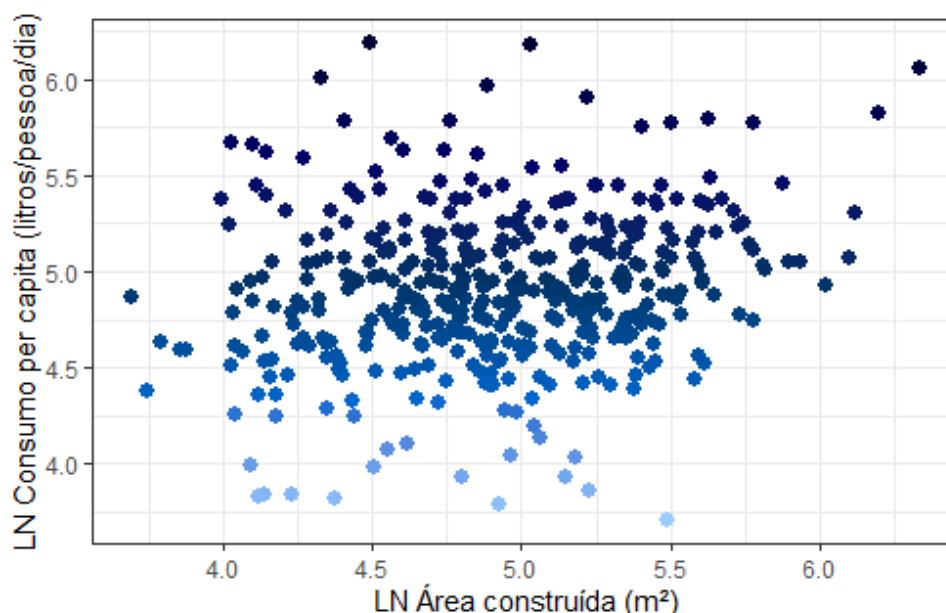
Figura 33 – Relação entre propriedade da residência e consumo *per capita* de água



Outra variável estatisticamente significativa nos modelos de regressão foi a área total do imóvel. Os resultados indicaram que o aumento da área construída eleva o consumo *per capita* de água das residências analisadas (Figura 34). Na Nigéria, Timothy e Paul (2014) também identificaram que o tamanho do imóvel é um fator relevante no aumento do consumo de água. No distrito Federal, a pesquisa de Sant'ana e Mazzega (2018) sugeriu que variáveis características da edificação (como área construída e área de jardim) elevaram o consumo doméstico de água. Em Jaffna, um distrito de Sri Lanka, o tamanho da edificação foi um fator que apresentou correlação positiva com o consumo de água (RAJEEVAN; MISHRA, 2020). Da mesma forma, Mayer e DeOreo (1999) obtiveram uma relação de aumento do consumo de

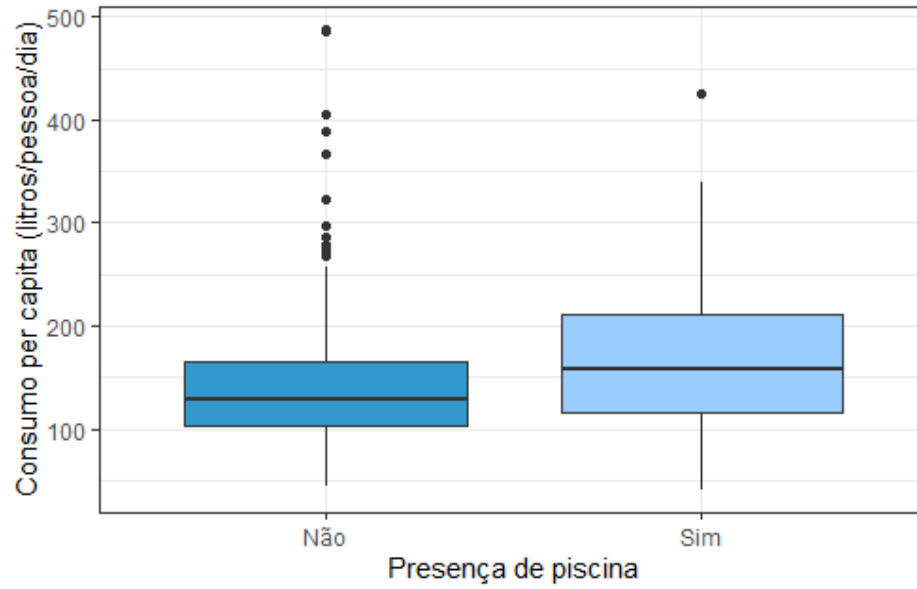
água em residências com maior área construída nos EUA, enquanto na Austrália, Loh e Coghlan (2003) encontraram uma diminuição do consumo em relação ao tamanho da residência.

Figura 34 – Relação entre área do imóvel e consumo *per capita* de água



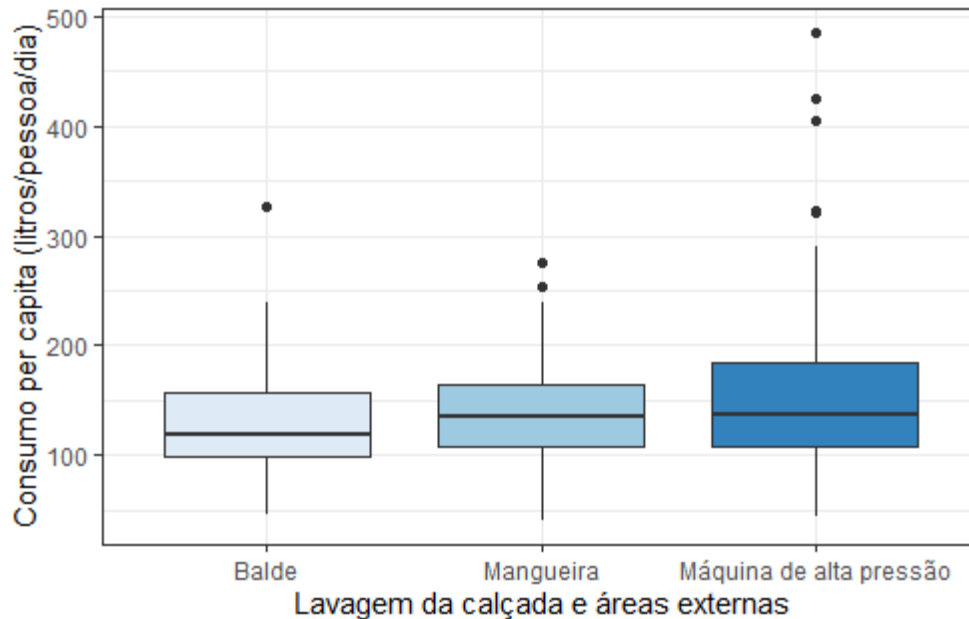
A área do terreno foi outro fator positivamente correlacionado com o consumo *per capita* de água na análise de correlação. No Distrito Federal, Sant’Ana e Mazzega (2018) obtiveram o mesmo resultado. Para os autores, grandes superfícies externas, muitas vezes jardins com piscina, necessitam de atividades rotineiras de manutenção como irrigação e limpeza, o que pode elevar o consumo de água. De fato, os resultados dos modelos de regressão mostraram que a presença de piscina provoca aumento no consumo *per capita* de água, como ilustra a Figura 35. Syme *et al.* (2004) observaram que em Perth, na Austrália, residências com piscina tendem a usar mais água. Segundo Guhathakurta e Gober (2010), é razoável esperar uma relação positiva entre a demanda de água e a existência de piscina em residências unifamiliares dado que as piscinas requerem quantidades significativas de água para compensar a perda de água por evaporação.

Figura 35 – Relação entre piscina e consumo *per capita* de água



Em relação aos hábitos de consumo dos usuários, o estudo de correlação revelou que o consumo *per capita* de água é maior nas casas cujos moradores utilizam máquina de alta pressão para lavagem da calçada e áreas externas. Por outro lado, os modelos de regressão mostraram que o uso de balde para lavagem da calçada e áreas externas tende a diminuir o consumo *per capita* (Figura 36). Portanto, conclui-se que o uso de balde é uma boa alternativa à máquina de alta pressão nas atividades de limpeza da área externa da casa.

Figura 36 – Relação entre equipamentos para lavagem de áreas externas e consumo *per capita*



O consumo *per capita* de água na China mostrou ser altamente afetado por investimentos em conservação de água. Por isso, os autores reforçam que essas medidas de intervenção devem ser mantidas para garantir a redução no uso de água (FAN *et al.*, 2017).

Diversas pesquisas abordam o potencial de conservação de água pelo uso de equipamentos economizadores (PÉREZ-URDIALES; GARCÍA-VALIÑAS, 2016; MARINOSKI; RUPP; GHISI, 2018) assim como fontes alternativas de água não potável (GHISI; FERREIRA, 2007; LEONG *et al.*, 2018), por exemplo a captação de água da chuva ou reúso de água. Porém, ao contrário do que se esperava, tais variáveis não apresentaram-se significativas na redução do consumo de água.

Um fato relevante nesse resultado é que somente 6,8% das residências avaliadas possuem algum tipo de sistema alternativo de abastecimento, seja aproveitamento água da chuva, captação de água subterrânea ou reúso de água. Sobre a utilização de equipamento economizadores de água, o percentual sobe para 43,5% das residências. Ainda assim, são valores baixos, possivelmente motivados pela falta de uma legislação nacional que regulamente a implementação de medidas para a conservação de água (PACHECO *et al.*, 2017).

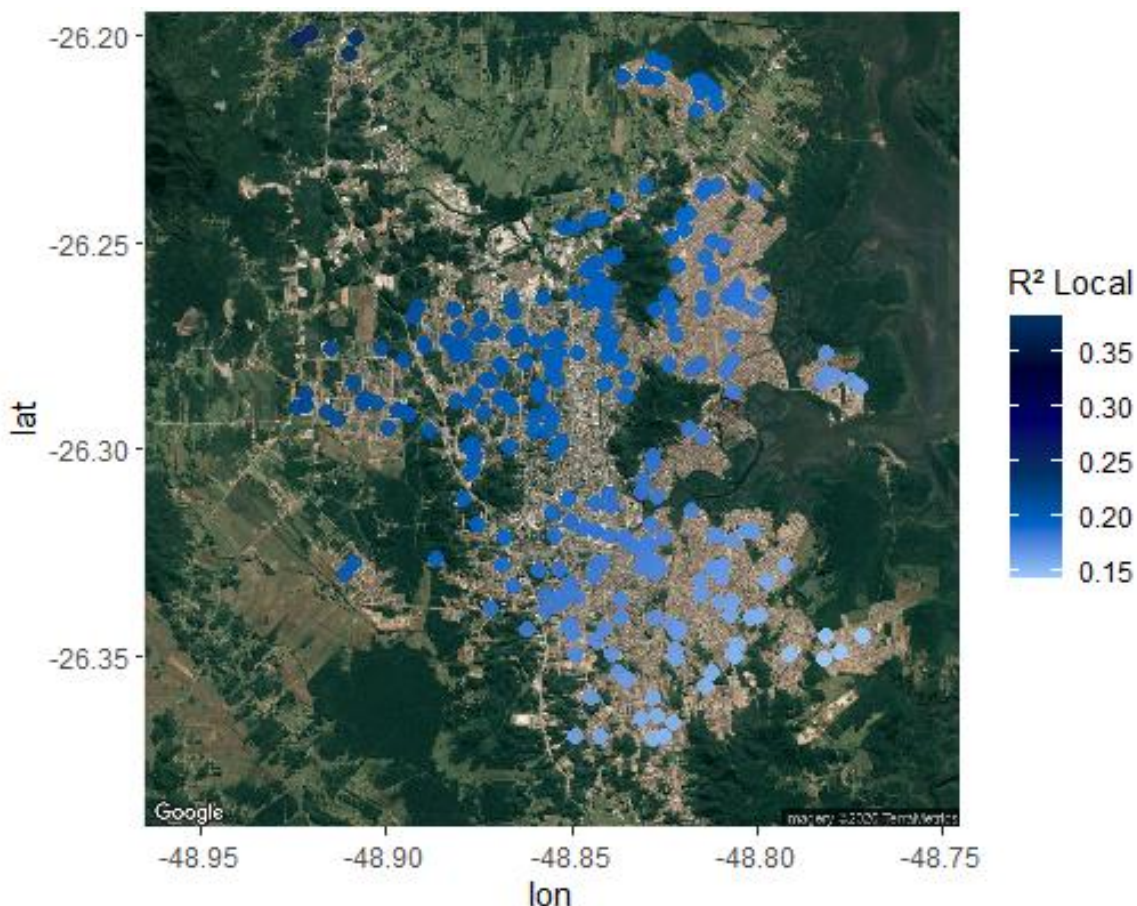
A análise de correlação possibilitou uma análise individual da influência de cada fator sobre o consumo *per capita* de água. A regressão linear múltipla e regressão robusta permitiram elaborar modelos que consideraram as variáveis explicativas de forma integrada e global. Entretanto, é necessário avaliar como os fatores variam espacialmente ao longo da área de estudo.

4.4 PADRÕES ESPACIAIS NO CONSUMO PER CAPITA DE ÁGUA

A análise espacial permitiu explorar espacialmente os padrões de consumo na área de estudo e foi realizada por meio da técnica *Geographically Weighted Regression* – GWR. Enquanto a regressão linear múltipla fornece um único modelo para a variável resposta, a análise GWR amplia essa regressão global produzindo uma equação de regressão para cada ponto do conjunto de dados em uma região pela variação dos coeficientes locais (ZHU, 2016). Dessa forma, a regressão geograficamente ponderada permite capturar as variações espaciais na relação entre um fenômeno ambiental e seus fatores influenciadores (FOTHERINGHAM; BRUNSDON; CHARLTON, 2002).

Os resultados mostraram que o modelo de regressão geograficamente ponderada foi estatisticamente significativo ao nível de 5%. Também pode-se considerar que o desempenho foi superior ao da regressão linear múltipla quando considerado o valor de *Akaike Information Criterion* – AIC, que reduziu de 255,53 no modelo OLS para 246,47 no modelo GWR (FOTHERINGHAN; BRUNSDON; CHARLTON, 2002). O aumento no poder explicativo do modelo (R^2 global=0,19833) corrobora o resultado anterior. A Figura 37 ilustra diferenças espaciais significativas nos coeficientes de determinação locais, confirmando a importância de considerar a heterogeneidade espacial nos modelos de regressão. Concluiu-se que o poder explicativo do modelo GWR é maior na região Norte do que da região Sul da cidade.

Figura 37 – GWR Coeficientes de determinação ao longo da área de estudo



Como mostra a Tabela 19 – Estimativa dos coeficientes do modelo GWR, as variáveis independentes que apresentaram alta variabilidade espacial foram: número de moradores, propriedade da residência (alugada ou financiada), área do imóvel, presença de piscina e uso de balde para lavagem de calçada e áreas externas. Os

quartis dos coeficientes locais de regressão mostraram que cada variável significativa manteve o mesmo efeito, positivo ou negativo, sobre todas as regiões da cidade. Por exemplo, o coeficiente do número de moradores variou de -0,10397 a -0,08998, revelando uma redução no consumo *per capita* em todas as regiões da cidade.

Tabela 19 – Estimativa dos coeficientes do modelo GWR

	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo	Global
Intercepto	4,26528	4,30302	4,33105	4,33105	4,59559	4,42309
Número total de moradores	-0,10397	-0,10145	-0,09461	-0,09203	-0,08998	-0,10266
Propriedade da residência (alugada ou financiada)	-0,158574	-0,14910	-0,14578	-0,14295	-0,13802	-0,14777
LN área construída (m²)	0,14155	0,17365	0,18794	0,19592	0,20089	0,17532
Presença de piscina	0,09876	0,11618	0,13302	0,16371	0,22266	0,14735
Uso de balde para lavagem de áreas externas	-0,11039	-0,10461	-0,10150	-0,09853	-0,09395	-0,10102

O diagnóstico de multicolinearidade na regressão é apresentado na Tabela 20. Como os valores mínimo e máximo de *VIF* – *Variance Inflation Factor* para as variáveis independentes foram menores que 5, os resultados indicaram ausência de multicolinearidade no modelo GWR (CONNOLLY; HAGELMAN III, 2015; FITZGERALD; AUERBACH, 2016; CRUZ *et al.*, 2017).

Tabela 20 – Valores mínimo e máximo de *VIF* – *Variance Inflation Factor* do modelo GWR

	VIF Mínimo	VIF Máximo
Número total de moradores	1,03980	1,10383
Propriedade da residência (alugada ou financiada)	1,02774	1,06229
LN área construída (m²)	1,13799	1,23529
Presença de piscina	1,05300	1,16125
Uso de balde para lavagem de áreas externas	1,00194	1,02773

A Figura 38 indica que a variável número de moradores varia espacialmente ao longo da área de estudo. Isto é, a quantidade de moradores por residência na região Sul de Joinville tem um impacto maior na redução do consumo do que na região Centro-Norte.

O modelo de regressão GWR apontou que o consumo *per capita* tende a diminuir quando a residência é financiada ou alugada. Essa redução é mais

significativa em regiões periféricas da cidade, principalmente nas regiões Nordeste, Sudeste, Leste e Oeste, como ilustra a Figura 39.

Figura 38 – Resultado GWR: número de moradores

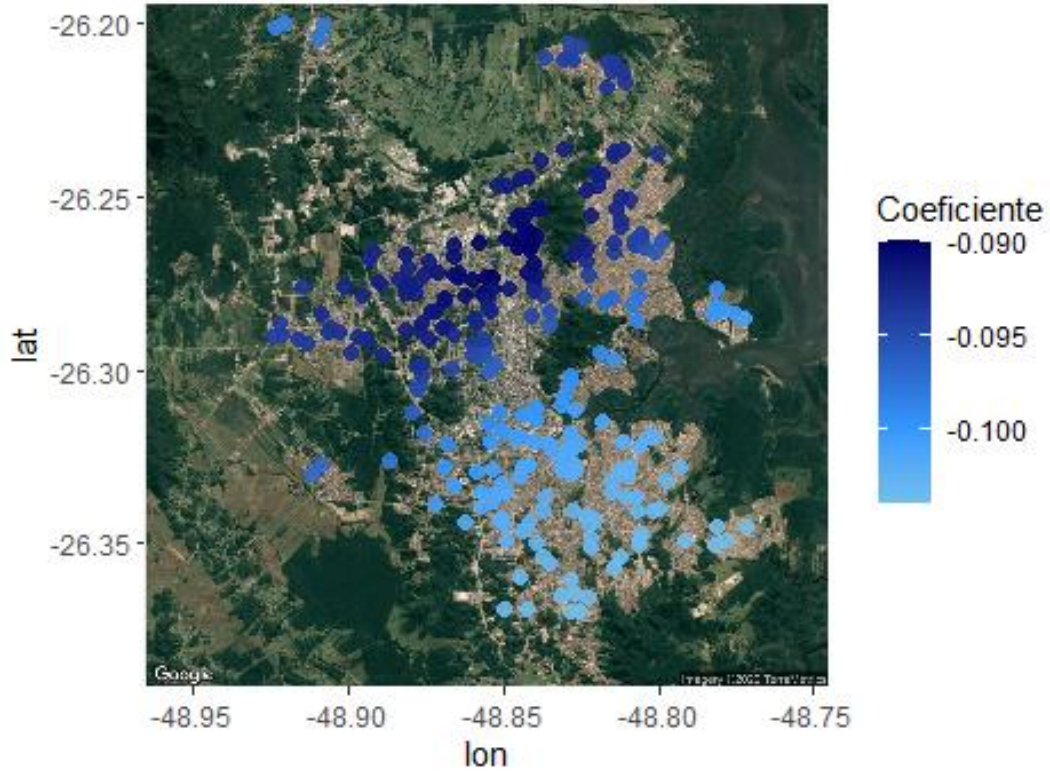
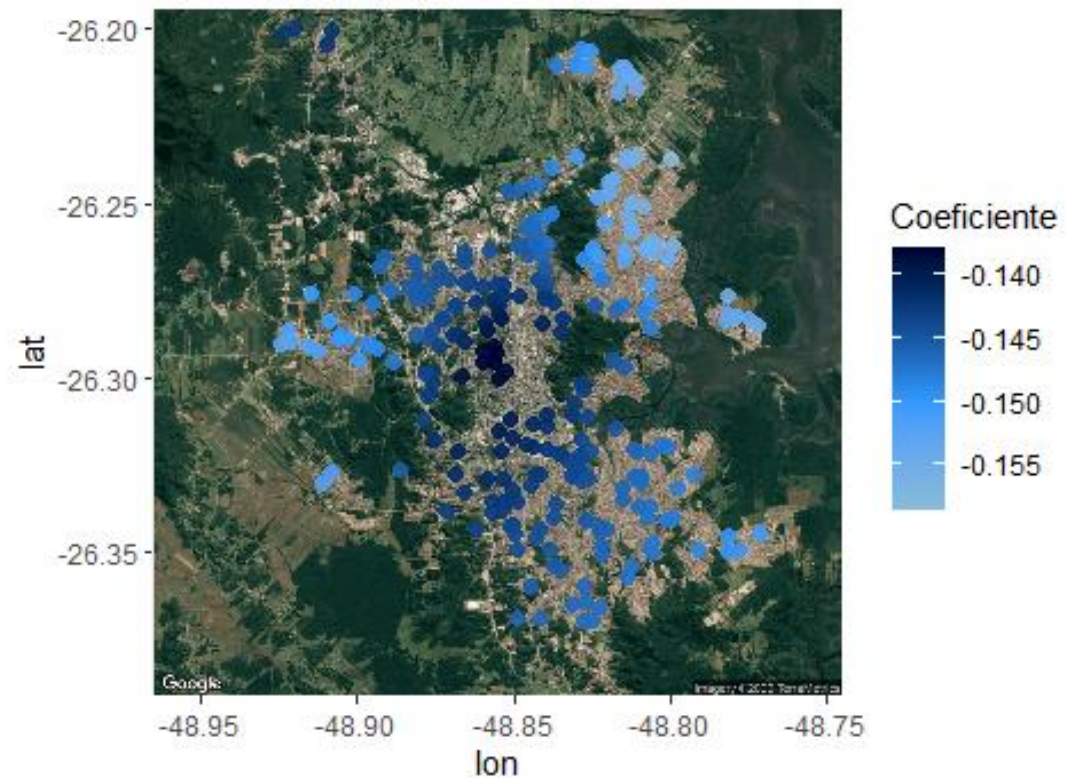
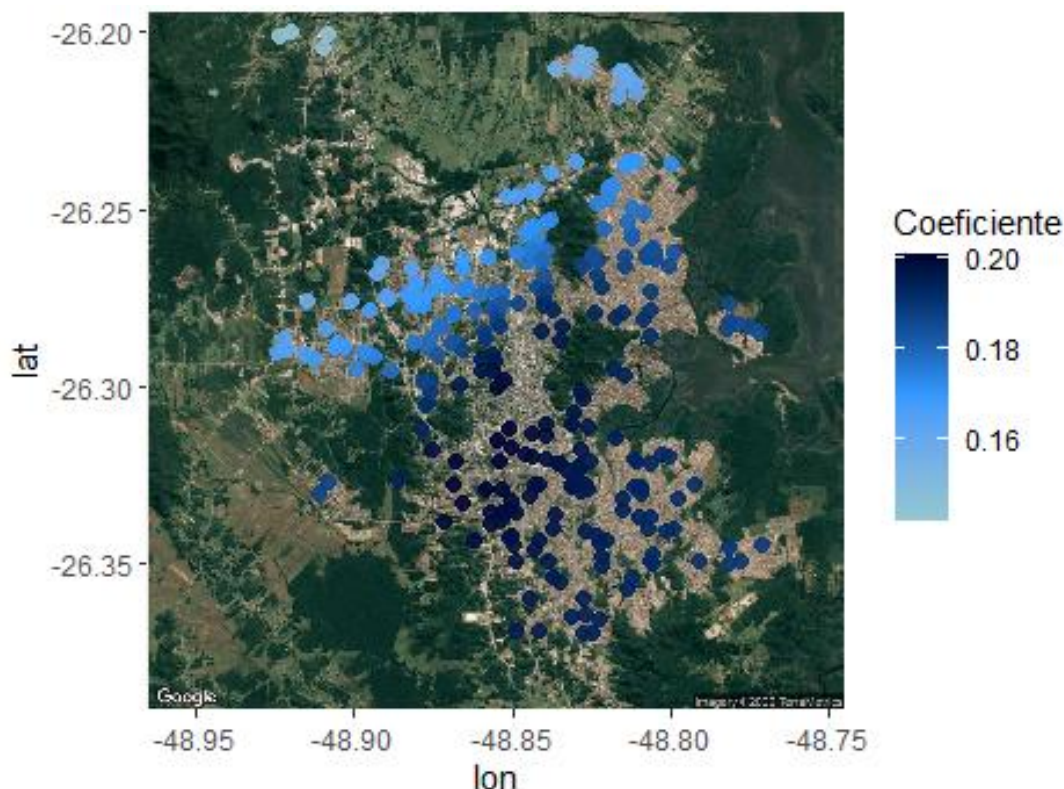


Figura 39 – Resultado GWR: Propriedade da residência



A análise espacial da variável área total do imóvel (Figura 40) revelou que o consumo de água da região Sul (pontos azul-escuros) é mais impactado pela variação na área da casa do que o da região Norte (pontos azul-claros).

Figura 40 – Resultado GWR: Área total do imóvel



Por fim, os resultados do modelo GWR indicaram que a presença de piscina na residência tende a elevar o consumo *per capita* de água e que seu efeito é maior na região Norte de Joinville (Figura 41).

Após o ajuste do modelo de regressão, é importante avaliar a qualidade do ajuste. Uma técnica útil é testar a capacidade do modelo em prever a resposta para novas observações, nesse caso a amostra de validação. Os resultados da análise de desempenho são apresentados na Tabela 21 e Apêndice C.

A diferença entre o consumo *per capita* medido pela Companhia Águas de Joinville e o previsto pelo modelo GWR foi, em média, 29,09%. O erro absoluto mediano foi 22,68%. Como esperado, os resultados indicaram que o erro na estimativa da regressão geograficamente ponderada foi menor que o da regressão linear múltipla e regressão robusta, isto é, o modelo GWR teve um desempenho melhor que os demais. Isso se deve ao fato de que a regressão geográfica considera as variações

espaciais dos fatores pela criação de equações locais de regressão, ao contrário da regressão linear múltipla, que gera uma única equação de regressão para todo o conjunto de dados.

Figura 41 – Resultado GWR: Presença de piscina

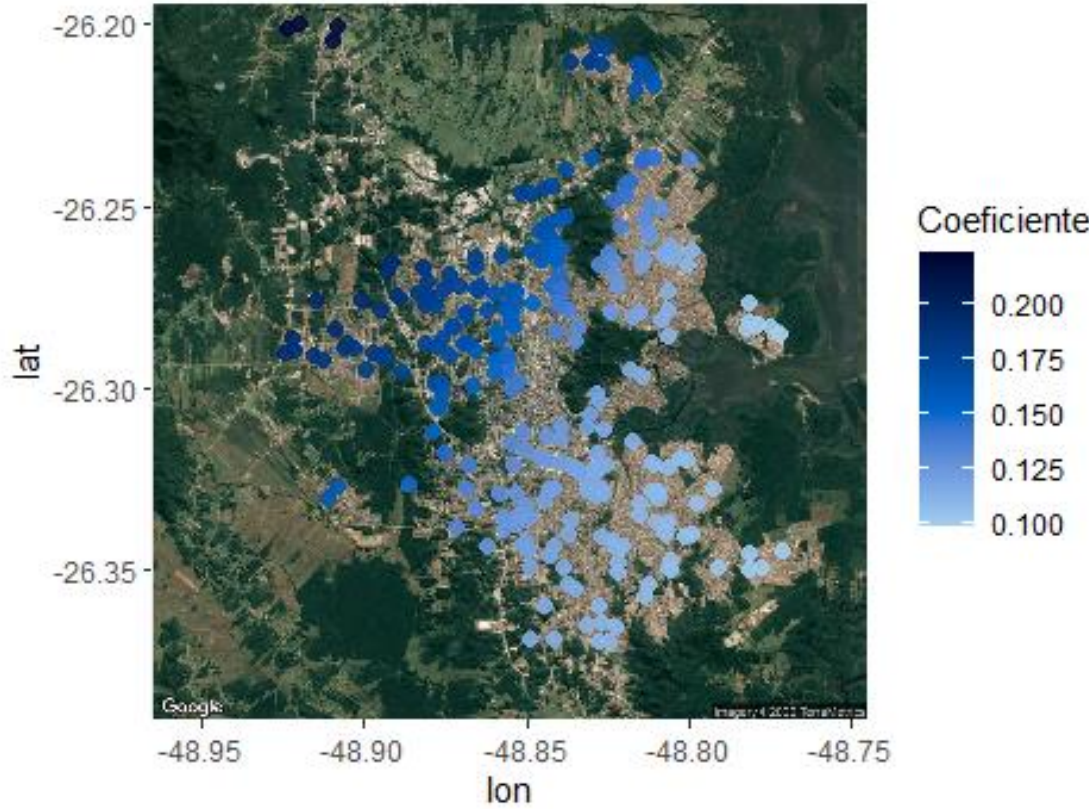


Tabela 21 – Indicadores da validação do modelo GWR

Método	MAE	MAPE	RMSE	R ² predição
Regressão geograficamente ponderada	42,24014	0,29097	58,19997	0,17691

Ao analisar as casas com erro absoluto percentual menor que 5% na estimativa do consumo *per capita*, percebeu-se que os modelos de regressão OLS, robusto e GWR foram mais assertivos nas residências localizadas na região norte da cidade, com número de moradores entre 2 e 5, com renda entre 1 e 9 SM, que não eram alugadas ou financiadas e que não possuíam piscina.

O valor do coeficiente de determinação R² (predição) foi 17,69%. Concluiu-se que o modelo possui uma capacidade de predição satisfatória uma vez que o valor de R² de predição (0,17691) está próximo ao R² do modelo original (0,19833). Dessa

forma, os resultados indicaram uma boa habilidade preditiva do modelo GWR pelo grau de aproximação entre o real consumo e o estimado.

Ao implementar a regressão geograficamente ponderada, percebeu-se a existência de lacunas na área, as quais necessitam ser explorados em pesquisas futuras para realizar uma análise estatística completa. Exemplo disso são: (i) como avaliar a confiabilidade dos coeficientes de um modelo GWR (ArcGIS, 2020) e (ii) como calcular intervalos de predição na validação de um modelo GWR (R SIG GEO, 2020).

A compreensão dos fatores determinantes e relações espaciais no uso de água pode fornecer uma visão crítica inicial para otimizar e direcionar políticas públicas, medidas de intervenção e incentivos voltados à economia de água em edificações. Além disso, a presente pesquisa também sugere que existem oportunidades para os gestores de recursos hídricos e planejadores urbanos trabalharem juntos a fim de alcançar a conservação e sustentabilidade da água (GOBER *et al.*, 2012), principalmente no projeto de novas construções uma vez que as características construtivas tiveram grande relevância na explicação da variação no consumo de água.

Dessa forma, conclui-se que os resultados dessa pesquisa podem ser úteis na explicação e previsão dos padrões de consumo em residências unifamiliares de Joinville. Além disso, os modelos propostos podem auxiliar na gestão da demanda de água ao fornecer estimativas de consumo e permitir a criação de cenários futuros.

As análises apresentadas representam uma etapa inicial na identificação das características construtivas, socioeconômicas e hábitos de consumo que mais afetam o consumo de água em residências unifamiliares de Joinville. As variações espaciais nos padrões de consumo sugerem que as estratégias de redução de consumo, incluindo campanhas de conscientização da população, precisam ser adaptadas às condições e realidades de cada região da cidade.

5 CONCLUSÕES

A gestão da demanda de água tem sido cada vez mais discutida, principalmente pela escassez hídrica em diversos países. Sendo assim, muitos estudos estão sendo elaborados com o intuito de desenvolver sistemas para redução do consumo de água. As pesquisas abrangem o entendimento dos parâmetros relevantes no consumo de água, medidas estratégicas de redução do consumo, incentivo às construções sustentáveis com a utilização de equipamentos economizadores de água e avaliação das práticas de conservação de água.

Dessa forma, este estudo buscou determinar os fatores que influenciam o consumo de água em residências unifamiliares na cidade de Joinville/SC. Para tanto, foi efetuada uma análise exploratória dos dados, examinando a representatividade da amostra. Posteriormente, investigou-se a presença de sazonalidade determinística no consumo de água. Também foram elaborados modelos de regressão linear múltipla e regressão robusta relacionando o consumo de água com variáveis construtivas das edificações, aspectos socioeconômicos e hábitos de consumo dos usuários. Por fim, foram avaliadas as variações espaciais no consumo de água de acordo com as regiões da cidade.

A revisão bibliográfica apresentou informações importantes relativas ao consumo urbano de água, além da descrição do método utilizado e resultados alcançados em pesquisas com foco semelhante ao da proposta desse trabalho. Essa exposição dos procedimentos metodológicos e análises estatísticas contribuiu na decisão acerca dos fatores a serem estudados e na definição do método adotado no presente estudo.

Na fase de levantamento dos dados, o envio dos formulários por correio eletrônico e redes sociais foi uma estratégia boa para alcançar o tamanho amostral almejado. Além disso, a aplicação dos questionários de forma presencial em bairros mais afastados do Centro da cidade e em terminais de ônibus permitiu obter dados de regiões com *déficit* de informação, aprimorando dessa forma a representatividade da amostra.

Ainda sobre a distribuição e representatividade da amostra, o teste qui-quadrado mostrou que o número de questionários alcançado em cada região se aproximou do estipulado no método de amostragem, resultando em uma boa distribuição espacial da amostra. A comparação entre os dados socioeconômicos amostrais e os oficiais revelou que a amostra conseguiu representar de forma satisfatória a população alvo da pesquisa. Todavia, a publicação de dados mais recentes como o próximo censo demográfico de 2020 poderá aprimorar essa análise.

A análise de sazonalidade do consumo de água foi fundamental uma vez que permitiu verificar a existência de padrões de consumo de acordo com as diferentes épocas do ano, auxiliando as empresas de abastecimento de água no planejamento adequado da demanda. A análise estatística indicou que o consumo de água nas residências unifamiliares é afetado pela componente sazonal, isto é, o uso de água varia conforme as épocas do ano, verão ou inverno.

Percebeu-se que não houve diferença significativa entre o consumo médio de água em 24 meses (dezembro/2017 a novembro/2019) e 48 meses (dezembro/2015 a novembro/2019) de medição. Sendo assim, foi definida a utilização da média do consumo de água como medida de tendência central do consumo *per capita* de cada residência, considerando um período de 24 meses (dezembro de 2017 a novembro de 2019). Estudos futuros podem incorporar a componente sazonal por meio do uso de dados em painel, que combinam a análise de séries temporais com dados de corte transversal.

Os resultados de correlação apontaram que das 57 variáveis independentes, quatorze tiveram relação numérica com o consumo *per capita*. Essas variáveis estatisticamente significativas aos níveis de 5% e 10% foram utilizadas na elaboração dos modelos de regressão linear múltipla, regressão robusta e regressão geograficamente ponderada. A amostra de 400 residências foi então dividida aleatoriamente em 70% (280 casas) para ajuste e 30% (120 casas) para validação do modelo.

A validação dos modelos foi realizada pela análise do desempenho dos modelos na predição do consumo *per capita* de água de 120 residências. Apenas 4,17% das observações ficou fora do intervalo de predição do modelo de regressão linear múltipla enquanto no caso da regressão robusta, esse percentual foi de 6,67%.

Os modelos de regressão sugeriram uma redução no consumo *per capita* de água baseado no aumento do número de moradores e no uso de balde para lavagem

de áreas externas. Outra conclusão relevante foi uma redução notória no consumo *per capita* quando a casa era alugada ou financiada. Por outro lado, o aumento na área construída da casa e a presença de piscina tendem a elevar o consumo de água.

A análise do modelo GWR permitiu ilustrar a variabilidade espacial que ocorre nas características construtivas (área da casa e presença de piscina) e aspectos socioeconômicos (número de moradores e propriedade da casa) ao longo da área de estudo. O estudo dos padrões de consumo apresentados aqui oferece suporte aos gestores públicos para a implementação de medidas e incentivos direcionados de acordo com as necessidades de cada região da cidade. Dessa forma, elaborar políticas públicas com base em dados científicos representa uma oportunidade valiosa de tornar as estratégias mais eficazes na consolidação do uso sustentável de água em edificações.

Por fim, a presente pesquisa representa um passo inicial no estudo do consumo de água em residências unifamiliares da cidade de Joinville. Sua principal contribuição é fornecer um entendimento claro dos parâmetros que mais influenciam o aumento ou a redução do consumo de água, inclusive da sua variabilidade espacial em toda a área de estudo, podendo ser útil para governos locais na criação de políticas sustentáveis.

5.1 LIMITAÇÕES DA PESQUISA

As limitações encontradas no desenvolvimento da pesquisa foram:

- Não foi possível obter informações sobre o consumo de energia elétrica com as Centrais Elétricas de Santa Catarina – CELESC visto que a empresa estava em fase de implementação de sua Política de Proteção aos Dados;
- Dificuldade em alcançar todas as faixas de renda, grau de escolaridade e faixas etárias. Por isso a importância da aplicação presencial dos questionários em terminais de ônibus e zonas mais afastadas do centro da cidade;
- Falta de dados oficiais atualizados sobre os aspectos socioeconômicos e demográficos de Joinville, sendo a maioria do último censo demográfico de 2010. Há uma defasagem de 10 anos, que afetou principalmente a análise de renda e grau de escolaridade;
- Não consideração da componente sazonal do consumo *per capita* de água nos modelos de regressão;

- Uso somente de questionários virtuais e presenciais. A falta de auditoria nas casas impossibilitou a detecção e análise de possíveis vazamentos visíveis;
- Precaução no uso do modelo para predição do consumo *per capita* de casas com histórico de consumo inferior a 60 litros por pessoa por dia pelo índice de falha nessa categoria;
- Ausência de função no software R para obtenção de parâmetros necessários à análise da confiabilidade dos coeficientes e para o cálculo dos intervalos de predição na validação do modelo GWR;
- Os resultados da pesquisa são válidos somente para a cidade de Joinville e tipologia de habitação analisada.

5.2 SUGESTÕES PARA TRABALHOS FUTUROS

Para complementação da pesquisa e possíveis trabalhos futuros, sugere-se:

- Incluir dados de consumo de energia elétrica nos modelos de regressão;
- Realizar auditoria nos sistemas hidráulicos das casas a fim de detectar qualquer vazamento visível, como no estudo desenvolvido por Sant'Ana e Mazzega (2018);
- Aprimorar a análise da representatividade da amostra utilizando dados do censo demográfico de 2020 e outros a serem publicados;
- Considerar a sazonalidade no consumo de água por meio da utilização de dados em painel ou criação de outros modelos cuja variável resposta seja a média do consumo de água somente no verão e inverno;
- Analisar a correlação espacial entre o consumo *per capita* de água e as regiões da cidade por meio do índice de Moran I;
- Uso de regressão quantílica para obter equações de regressão de acordo com determinadas faixas de consumo *per capita*;
- Investigar a influência do tipo de sistema de aquecimento no consumo de água;
- Expandir o estudo dos hábitos de consumo investigando, por exemplo, o tempo e frequência de banho, fechamento das torneiras quando não estão em uso, hábitos de lavagem de roupas e louças, além do monitoramento e conserto de vazamentos;
- Ampliar a área de estudo para além das fronteiras de uma cidade, considerando por exemplo os municípios da região Norte de Santa Catarina.

REFERÊNCIAS

ABNT – ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 5626:** Instalação predial de água fria. Set. 1998.

AKOGLU, Haldun. User's guide to correlation coefficients. **Turkist Journal of Emergency Medicine**, v. 18, n. 3, p. 91-93, set. 2018.

ALEXANDRE, Ananda C.; KALBUSCH, Andreza; HENNING, Elisa. Avaliação do impacto da substituição de equipamentos hidrossanitários convencionais por equipamentos economizadores no consumo de água. **Engenharia Sanitária e Ambiental**, Rio de Janeiro, v. 22, n. 5, p. 1005-1015, set./out. 2017.

ALI, Alireza Mashhadi; SHAFIEE, M. Ehsan; BERGLUND, Emily Zechman. Agent-based modeling to simulate the dynamics of urban water supply: Climate, population growth, and water shortages. **Sustainable Cities and Society**, v. 28, p. 420-434, jan. 2017.

ALMETWALLY, Ehab M.; ALMONGY, Hisham M. Comparison between M-estimation, S-estimation, and MM-estimation methods of robust estimation with application and simulation. International. **Journal of Mathematical Archive**, v. 9, n. 11, p. 55-63, nov. 2018.

_____. **Conjuntura dos recursos hídricos no Brasil 2018:** relatório pleno. Brasília: ANA, 2018. 72 p. Disponível em: <http://arquivos.ana.gov.br/portal/publicacao/Conjuntura2018.pdf>. Acesso em: 22 mar. 2019.

ArqGIS. **How Geographically Weighted Regression (GWR) works**, 2020. Disponível em: <https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/how-geographicallyweightedregression-works.htm>. Acesso em: 30 abr. 2020.

BARBETTA, Pedro Alberto; REIS, Marcelo Menezes; BORNIA, Antonio Cezar. **Estatística para Cursos de Engenharia e Informática**. 3. ed. São Paulo: Atlas, 2010. 416 p.

BARROS, Ana Luiza Bessa de Paula. **Revisitando o problema de classificação de padrões na presença de outliers usando técnicas de regressão robusta**. 2013. 132 f. Tese (Doutorado em Engenharia de Teleinformática) – Universidade do Ceará, Fortaleza, 2013.

BEAL, Cara D.; JACKSON, Melissa; STEWART, Rodney A.; RAYMENT, Cail; MILLER, Adrian. Identifying and understanding the drivers of high water consumption in remote Australian Aboriginal and Torres Strait Island communities. **Journal of Cleaner Production**, v. 172, p. 2425-2434, jan. 2018.

BEAL, Cara D.; STEWART, Rodney A.; FIELDING, Kelly. A novel mixed method smart metering approach to reconciling differences between perceived and actual residential end use water consumption. **Journal of Cleaner Production**, v. 60, p. 116-128, dez. 2013.

BOX, George E.P.; JENKINS, Gwilym M.; REINSEL, Gregory C. **Time Series Analysis: forecasting and control**. 3. ed. New Jersey: Prentice-Hall, 1994. 598 p.

BOWMAN, Adrian W. An Alternative Method of Cross-Validation for the Smoothing of Density Estimates. **Biometrika**, v. 71, n. 2, p. 353–360, ago. 1984.

BRASIL. **Lei n. 8.069 de 13 de julho de 1990**. Dispõe sobre o estatuto da criança e do adolescente e dá outras providências. Brasília, 1990.

_____. **Lei n. 10.741 de 1º de outubro de 2003**. Dispõe sobre o estatuto do idoso e dá outras providências. Brasília, 2003.

BRENNER, Johannes; GENOVA, Giulio; BERTOLDI, Giacomo; NIEDRIST, Georg; CHIESA, Stefano D. SWCalibrateR: Interactive, Web – Based Calibration of Soil Moisture Sensors. **Journal of Open Research Software**, v. 20 p. 1-7, jun. 2019.

BRUNSDON, Chris; FOTHERINGHAM, A. Stewart; CHARLTON, Martin E. Geographically weighted regression: A method for exploring spatial nonstationarity. **Geographical Analysis**, v. 28, n. 4, p. 281-298, out. 1996.

BUREK, Peter; SATOH, Yusuke; FISCHER, Günther; KAHIL, Taher; JIMENEZ, Luzma Nava; SCHERZER, Angelika; TRAMBEREND, Sylvia; WADA, Yoshihide; EISNER, Stefanie; FLÖRKE, Martina; HANASAKI, Naota; MAGNUSZIEWSKI, Piotr; COSGROVE, William; WIBERG, David. **Water futures and solution**: Fast track initiative. Áustria: IIASA, 2016. 88 p. Disponível em: <http://pure.iiasa.ac.at/id/eprint/13008/1/WP-16-006.pdf>. Acesso em: 18 abr. 2018.

CALDANA, Graziela; GABRIEL, Carmen Silvia. Avaliação do programa de acreditação hospitalar: validação de face e conteúdo. **Revista Brasileira de Enfermagem**, v.70, n.1, p.47-53, jan./fev. 2017.

CARLO, Joyce Correna. **Desenvolvimento de metodologia de avaliação da eficiência energética do envoltório de edificações não-residenciais**. 2008. 196 f. Tese (Doutorado em Engenharia Civil) – Universidade Federal de Santa Catarina, Florianópolis, 2008.

CHANG, Heejun; BONNETTE, Matthew Ryan; STOKER, Philip; CROW-MILLER, Britt; WENTZ, Elizabeth. Determinants of single family residential water use across scales in four western US cities. **Science of the Total Environment**, v. 596-597, p. 451-464, out. 2017.

CHANG, Heejun; PARANDVASH, G. Hossein; SHANDAS, Vivek. Spatial variations of single-family residential water consumption in Portland, Oregon. **Urban Geography**, v. 31, n. 7, p. 953-972, 2010.

CHENOWETH, Jonathan; LÓPEZ-AVILÉS, Alma; MORSE, Stephen; DRUCKMAN, Angela. Water consumption and subjective wellbeing: An analysis of British households. **Ecological Economics**, v. 130, p. 186-194, out. 2016.

COELHO, Marcelo; FERNANDES, Cristovão Vicente Scapulatempo; DETZEL, Daniel Henrique Marco; MANNICH, Michael. Statistical validity of water quality time series in urban watersheds. **Brazilian Journal of Water Resources**, Porto Alegre, v. 22, n. 51, jul. 2017.

COMUNELLO, Ana Carla; DAMINELLI, Lais Marques; BORTOLOTTI, Silvana Ligia Vincenzi. Análise e previsão da demanda de água no município de Medianeira-PR usando o modelo de Holt-Wintersthematic. **Revista Eletrônica Científica Inovação e Tecnologia**, v. 2, n. 8, p. 10-19, 2013.

CONNOLLY, Matthew H.; HAGELMAN III, Ronald R. Management implications of spatiotemporal non-stationarity in municipal water consumption patterns and drivers. **Applied Geography**, v.65, p. 70-81, dez. 2015.

COOK, R. Dennis; WEISBERG, Sanford. **Residuals and Influence in Regression**. New York: Chapman and Hall, 1982.

CROUX, Christophe; DEHON, Catherine. Influence functions of the Spearman and Kendall correlation measures. **Statistical Methods & Applications**, v. 19, p. 497-515, maio 2010.

CRUZ, Arturo O. de la; ALVAREZ-CHAVEZ, Clara R.; RAMOS-CORELLA, Marco A.; SOTO-HERNANDEZ, Fernando. Determinants of domestic water consumption in Hermosillo, Sonora, Mexico. **Journal of Cleaner Production**, v. 142, n. 4, p. 1901-1910, jan. 2017.

CUI, Zhe; BADAM, Sriram Karthik; YALÇIN, Adil; ELMQVIST, Niklas. DataSite: Proactive visual data exploration with computation of insight-based recommendations. **Information Visualization**, v. 18, n. 2, p. 251-267, abr. 2019.

CUSTÓDIO, Diego Antônio. **Influência do aproveitamento de água pluvial em edificações residenciais na drenagem urbana de Joinville/SC**. 2017. 168 f. Dissertação (Mestrado em Engenharia Civil) – Universidade Federal de Santa Catarina, Florianópolis, 2017.

CUSTÓDIO, Diego Antônio; GHISI, Enedir. Assessing the potential for potable water savings in the residential sector of a city: A case study of Joinville city. **Water**, v. 11, n. 10, p. 2074-2092, out. 2019.

DAY, Denise; HOWE, Carol. Forecasting peak demand: What do we need to know? **Water Science and Technology: Water Supply**, v. 3, p. 177-184, jun. 2003.

DEVORE, Jay L. **Probabilidade e Estatística para Engenharia e Ciências**. 8. ed. São Paulo: Cengage Learning, 2016. 633 p.

DIAS, Talita F.; KALBUSCH, Andreza; HENNING, Elisa. Factors influencing water consumption in buildings in southern Brazil. **Journal of Cleaner Production**, v. 184, p. 160-167, maio 2018.

DOMENE, Elena; SAURÍ, David. Urbanisation and Water Consumption: Influencing Factors in the Metropolitan Area of Barcelona. **Urban Studies**, v. 43, n. 9, p. 1605-1623, ago. 2006.

ERNST; Anja F.; ALBERS, Casper. Regression assumption in clinical psychology research practice: A systematic review of common misconceptions. **PeerJ**, v. 5, p. 1-16, 2017.

FACHIN, Odília. **Fundamentos de metodologia**. 3. ed. São Paulo: Saraiva, 2001. 200 p.

FAN, Liangxin; GAI, Lingtong; TONG, Yan; LI, Ruihua. Urban water consumption and its influencing factors in China: Evidence from 286 cities. **Journal of Cleaner Production**, v. 166, p. 124-133, nov. 2017.

FARCOMENI, Alessio; VENTURA, Laura. An overview of robust methods in medical research. **Statistical Methods in Medical Research**, v. 21, n. 2, p. 111-133, out. 2010.

FITZGERALD, Jared B.; AUERBACH, Daniel. The political economy of the water footprint: A cross-national analysis of ecologically unequal exchange. **Sustainability**, v. 8, n. 12, p. 1-16, dez. 2016.

FOTHERINGHAM, A. Stewart; BRUNSDON, Chris; CHARLTON, Martin E. **Geographically weighted regression: The analysis of spatially varying relationships**. Estados Unidos da América: John Wiley & Sons, 2002. 284 p.

FOTHERINGHAM, A. Stewart; CHARLTON, Martin E.; BRUNSDON, Chris. Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. **Environment and planning**, v. 30, p. 1905-1927, nov. 1998.

FOX, John; WEISBERG, Sanford; PRICE, Brad. **An R Companion to Applied Regression**. 3. ed. Thousand Oaks, CA, USA: SAGE Publications, 2019. Disponível em: <<https://socialsciences.mcmaster.ca/jfox/Books/Companion/>>. Acesso em: 10 jan. 2020.

GARCIA, Ana Paula Arruda de Almeida. **Fatores associados ao consumo de água em residências de baixa renda**. 2011. 122 f. Dissertação (Mestrado em Engenharia Industrial) – Universidade Federal da Bahia, Salvador, 2011.

GARCIA-CUERVA, Laura; BERGLUND, Emily Z.; BINDER, Andrew R. Public perceptions of water shortages, conservation behaviors, and support for water reuse in the U.S. **Resources, Conservation and Recycling**, v. 113, p. 106-115, out. 2016.

GHISI, Enedir; FERREIRA, Daniel F. Potential for potable water savings by using rainwater and greywater in a multi-storey residential building in southern Brazil. **Building and Environment**, v. 42, n. 7, p. 2512–2522, jul. 2007.

GHISI, Enedir; OLIVEIRA, Sulayre Mengotti de. Potential for potable water savings by combining the use of rainwater and greywater in houses in southern Brazil. **Building and Environment**, v. 42, n. 4, p. 1731-1742, abr. 2007.

GHOSH, Aindrila; NASHAAT, Mona; MILLER, James; QUADER, Shaikh; MARSTON, Chad. A comprehensive review of tools for exploratory analysis of tubular industrial datasets. **Visual Informatics**, v. 2, n. 4, p. 235-253, dez. 2018.

GOBER, Patricia; MIDDEL, Ariane; BRAZEL, Anthony; MYINT, Soe; CHANG, Heejun; DUH, Jiunn-Der; HOUSE-PETERS, Lily. Tradeoffs between water conservation and temperature amelioration in Phoenix and Portland: implications for urban sustainability. **Urban Geography**, v. 33, n. 7, p. 1030-1054, 2012.

GOLLINI, Isabella; LU, Binbin; CHARLTON, Martin; BRUNSDON, Chris; HARRIS, Paul. GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. **Journal of Statistical Software**, v. 63, n. 17, p. 1–50, 2015. Disponível em: <<http://www.jstatsoft.org/v63/i17/>>. Acesso em: 14 jan. 2020.

GOMES, Heber Pimentel. **Sistemas de abastecimento de água**: dimensionamento econômico e operação de redes e elevatórias. 3. ed. João Pessoa: Editora Universitária da UFPB, 2009. 277 p.

GONZÁLEZ, Francisco Cubillo; RUEDA, Teresa Moreno; LES, Silvia Ortega. **Cuadernos de I+D+i 4: Microcomponentes y Factores Explicativos del Consumo Doméstico de Agua en la Comunidad de Madrid**. Canal de Isabel II. Disponível em: <<http://www.madrid.org/bvirtual/BVCM008675.pdf>>. Acesso em: 6 jun. 2019.

GOOGLE EARTH. **Google Earth Pro**: computer program that renders a 3D representation of Earth based on satellite imagery. Version 7.3.1.4507. Estados Unidos da América: Google, 2018. Disponível em: <https://www.google.com/earth/>. Acesso em: 04 jun. 2018.

GRECO, Linda Del; WALOP, Wikke; MCCARTHY, Richard H. Questionnaire development 2: Validity and reliability. **CMAJ: Canadian Medical Association Journal**, v. 136, p. 699-700, abr. 1987.

GUHATHAKURTA, Subhrajit; GOBER, Patricia. Residential land use, the urban heat island, and water use in Phoenix: a path analysis. *Journal of Planning Education and Research*, v. 30, p. 40-51, 2010.

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

HERITIER, S. **Robust Methods in Biostatistics**. John Wiley & Sons, 2009.

HLPW – HIGH LEVEL PANEL ON WATER. **Joint Statement of the High Level Panel on Water**. Nova Iorque: United Nations, 2016. 2 p. Disponível em: <https://sustainabledevelopment.un.org/content/documents/11296Joint%20Statement%20HLPW.pdf>. Acesso em: 21 abr. 2018.

HUSSIEN, Wa'el A.; MEMON, Fayyaz A.; SAVIC, Dragan A. Assessing and modelling the influence of household characteristics on per capita water consumption. **Water Resources Management**, v. 30, n. 9, p. 2931-2955, jul. 2016.

_____. A risk-based assessment of the household water-energy-food nexus under the impact of seasonal variability. **Journal of Cleaner Production**, v. 171, p. 1275-1289, jan. 2018.

HYNDMAN, Rob J.; ATHANASOPOULOS, George. **Forecasting: principles and practice**. Austrália: Springer, 2018. Disponível em: <<https://otexts.org/fpp2/>>. Acesso em: 04 nov. 2018.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Resultados preliminares do universo do censo demográfico 2010**. Rio de Janeiro: IBGE, 2010. Disponível em: <<http://www2.datasus.gov.br/DATASUS/index.php?area=0206&id=7317548&VObj=http://tabnet.datasus.gov.br/cgi/defthtm.exe?ibge/censo/cnv/renda>>. Acesso em: 06 maio 2018.

_____. **Panorama das cidades e estados do Brasil**. Versão 4.3.8.1. Rio de Janeiro: IBGE, 2017. Disponível em: <<https://cidades.ibge.gov.br/brasil/sc/joinville/panorama>>. Acesso em: 06 maio 2018.

_____. **Panorama das cidades e estados do Brasil**. Versão 4.3.8.1. Rio de Janeiro: IBGE, 2019. Disponível em: <<https://cidades.ibge.gov.br/brasil/sc/joinville/panorama>>. Acesso em: 15 fev. 2020.

IPPUJ - INSTITUTO DE PESQUISA E PLANEJAMENTO URBANO DE JOINVILLE. **Mapa das regiões de abrangência das subprefeituras do município de Joinville. Joinville: IPPUJ, 2013**. Disponível em: <<https://www.joinville.sc.gov.br/wp-content/uploads/2017/04/Mapa-das-regi%C3%B5es-de-abrang%C3%Aancia-das-Subprefeituras-do-Munic%C3%ADpio-de-Joinville.pdf>>. Acesso em: 19 mar. 2019.

INEP – INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Questionário socioeconômico ENCCEJA 2013**. Brasília: INEP, 2013. Disponível em: <http://download.inep.gov.br/educacao_basica/encceja/questionario_socioeconomico/2013/questionario_socioeconomico_encceja_2013.p>. Acesso em: 14 maio 2018.

JANMAAT, Johannus John. Spatial patterns and policy implications for residential water use: An example using Kelowna, British Columbia. **Water Resources and Economics**, v. 1, p. 3-19, jan. 2013.

KAHLE, David; WICKHAM, Hadley. ggmap: Spatial Visualization with ggplot2. **The R Journal**, v. 5, n. 1, p. 144-161, jun. 2013. Disponível em <<http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>>. Acesso em: 25 jan. 2020.

KRASKA, Tim. Northstar: An interactive data science system. **Proceedings of the VLDB Endowment**, v. 11, n. 12, p. 2150-2164, ago. 2018.

KELLEY, Kate; CLARK, Belinda; BROWN, Vivienne; SITZIA, John. Good practice in the conduct and reporting of survey research. **International Journal for Quality in Health Care**, v. 15, n. 3, p. 261-266, 2003.

KONTOKOSTA, Constantine E.; JAIN, Rishree K. Modeling the determinants of large-scale building water use: Implications for data-driven urban sustainability policy. **Sustainable Cities and Society**, v. 18, p. 44-55, nov. 2015.

KOTHARI, Chakravanti Rajagopalachari. **Research Methodology: Methods and Techniques**. Índia: New Age International, 2004. 401 p.

KRZYWINSKI, Martin; ALTMAN, Naomi. Points of significance: Comparing samples – part I. **Nature Methods**, v. 11, n. 3, p. 215-216, mar. 2014.

KUHN, Max. **Caret: Classification and Regression Training**. R package version 6.0-86. Nova Zelândia: The R Foundation, 2020. Disponível em: <<https://CRAN.R-project.org/package=caret>>. Acesso em: 06 jan. 2020.

KUTNER, Michael H.; NACHTSHEIM, Christopher J.; NETER, John. **Applied Linear Regression Models**. 4. ed. Nova York: McGraw-Hill/Irwin, 2004. 701 p.

LADIRAY, Dominique; QUENNEVILLE, Benoit. **Seasonal Adjustment with the X-11 Method**. New York: Springer, 2001. 225 p.

LEE, Seung-Jae; CHANG, Heejun; GOBER, Patricia. Space and time dynamics of urban water demand in Portland, Oregon and Phoenix, Arizona. **Stochastic Environmental Research and Risk Assessment**, v. 29, n. 4, p. 1135-1147, maio 2015.

LEONG, Janet Yip Cheng; CHONG, Meng Nan; POH, Phaik Eong; VIERITZ, Alison; TALEI, Amin; CHOW, Ming Fai. Quantification of mains water savings from decentralised rainwater, greywater, and hybrid rainwater-greywater systems in tropical climatic conditions. **Journal of Cleaner Production**, v. 176, p. 946-958, mar. 2018.

LEWIS-BECK, Michael S.; BRYMAN, Alan; LIAO, Tim F. **The SAGE encyclopedia of social science research methods**. 3. ed. Thousand Oaks, CA, USA: SAGE Publications, 2004.

LI, Yao; TANG, Zhenghong; LIU, Can; KILIC, Ayse. Estimation and investigation of consumptive water use in residential area - case cities in nebraska, USA. **Sustainable Cities and Society**, v. 35, p. 637-644, nov. 2017.

LINDSAY, Jo; SUPSKI, Sian. Changing household water consumption practices after drought in three Australian cities. **Geoforum**, v. 84, p. 51-58, ago. 2017.

LIRA, Sachiko Araki; CHAVES NETO, Anselmo. Coeficientes de correlação para variáveis ordinais e dicotômicas derivados do coeficiente linear de Pearson. **Ciência e Engenharia**, Uberlândia, v. 14, n. 1/2, p. 45-53, jan./dez. 2006.

LOH, M.; COGHLAN, P. **Domestic Water Use Study**. Perth, Australia: Water Corporation, 2003.

LU, Binbin; CHARLTON, Martin; HARRIS, Paul; FOTHERINGHAM, A. Stewart. Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data. **International Journal of Geographical Information Science**, v. 28, n. 4, p. 660-681, jan. 2014.

LU, Binbin; HARRIS, Paul; CHARLTON, Martin; BRUNSDON, Chris. GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. **Geo-spatial Information Science**, v. 17, n. 2, p. 85–101, 2014. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/10095020.2014.917453>>. Acesso em: 18 fev. 2020.

MAECHLER, Martin; ROUSSEUW, Peter; CROUX, Christophe; TODOROV, Valentin; RUCKSTUHL, Andreas; SALIBIAN-BARRERA, Matias; VERBEKE, Tobias; KOLLER, Manuel; Conceicao, Eduardo L. T.; Anna di Palma, Maria. **robustbase: Basic Robust Statistics**. R package version 0.93-6. Nova Zelândia: The R Foundation, 2020. Disponível em <<http://robustbase.r-forge.r-project.org/>>. Acesso em: 09 jan. 2020.

MAKKI, Anas A.; STEWART, Rodney A.; BEAL, Cara D.; PANUWATWANICH, Kriengsak. Novel bottom-up urban water demand forecasting model: Revealing the determinants, drivers and predictors of residential indoor end-use consumption. **Resources, Conservation and Recycling**, v. 95, p. 15-37, fev. 2015.

MAKWIZA, Chikondi; FUAMBA, Musandji; HOUSSA, Fadoua; JACOBS, Heinz Erasmus. Estimating the impact of climate change on residential water use using panel data analysis: a case study of Lilongwe, Malawi. **Journal of Water, Sanitation and Hygiene for Development**, v. 8, n. 2, p. 217-226, set. 2018.

MANOUSELI, D.; ANDERSON, B.; NAGARAJAN, M. Domestic water demand during droughts in temperate climates: Synthesising evidence for an integrated framework. **Water Resources Management**, v.32, n. 2, p. 433-447, jan. 2018.

MARINOSKI, Ana Kelly; RUPP, Ricardo Forgiarini; GHISI, Enedir. Environmental benefit analysis of strategies for potable water savings in residential buildings. **Journal of Environmental Management**, v. 206, p. 28-39, jan. 2018.

MATOS, Cristina; TEIXEIRA, Carlos A.; BENTO, Ricardo; VARAJÃO, João; BENTES, Isabel. An exploratory study on the influence of socio-demographic characteristics on water end uses inside buildings. **Science of the Total Environment**, v. 466-467, p. 467-474, jan. 2014.

MATOS, Cristina; TEIXEIRA, Carlos A.; DUARTE, A. A. L. S.; BENTES, Isabel. Domestic water uses: Characterization of daily cycles in the north region of Portugal. **Science of the Total Environment**, v. 458-460, p. 444-450, ago. 2013.

MAYER, P. W.; DeOREO, W. B. **Residential End Uses of Water**. Denver, CO, USA: AWWARF, 1999.

MEDINA, Fabio Augusto Scalet. **Regressão logística geograficamente ponderada aplicada a modelos de *credit scoring***. 2016. 91 f. Dissertação (Mestrado em Administração) – Universidade de Brasília, Brasília, 2016.

MIOT, Hélio Amante. Correlation analysis in clinical and experimental studies. **Jornal Vascular Brasileiro**, Porto Alegre, v. 17, n. 4, p. 275-279, out./dez. 2018.

MONTGOMERY, Douglas C.; PECK, Elizabeth A.; VINING, G. Geoffrey. **Introduction to linear regression analysis**. 5. ed. United States of America: John Wiley & Sons, 2012. 645 p.

MONTGOMERY, Douglas C.; RUNGER, George C. **Applied statistics and probability for engineers**. 6. ed. United States of America: John Wiley & Sons, 2010. 811 p.

MONTGOMERY, Douglas C.; RUNGER, George C. **Estatística aplicada e probabilidade para engenheiros**. 6. ed. Rio de Janeiro: LTC, 2016. 629 p.

MORAN, P.A.P. Notes on Continuous Stochastic Phenomena. **Biometrika**, v. 37, p. 17-23, 1950.

MORETTIN, Pedro A.; BUSSAB, Wilton O. **Estatística Básica**. 8. ed. São Paulo: Saraiva, 2013. 548 p.

MORETTIN, Pedro A.; TOLOI, Clélia Maria de Castro. **Análise de séries temporais**. 2. ed. São Paulo: Blucher, 2006. 538 p.

NGUYEN, Linh; NOVÁK, Vilém. Forecasting seasonal time series based on fuzzy techniques. **Fuzzy Sets and Systems**, v. 361, p. 114-129, abr. 2019.

OGU, A. I.; INYAMA, S. C.; ACHUGAMONU, P. C. Methods of detecting outliers in a regression analysis model. **West African Journal of Industrial and Academic Research**, v. 7, n. 1, p. 105-113, jun. 2013.

OUYANG, Yun; WENTZ, Elizabeth A.; RUDELL, Benjamin L.; HARLAN, Sharon L. A multi-scale analysis of single family residential water use in the Phoenix metropolitan area. **Journal of the American Water Resources Association**, v. 50, n. 2, p. 448-467, abr. 2014.

ÖZTUNA, Derya; ELHAN, Atilla Halil; TÜCCAR, Ersöz. Investigation of four different normality test in terms of type 1 error rate and power under different distributions. **Turkish Journal of Medical Sciences**, v. 36, n. 3, p. 171-176, jan. 2006.

PACHECO, Paulo R. da Costa; GÓMEZ, Yapur Dumit; OLIVEIRA, Izabelle Ferreira de; TEIXEIRA, Luiza C. Girard. A view of the legislative scenario for rainwater harvesting in Brazil. **Journal of Cleaner Production**, v. 141, p. 290-294, jan. 2017.

PÉREZ-URDIALES, María; GARCÍA-VALIÑAS, María Ángeles. Efficient water-using technologies and habits: A disaggregated analysis in the water sector. **Ecological Economics**, v. 128, p. 117-129, ago. 2016.

PREFEITURA MUNICIPAL DE JOINVILLE. **Lei Ordinária Municipal nº 2260:** Estabelece a obrigatoriedade de existência de reservatório de água potável, 1988.

PUTH, Marie-Therese; NEUHÄUSER, Markus; RUXTON, Graeme D. Effective use of Pearson's product-moment correlation coefficient. **Animal Behaviour**, v. 93, p. 183-189, jul. 2014.

R CORE TEAM. **R: A language and environment for statistical computing and graphics**. Version 3.4.3. Nova Zelândia: The R Foundation, 2018. Disponível em: <<https://www.r-project.org/>>. Acesso em: 18 abr. 2018.

R SIG GEO. **GWR**: Confidence intervals on prediction, 2020. Disponível em: <<https://stat.ethz.ch/pipermail/r-sig-geo/2008-October/004270.html>>. Acesso em: 30 abr. 2020.

RAJEEVAN, Usitha; MISHRA, Binaya Kumar. Sustainable management of the groundwater resource of Jaffna, Sri Lanka with the participation of households: Insights from a study on household water consumption and management. **Groundwater for Sustainable Development**, v. 10, abr. 2020.

RAZALI, Nornadiah Mohd; WAH, Yap Bee. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. **Journal of Statistical Modeling and Analytics**, v. 2, n. 1, p. 21-33, 2011.

REIS, Marcelo Menezes. **Estatística aplicada à administração**. Florianópolis: Departamento de Ciências da Administração/UFSC, 2008. 280 p.

RIANI, Marco; CERIOLI, Andrea; ATKINSON, Anthony C.; PERROTTA, Domenico. Monitoring robust regression. **Electronic Journal of Statistics**, v. 8, p. 646-677, 2014. RIBEIRO, João Marcelo Pereira; BOCASANTA, Stephane Louise; ÁVILA, Brisa Oliveira; MAGTOTO, Mica; JONCK, Ana Valquiria; GABRIEL, George Matthew; GUERRA, José Baltazar Salgueirinho Osório de Andrade. The adoption of strategies for sustainable cities: A comparative study between Seattle and Florianopolis legislation for energy and water efficiency in buildings. **Journal of Cleaner Production**, v. 197, p. 366-378, out. 2018

ROMANO, Giulia; SALVATI, Nicola; GUERRINI, Andrea. Estimating the determinants of residential water demand in Italy. **Water**, v. 6, n. 10, p. 2929-2945, set. 2014.

_____. An empirical analysis of the determinants of water demand in Italy. **Journal of Cleaner Production**, v. 130, p. 74-81, set. 2016.

ROUSSEEUW, Peter J.; van ZOMEREN, Bert C. Unmasking Multivariate Outliers and Leverage Points. **Journal of the American Statistical Association**, v. 85, n. 411, p. 633-639, set. 1990.

SABINO, Claudia Vilhena Schayer; LAGE, Ludmila Vieira; NORONHA, César Vinicius. Variação sazonal e temporal da qualidade das águas em um ponto do Córrego Gameleiras usando técnicas quimiométricas robustas, **Engenharia Sanitária e Ambiental**, vol. 22, n. 5, p. 969-983, set./out., 2017.

SALKIND, N. J. **Encyclopedia of measurement and statistics**. 3. ed. Thousand Oaks, CA, USA: SAGE Publications, 2007.

SANT'ANA, Daniel; MAZZEGA, Pierre. Socioeconomic analysis of domestic water end-use consumption in the Federal District, Brazil. **Sustainable Water Resources Management**, v. 4, n. 4, p. 921-936, dez. 2018.

SAVENIJE, Hubert H. G. Why water is not an ordinary economic good, or why the girl is special. **Physics and Chemistry of the Earth**, v. 27, n. 11-12, p. 741-744, 2002.

SCHLEICH, Joachim; HILLENBRAND, Thomas. Determinants of residential water demand in Germany. **Ecological Economics**, v. 68, n. 6, p. 1756-1769, abr. 2009.

SCHMIDT, Amand F.; FINAN, Chris. Linear regression and the normality assumption. **Journal of Clinical Epidemiology**, v. 98, p. 146-151, jun. 2018.

SEPUD – SECRETARIA DE PLANEJAMENTO URBANO E DESENVOLVIMENTO SUSTENTÁVEL. **Joinville cidade em dados 2017**. Joinville: Prefeitura Municipal, 2017. 73 p. Disponível em: <https://www.joinville.sc.gov.br/wp-content/uploads/2016/01/Joinville-Cidade-em-Dados-2017.pdf>. Acesso em: 06 maio 2018.

_____. **Joinville cidade em dados 2018**: Ambiente construído. Joinville: Prefeitura Municipal, 2018. 50 p. Disponível em: <https://www.joinville.sc.gov.br/wp-content/uploads/2018/09/Joinville-Cidade-em-Dados-2018-Ambiente-Constru%C3%ADdo.pdf>. Acesso em: 17 mar. 2019.

_____. **Joinville cidade em dados 2019**: Ambiente construído. Joinville: Prefeitura Municipal, 2019. 57 p. Disponível em: <https://www.joinville.sc.gov.br/wp-content/uploads/2019/08/Joinville-Cidade-em-Dados-2019-Ambiente-Constru%C3%ADdo.pdf>. Acesso em: 29 jan. 2020.

SHI, Linrui; XU, Ping; WANG, Changzheng; GUAN, Tong; ZHANG, Yajun; XU, Huiwei. A review of applying spatial modelling and GIS in residential water use. **Materials Science and Engineering**, v. 392, 2018.

SILVA, Gisele Sanches da; TAMAKI, Humberto Oyamada; GONÇALVES, Orestes Marraccini. Implementação de programas de uso racional da água em campi universitários. **Ambiente Construído**, Porto Alegre, v. 6, n. 1, p. 49-61, jan./mar. 2006.

SILVA, Sérgio Ricardo dos Santos. **Avaliação do sistema de medição individualizada de água em prédios populares situados na cidade do Salvador – Bahia**. 2010. 146 f. Dissertação (Mestrado Profissional em Gerenciamento e Tecnologia Ambiental no Processo Produtivo) – Universidade Federal da Bahia, Salvador, 2010.

SNEE, Ronald D. Validation of Regression Models: Methods and Examples. **Technometrics**, v. 19, n. 4, p. 415-428, nov. 1977.

SNIS – SISTEMA NACIONAL DE INFORMAÇÕES SOBRE SANEAMENTO. **SNIS: Série Histórica**. 2018. Disponível em: <<http://app4.mdr.gov.br/serieHistorica/#>>. Acesso em: 25 jan. 2020.

SOARES, Gerson de Oliveira. **Avaliação da redução do consumo d'água pelo emprego de arejadores**. 2010. 120 f. Dissertação (Mestrado em Engenharia Civil) – Universidade Federal de Santa Catarina, Florianópolis, 2010.

STEVENS, James P. Outliers and influential data points in regression analysis. **Psychological Bulletin**, v. 95, n. 2, p. 334-344, mar. 1984.

SUSANTI, Yuliana; PRATIWI, Hasih; SULISTIOWATI, Sri; LIANA, Twenty. M estimation, S estimation, and MM estimation in robust regression. **International Journal of Pure and Applied Mathematics**, v. 91, n. 3, p. 349-369, 2014.

SYME, Geoffrey J.; SHAO, Quanxi; PO, Murni; CAMPBELL, Eddy. Predicting and understanding home garden water use. *Landscape and Urban Planning*, v. 68, n. 1, p. 121-128, maio 2004.

TIMOTHY, Ogunbode O.; PAUL, Ifabiyi I. Determinants of domestic water consumption in a growing urban centre in Osun state, Nigeria. **African Journal of Environmental Science and Technology**, v. 8, n. 4, p. 247-255, 2014.

TORTELLA, Bartolomé Deyà; TIRADO, Dolores. Hotel water consumption at a seasonal mass tourist destination: The case of the island of Mallorca. **Journal of Environmental Management**, v. 92, n. 10, p. 2568-2579, out. 2011.

TUFFÉRY, Stéphane. **Data mining and statistics for decision making**. United Kingdom: John Wiley & Sons, 2011. 689 p.

TUKEY, John W. **Exploratory data analysis**. Reading, PA: Addison-Wesley: 1977.

WALPOLE, Ronald E.; MYERS, Raymond H.; MYERS, Sharon L; YE, Keying. **Probabilidade e estatística para engenharia e ciências**. 8. ed. São Paulo: Pearson, 2009.

WICKHAM, Hadley. **ggplot2: Elegant Graphics for Data Analysis**. New York: Springer-Verlag, 2016. Disponível em: <<https://ggplot2.tidyverse.org>>. Acesso em: 10 fev. 2020.

WILLIS, Rachelle M.; STEWART, Rodney A.; GIURCO, Damien P.; TALEBPOUR, Mohammad R.; MOUSAVINEJAD, Alireza. End use water consumption in households: impact of socio-demographic factors and efficient devices. **Journal of Cleaner Production**, v. 60, p. 107-115, dez. 2013.

WILLIS, Rachelle M.; STEWART, Rodney A.; PANUWATWANICH, Kriengsak; WILLIAMS, Philip R.; HOLLINGSWORTH, Anna L. Quantifying the influence of environmental and water conservation attitudes on household end use water consumption. **Journal of Environmental Management**, v. 92, n. 8, p. 1996-2009, ago. 2011.

WWAP - World Water Assessment Programme. **The United Nations World Water Development Report 2018**: Nature-based solutions for water. Paris: UNESCO, 2018. 139 p. Disponível em: <<http://unesdoc.unesco.org/images/0026/002614/261424e.pdf>>. Acesso em: 06 abr. 2018.

XU, Guoce; LI, Peng; LU, Kexin; TANTAI, Zhan; ZHANG, Jiaxin; REN, Zongping; WANG, Xiukang; YU, Kunxia; SHI, Peng; CHENG, Yuting. Seasonal changes in water quality and its main influencing factors in the Dan River basin. **Catena**, v. 173, p. 131-140, fev. 2019.

ZEILEIS, Achim; HOTHORN, Torsten. Diagnostic Checking in Regression Relationships. **R News**, v. 2, n. 3, p. 7-10, dez. 2002 . Disponível em: <<https://CRAN.R-project.org/doc/Rnews/>>. Acesso em: 15 jan. 2020.

ZHANG, Yuexia; QIN, Guoyou; ZHU, Zhongyi; ZHANG, Jiajia. Robust estimation in linear regression models for longitudinal data with covariate measurement errors and outliers. **Journal of Multivariate Analysis**, v. 168, p. 261-275, nov. 2018.

ZHU, Xuan. **GIS for environmental applications**: A practical approach. Abingdon, Oxon, New York, NY: Routledge, 2016. 764 p.

APÊNDICES

APÊNDICE A – Modelo de questionário

Pesquisa sobre consumo de água nas residências unifamiliares de Joinville

Caro(a) morador(a), você está sendo convidado a participar de uma pesquisa de mestrado intitulada "Determinação dos fatores influentes no consumo de água em residências unifamiliares de Joinville", respondendo ao questionário solicitado.

Essa pesquisa leva aproximadamente 10 minutos e abrange questões relacionadas a aspectos socioeconômicos e demográficos, características construtivas da edificação, além dos equipamentos instalados e hábitos de consumo.

Sua opinião é fundamental para enriquecer a pesquisa e suas respostas serão anônimas.

Esperamos que isso não aconteça, mas caso você sinta algum desconforto em responder às perguntas, fique a vontade para abandonar a pesquisa.

Se você estiver de acordo, assinale a opção "concordo" a seguir e vamos às questões! Obrigado pela participação!

***Obrigatório**

Você concorda em responder esta pesquisa de forma voluntária? *

Ao concordar com a pesquisa, você afirma ser maior de 18 anos e aceita participar da pesquisa sobre "Consumo de água nas residências unifamiliares de Joinville". Os dados de consumo de água, consumo de energia elétrica e informações do projeto arquitetônico (área da casa, área do terreno e idade do imóvel) serão coletados diretamente com os órgãos responsáveis. A sua identidade será preservada, pois cada residência será identificada por um número. Caso você tenha interesse em ver o 'Termo de Consentimento Livre e Esclarecido' clique nesse link: <https://drive.google.com/file/d/1pXZLbCB3VfWVxvIbN2xIUe3zp7C50jRC/view?usp=sharing>. As informações neste termo são importantes e incluem o contato com o pesquisador responsável pela pesquisa. Oriente-se que você salve um print de tela e/ou imprima este documento.

☐ Sim, estou de acordo.

☐ Não, discordo.

Você mora em: *

☐ Joinville - Casa ou Geminado

☐ Joinville - Apartamento

☐ Outra cidade

Infraestrutura do entorno

Qual o endereço da sua casa (rua)? *

Exemplo: Rua Paulo Malschitzki

Sua resposta

Qual o número da sua casa? *

Exemplo: 200. Se a casa não possui número S/N, digite 0.

Sua resposta

Qual o bairro da sua casa? *

Exemplo: Zona industrial norte

Sua resposta

Nome do responsável pela conta de água: *

Sua resposta

A rua da sua casa é pavimentada? *

Pavimento: asfalto, paver, lajota, etc. Caso sua rua seja de "estrada de chão", responda Não.

☐ Sim

☐ Não

A rua da sua casa possui rede coletora de esgoto sanitário? *

☐ Sim

☐ Não

Informações dos moradores

Número de moradores da residência (escreva apenas números): *

Exemplo: 2 (caso morem 2 pessoas na residência)

Sua resposta

Gênero de cada morador:

	Feminino	Masculino	Outro
Morador 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Morador 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Morador 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Morador 4	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Morador 5	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Morador 6	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Idade morador 1 (escreva apenas números): *

Exemplo: 23 (caso o morador tenha 23 anos)

Sua resposta

Idade morador 2 (escreva apenas números):

Sua resposta

Idade morador 3 (escreva apenas números):

Sua resposta

Características socioeconômicas dos moradores

Grau de escolaridade do responsável pela conta de água: *

- ☐ Ensino fundamental incompleto
- ☐ Ensino fundamental completo
- ☐ Ensino médio
- ☐ Ensino superior
- ☐ Pós-graduação

Renda domiciliar mensal total: *

Considere a soma da renda de todos os membros da família. Para conhecimento, o valor do salário mínimo SM em 2018 é R\$954,00.

- ☐ Até 1 SM (até R\$954,00)
- ☐ De 1 a 3 SM (de R\$954,01 até R\$2.862,00)
- ☐ De 3 a 6 SM (de R\$2.862,01 até R\$5.724,00)
- ☐ De 6 a 9 SM (de R\$5.724,01 até R\$8.586,00)
- ☐ De 9 a 12 SM (de R\$8.586,01 até R\$11.448,00)
- ☐ De 12 a 15 SM (de R\$11.448,01 até R\$14.310,00)
- ☐ Mais de 15 SM (mais de R\$14.310,00)

Situação da residência: *

- ☐ Alugada ou cedida
- ☐ Própria (financiada)
- ☐ Própria (quitada)

Características construtivas da residência

Número de dormitórios (incluindo suítes): *

Escreva apenas números. Exemplo: 2, se a casa possui 2 dormitórios.

Sua resposta

Número de banheiros (incluindo lavabos): *

Escreva apenas números. Exemplo: 2, se a casa possui 2 banheiros.

Sua resposta

Tipologia: *

- ☐ Térrea
- ☐ Sobrado
- ☐ Geminado térreo
- ☐ Geminado sobrado

Possui piscina? *

- ☐ Sim
- ☐ Não

Possui reservatório (caixa d'água)? *

- ☐ Sim
- ☐ Não

Possui pressurizador de água? *

- ☐ Sim
- ☐ Não

Costuma faltar água na sua residência? *

- ☐ Sim
- ☐ Não

Falta de água

Com que frequência há falta de água? *

- ☐ Diária
- ☐ Semanal
- ☐ Mensal

Características construtivas da residência

Possui dependência de empregada? *

- ☐ Sim
- ☐ Não

Possui banheiro? *

- ☐ Sim
- ☐ Não

Uso da banheiro

Qual é a frequência de uso da banheiro? *

- ☐ Diária
- ☐ Semanal
- ☐ Mensal
- ☐ Nunca utiliza

Características construtivas da residência

Possui sistema alternativo de abastecimento de água? *

Exemplo: poço, aproveitamento água da chuva, reúso de água, etc.

☐ Sim

☐ Não

Qual(is) sistema(s) alternativo(s) é(são) utilizado(s)? *

☐ Aproveitamento de água da chuva

☐ Reúso de água

☐ Poço

☐ Outro: _____

Para qual(is) finalidade(s)? *

☐ Consumo de água de toda a residência

☐ Lavagem da calçada/áreas externas

☐ Lavagem do carro

☐ Lavagem de roupas

☐ Descarga do vaso sanitário

☐ Outro: _____

Há medidor do consumo de água para o sistema alternativo? *

☐ Sim

☐ Não

Equipamentos instalados

Sua família costuma comprar água mineral para beber? *

☐ Sim

☐ Não

Possui algum dos seguintes aparelhos?

☐ Tanquinho

☐ Máquina de lavar roupa

☐ Máquina de lavar louça

Possui algum(ns) equipamento(s) economizador(es) de água? *

Exemplo: vaso com 6 litros por fluxo (6lpf), vaso com duplo acionamento, arejador nas torneiras, etc.

☐ Sim

☐ Não

☐ Não sei

Exemplo: Vaso com 6lpf, vaso com duplo acionamento e torneira com arejador



Equipamentos economizadores de água

A residência possui qual(is) equipamento(s) economizador(es)? *

☐ Vaso sanitário com duplo fluxo

☐ Vaso sanitário com 6L/fluxo

☐ Arejador nas torneiras

☐ Outro: _____

Hábitos de consumo

Sua família reusa a água da máquina de lavar roupa? *

☐ Sim

☐ Não

Reuso água maquina lavar roupa

Para qual(is) finalidades(s) sua família reusa a água da máquina de lavar roupa? *

Sua resposta

Hábitos de consumo

Lava as roupas em casa? *

☐ Sim

☐ Não

Lava o carro em casa? *

☐ Sim

☐ Não

Lavagem do carro

Utiliza qual(is) equipamento(s) para lavar o carro? *

- ☐ Balde
- ☐ Mangueira
- ☐ Lavadora de alta pressão (Vap)
- ☐ Outro: _____

Com que frequência? *

- ☐ Semanal
- ☐ Quinzenal
- ☐ Mensal

Hábitos de consumo

Lava a calçada/áreas externas? *

- ☐ Sim
- ☐ Não

Lavagem da calçada/áreas externas

Utiliza qual(is) equipamento(s) para lavar a calçada/áreas externas? *

- ☐ Balde
- ☐ Mangueira
- ☐ Lavadora de alta pressão (Vap)
- ☐ Outro: _____

Com que frequência? *

- ☐ Semanal
- ☐ Mensal
- ☐ Semestral
- ☐ Outro: _____

Hábitos de consumo

É feita irrigação de jardim? *

- ☐ Sim
- ☐ Não

Irrigação de jardim

Utiliza qual(is) equipamento(s) para irrigar o jardim? *

- ☐ Mangueira
- ☐ Irrigador
- ☐ Aspersor
- ☐ Outro: _____

Com que frequência? *

- ☐ Diária
- ☐ Semanal
- ☐ Mensal

Obrigado pela sua participação!

Não esqueça de clicar em 'Enviar'. Sua resposta será armazenada e utilizada em análises estatísticas. Agradecemos seu tempo e apoio para tornar essa pesquisa possível!

Deixe seu email abaixo para receber um relatório final do estudo. Comentários também são bem-vindos!

APÊNDICE B – Dicionário de variáveis

Nº	Nome	Tipo	Descrição	Categoria	Fonte
YCB	CONSUMO_BRUTO	Numérica	Consumo bruto de água (m³) - média últimos 24 meses	-	Companhia Águas de Joinville
YIC	INDICADOR_CONSUMO	Numérica	Indicador de consumo de água (litros por pessoa por dia)	-	Companhia Águas de Joinville
INFRAESTRUTURA DO ENTORNO					
	LAT	Numérica	Latitude da residência	-	Google Maps
	LONG	Numérica	Longitude da residência	-	Google Maps
1	INF_PAVIMENTO_RUA	Dummy	Rua é pavimentada?	0 = Não 1 = Sim	Google Maps
2	INF_COLETA_ESGOTO	Dummy	Possui rede coletora esgoto?	0 = Não 1 = Sim	Companhia Águas de Joinville
CARACTERÍSTICAS SOCIOECONÔMICAS E DEMOGRÁFICAS					
3	SE_NUM_MORADORES	Numérica	Número total de moradores	-	Moradores
4	SE_NUM_MULHERES	Numérica	Número de mulheres	-	Moradores
5	SE_NUM_CRIANCAS	Numérica	Número de crianças (idade menor que 12 anos)	-	Moradores
6	SE_NUM_JOVENS	Numérica	Número de adolescentes e jovens (idade entre 12 e 17 anos)	-	Moradores
7	SE_NUM_ADULTOS	Numérica	Número de adultos (idade entre 18 e 59 anos)	-	Moradores
8	SE_NUM_IDOSOS	Numérica	Número de idosos (idade maior que 60 anos)	-	Moradores

CARACTERÍSTICAS SOCIOECONÔMICAS E DEMOGRÁFICAS					
9	SE_ESCOLARIDADE	Categórica	Grau escolaridade responsável pela conta água	1 = Fundamental incompleto 2 = Fundamental completo 3 = Ensino médio 4 = Ensino superior 5 = Pós-graduação	Moradores
10	SE_RENDA	Categórica	Renda domiciliar mensal total	1 = Até 1 SM 2 = De 1 a 3 SM 3 = De 3 a 6 SM 4 = De 6 a 9 SM 5 = De 9 a 12 SM 6 = De 12 a 15 SM 7 = Mais de 15 SM	Moradores
11	SE_PROPRIETARIO	Categórica	Situação da residência	1 = Alugada ou cedida 2 = Própria (financiada) 3 = Própria (quitada)	Moradores
CARACTERÍSTICAS CONSTRUTIVAS					
12	CO_NUM_QUARTOS	Numérica	Número de quartos	-	Moradores
13	CO_NUM_BANHEIROS	Numérica	Número de banheiros	-	Moradores
14	CO_IDADE_IMOVEL	Numérica	Idade do imóvel (anos)	-	Prefeitura Municipal de Joinville
15	CO_AREA_TERRENO	Numérica	Área do terreno (m²)	-	Prefeitura Municipal de Joinville
16	CO_AREA_TOTAL_IMOVEL	Numérica	Área total construída do imóvel (m²) - inclui área da casa, garagem, edícula	-	Prefeitura Municipal de Joinville
17	CO_AREA_CASA	Numérica	Área da casa (m²) - considera apenas a área da casa	-	Prefeitura Municipal de Joinville
18	CO_VALOR_VENAL	Numérica	Valor venal da residência (R\$)	-	Prefeitura Municipal de Joinville

CARACTERÍSTICAS CONSTRUTIVAS					
19	CO_TIPOLOGIA	Categórica	Tipologia	1 = Térrea 2 = Sobrado 3 = Geminado térreo 4 = Geminado sobrado	Google Maps
20	CO_PISCINA	Dummy	Possui piscina?	0 = Não 1 = Sim	Moradores
21	CO_RESERVATORIO	Dummy	Possui reservatório (caixa'água)?	0 = Não 1 = Sim	Moradores
22	CO_PRESSURIZADOR	Dummy	Possui pressurizador de água?	0 = Não 1 = Sim	Moradores
23	CO_FALTA_AGUA	Dummy	Há falta de água na residência?	0 = Não 1 = Sim	Moradores
24	CO_FREQ_AGUA	Categórica	Com que frequência há falta de água?	0 = Não falta água 1 = Diária 2 = Semanal 3 = Mensal	Moradores
25	CO_DEP_EMPREGADA	Dummy	Possui dependência de empregada?	0 = Não 1 = Sim	Moradores
26	CO_BANHEIRA	Dummy	Possui banheiro?	0 = Não 1 = Sim	Moradores
27	CO_FREQ_BANHEIRA	Categórica	Qual é a frequência de uso da banheiro?	0 = Não possui banheiro 1 = Diária 2 = Semanal 3 = Mensal 4 = Nunca utiliza	Moradores
28	CO_SIST_ALTERNATIVO	Dummy	Possui sistema alternativo abastecimento de água?	0 = Não 1 = Sim	Moradores

CARACTERÍSTICAS CONSTRUTIVAS					
29	CO_TIPO_SIST	Categórica	Qual sistema alternativo abastecimento é utilizado?	0 = Não possui sistema alternativo 1 = Aproveitamento água da chuva 2 = Captação água subterrânea (poço) 3 = Reúso de água	Moradores
30	CO_FINALIDADE_SIST	Categórica	Para qual finalidade é utilizado sistema alternativo?	0 = Não possui sistema alternativo 1 = Consumo água toda a residência 2 = Lavagem da calçada, áreas externas e carro 3 = Lavagem de roupas 4 = Descarga da bacia sanitária	Moradores
31	CO_MEDICAO_SIST	Dummy	Há medidor do consumo de água para o sistema alternativo?	0 = Não 1 = Sim	Moradores
EQUIPAMENTOS INSTALADOS					
32	EQ_TANQUINHO	Dummy	Possui tanquinho?	0 = Não 1 = Sim	Moradores
33	EQ_MAQ_ROUPA	Dummy	Possui máquina de lavar roupa?	0 = Não 1 = Sim	Moradores
34	EQ_MAQ_LOUCA	Dummy	Possui máquina de lavar louça?	0 = Não 1 = Sim	Moradores
35	EQ_ECONOMIZADOR	Dummy	Possui equipamento(s) economizador(es) de água?	0 = Não 1 = Sim	Moradores

EQUIPAMENTOS INSTALADOS					
36	EQ_BACIA_DUPLO	Dummy	Possui bacia sanitário com duplo acionamento?	0 = Não 1 = Sim	Moradores
37	EQ_BACIA_6L	Dummy	Possui bacia sanitária com 6L por fluxo?	0 = Não 1 = Sim	Moradores
38	EQ_AREJADOR	Dummy	Possui arejador em alguma torneira?	0 = Não 1 = Sim	Moradores
HÁBITOS DE CONSUMO					
39	HC_AGUA_MINERAL	Dummy	Sua família costuma comprar água mineral para beber?	0 = Não 1 = Sim	Moradores
40	HC_AGUA_MAQ	Dummy	Sua família reusa a água da máquina de lavar?	0 = Não 1 = Sim	Moradores
41	HC_FINALIDADE	Categórica	Para qual finalidade é reutilizada a água da máquina de lavar?	1 = Lavagem da calçada e áreas externas 2 = Lavagem do carro 3 = Lavagem de roupas 4 = Descarga da bacia sanitária 5 = Lavagem da casa 6 = Rega de jardim	Moradores
42	HC_ROUPA	Dummy	Lava as roupas em casa?	0 = Não 1 = Sim	Moradores
43	HC_CARRO	Dummy	Lava o carro em casa?	0 = Não 1 = Sim	Moradores
44	HC_BALDE_CARRO	Dummy	Utiliza balde para lavar o carro?	0 = Não 1 = Sim	Moradores
45	HC_MANGUEIRA_CARRO	Dummy	Utiliza mangueira para lavar o carro?	0 = Não 1 = Sim	Moradores

HÁBITOS DE CONSUMO					
46	HC_VAP_CARRO	Dummy	Utiliza máquina de alta pressão (Vap) para lavar o carro?	0 = Não 1 = Sim	Moradores
47	HC_FREQ_CARRO	Categórica	Com que frequência lava o carro?	0 = Não lava o carro 1 = Semanal 2 = Quinzenal 3 = Mensal	Moradores
48	HC_CALCADA	Dummy	Lava a calçada/áreas externas?	0 = Não 1 = Sim	Moradores
49	HC_BALDE_CALCADA	Dummy	Utiliza balde para lavar a calçada/áreas externas?	0 = Não 1 = Sim	Moradores
50	HC_MANGUEIRA_CALCADA	Dummy	Utiliza mangueira para lavar a calçada/áreas externas?	0 = Não 1 = Sim	Moradores
51	HC_VAP_CALCADA	Dummy	Utiliza máquina de alta pressão (Vap) para lavar a calçada/áreas externas?	0 = Não 1 = Sim	Moradores
52	HC_FREQ_CALCADA	Categórica	Com que frequência lava a calçada/áreas externas?	0 = Não lava a calçada 1 = Semanal 2 = Mensal 3 = Semestral	Moradores
53	HC_JARDIM	Dummy	É feita irrigação de jardim?	0 = Não 1 = Sim	Moradores
54	HC_MANGUEIRA_JARDIM	Dummy	Utiliza mangueira para irrigar o jardim?	0 = Não 1 = Sim	Moradores
55	HC_IRRIGADOR_JARDIM	Dummy	Utiliza irrigador para irrigar o jardim?	0 = Não 1 = Sim	Moradores
56	HC_ASPERSOR_JARDIM	Dummy	Utiliza aspersor para irrigar o jardim?	0 = Não 1 = Sim	Moradores

HÁBITOS DE CONSUMO					
57	HC_FREQ_JARDIM	Categórica	Com que frequência é irrigado o jardim?	0 = Não irriga o jardim 1 = Diária 2 = Semanal 3 = Mensal	Moradores

APÊNDICE C – Validação dos modelos

N	Casa	Consumo medido CAJ (l/p/d)	Regressão linear múltipla				Regressão robusta				Regressão geográfica	
			Consumo previsto (l/p/d)	Intervalo de predição (95%)		Erro absoluto (%)	Consumo previsto (l/p/d)	Intervalo de predição (95%)		Erro absoluto (%)	Consumo previsto (l/p/d)	Erro absoluto (%)
1	1	123,85	119,44	56,62	251,93	3,56	120,84	62,95	231,97	2,43	119,88	3,21
2	4	142,57	180,53	85,28	382,16	26,63	182,37	94,49	352,01	27,92	174,04	22,07
3	6	127,99	117,85	55,85	248,68	7,93	115,78	60,18	222,74	9,54	119,29	6,80
4	10	62,75	109,48	51,72	231,75	74,47	110,11	57,12	212,25	75,47	108,67	73,17
5	12	157,09	115,90	54,85	244,91	26,22	120,84	62,68	232,97	23,08	115,36	26,56
6	13	165,26	126,03	59,86	265,37	23,74	124,94	65,13	239,68	24,40	128,11	22,48
7	21	85,62	134,22	63,77	282,52	56,77	134,18	70,02	257,15	56,71	132,72	55,01
8	22	101,26	110,83	51,76	237,28	9,45	111,48	57,21	217,23	10,09	103,95	2,66
9	23	85,67	138,97	65,99	292,67	62,22	139,57	72,82	267,50	62,92	140,45	63,94
10	28	81,90	140,05	66,49	294,99	70,99	140,79	73,45	269,87	71,90	139,27	70,04
11	30	147,61	180,95	85,48	383,05	22,58	182,85	94,74	352,93	23,87	180,47	22,26
12	31	170,50	137,20	65,17	288,87	19,53	137,56	71,78	263,62	19,32	136,40	20,00
13	33	277,77	143,62	67,69	304,73	48,30	144,66	74,65	280,32	47,92	135,93	51,06
14	35	180,48	155,45	73,75	327,64	13,87	149,97	78,00	288,37	16,90	152,59	15,45
15	36	105,12	138,54	65,79	291,74	31,79	139,08	72,57	266,55	32,30	138,89	32,13
16	40	158,88	163,07	77,25	344,23	2,64	162,75	84,74	312,59	2,44	162,87	2,51
17	45	183,18	127,30	59,94	270,32	30,51	127,90	65,86	248,37	30,18	125,76	31,34
18	50	98,85	147,46	70,07	310,31	49,18	145,21	75,74	278,43	46,91	145,96	47,66
19	51	104,30	144,59	68,72	304,24	38,64	142,02	74,06	272,33	36,17	141,97	36,12
20	61	165,72	160,93	76,35	339,21	2,89	155,98	81,15	299,80	5,88	159,88	3,53
21	62	155,15	143,67	68,05	303,35	7,39	137,16	71,15	264,42	11,59	139,77	9,91

N	Casa	Consumo medido CAJ (l/p/d)	Regressão linear múltipla				Regressão robusta				Regressão geográfica	
			Consumo previsto (l/p/d)	Intervalo de predição (95%)		Erro absoluto (%)	Consumo previsto (l/p/d)	Intervalo de predição (95%)		Erro absoluto (%)	Consumo previsto (l/p/d)	Erro absoluto (%)
22	65	184,75	144,78	68,66	305,29	21,64	146,20	76,22	280,41	20,87	145,02	21,50
23	68	102,91	113,75	53,85	240,28	10,53	118,30	61,39	227,99	14,96	114,75	11,51
24	75	167,91	137,31	65,22	289,11	18,22	137,69	71,85	263,87	18,00	136,15	18,91
25	79	118,80	146,24	68,84	310,65	23,10	148,48	76,88	286,77	24,99	143,76	21,01
26	81	133,77	164,35	77,95	346,51	22,85	159,73	83,10	307,02	19,41	162,89	21,76
27	82	171,04	146,57	69,65	308,42	14,31	144,22	75,22	276,53	15,68	146,52	14,33
28	88	95,69	133,32	62,92	282,47	39,32	133,90	69,43	258,26	39,93	133,88	39,91
29	89	116,71	130,14	61,18	276,85	11,51	136,71	70,48	265,16	17,14	139,60	19,61
30	98	322,27	159,22	75,03	337,87	50,59	167,14	86,51	322,93	48,14	163,41	49,29
31	100	97,90	115,48	54,73	243,68	17,96	116,32	60,55	223,44	18,81	114,96	17,43
32	101	176,55	154,51	73,00	327,05	12,48	157,16	81,37	303,52	10,99	153,03	13,32
33	103	323,47	147,00	69,68	310,14	54,55	140,77	73,09	271,10	56,48	147,14	54,51
34	104	111,55	141,37	67,18	297,47	26,73	138,43	72,16	265,54	24,09	139,41	24,97
35	107	184,96	129,67	61,33	274,20	29,89	129,88	67,15	251,21	29,78	127,51	31,06
36	108	273,90	158,99	75,43	335,09	41,95	153,84	80,03	295,72	43,83	158,35	42,19
37	109	163,12	173,20	81,22	369,35	6,18	180,03	92,29	351,20	10,37	179,22	9,87
38	110	192,05	175,35	82,75	371,59	8,70	181,38	93,99	350,02	5,55	170,47	11,24
39	116	45,65	102,77	48,52	217,70	125,13	103,16	53,36	199,43	125,98	100,47	120,09
40	117	84,52	102,86	48,54	217,96	21,70	106,14	54,87	205,33	25,58	102,06	20,76
41	118	123,69	99,14	46,76	210,21	19,85	99,04	51,20	191,56	19,93	96,49	21,98
42	119	134,62	127,95	60,60	270,15	4,96	124,33	64,64	239,16	7,65	126,49	6,04
43	124	127,56	151,60	72,01	319,16	18,84	149,84	78,14	287,32	17,46	150,50	17,98
44	126	120,47	110,56	52,34	233,57	8,22	110,71	57,54	213,01	8,09	114,31	5,11

N	Casa	Consumo medido CAJ (l/p/d)	Regressão linear múltipla				Regressão robusta				Regressão geográfica	
			Consumo previsto (l/p/d)	Intervalo de predição (95%)		Erro absoluto (%)	Consumo previsto (l/p/d)	Intervalo de predição (95%)		Erro absoluto (%)	Consumo previsto (l/p/d)	Erro absoluto (%)
45	127	120,93	106,76	50,33	226,46	11,72	110,71	57,15	214,46	8,45	106,80	11,68
46	129	151,44	144,18	68,39	303,99	4,79	145,52	75,88	279,08	3,91	145,28	4,07
47	141	156,74	116,06	54,96	245,07	25,96	117,73	61,04	227,08	24,89	113,93	27,31
48	149	202,04	99,35	46,90	210,44	50,83	102,05	52,79	197,26	49,49	98,04	51,48
49	152	116,61	127,63	60,35	269,92	9,45	127,45	66,19	245,43	9,30	127,60	9,43
50	156	101,79	144,59	68,72	304,24	42,05	142,02	74,06	272,33	39,52	142,56	40,06
51	159	119,82	132,05	62,74	277,93	10,20	131,72	68,73	252,46	9,93	133,62	11,51
52	163	216,14	136,21	64,31	288,53	36,98	141,16	72,95	273,17	34,69	134,94	37,57
53	165	142,48	154,51	73,30	325,67	8,44	148,94	77,45	286,42	4,53	151,53	6,35
54	168	108,60	141,16	67,00	297,40	29,98	142,07	74,11	272,34	30,81	141,39	30,19
55	169	85,85	126,87	60,09	267,86	47,79	122,46	63,56	235,93	42,65	125,00	45,61
56	170	144,56	128,55	60,88	271,42	11,08	125,00	64,98	240,43	13,53	126,37	12,59
57	172	92,51	151,71	72,06	319,40	63,99	149,97	78,21	287,57	62,10	150,48	62,66
58	173	101,14	101,18	47,81	214,15	0,04	106,50	55,23	205,36	5,30	103,98	2,81
59	175	54,17	108,26	51,17	229,06	99,84	108,81	56,32	210,21	100,86	105,14	94,08
60	177	119,87	118,67	56,00	251,46	1,00	117,46	60,62	227,58	2,01	118,53	1,12
61	186	94,42	182,75	86,32	386,92	93,55	184,92	95,81	356,91	95,84	180,34	90,99
62	191	56,62	143,31	67,61	303,75	153,11	148,34	76,74	286,74	161,99	149,52	164,08
63	194	105,05	126,53	60,10	266,39	20,45	125,50	65,42	240,72	19,47	127,61	21,48
64	200	102,37	117,23	55,51	247,60	14,52	115,75	60,17	222,67	13,07	115,49	12,82
65	203	425,92	215,68	101,03	460,44	49,36	223,11	115,00	432,84	47,62	218,25	48,76
66	206	190,74	114,17	54,05	241,19	40,14	118,80	61,64	228,97	37,71	113,27	40,61
67	208	100,90	132,76	62,78	280,73	31,58	133,39	68,96	258,00	32,20	131,63	30,46

N	Casa	Consumo medido CAJ (l/p/d)	Regressão linear múltipla				Regressão robusta				Regressão geográfica	
			Consumo previsto (l/p/d)	Intervalo de predição (95%)		Erro absoluto (%)	Consumo previsto (l/p/d)	Intervalo de predição (95%)		Erro absoluto (%)	Consumo previsto (l/p/d)	Erro absoluto (%)
68	209	112,11	97,92	46,23	207,42	12,65	100,39	51,95	194,02	10,45	97,83	12,73
69	210	101,17	129,27	61,16	273,21	27,78	133,03	68,90	256,88	31,50	128,38	26,90
70	211	127,16	150,10	71,31	315,95	18,05	148,17	77,28	284,10	16,52	149,02	17,19
71	216	134,65	150,89	71,68	317,63	12,06	149,04	77,73	285,79	10,69	150,31	11,63
72	220	231,98	144,19	68,39	304,01	37,84	145,53	75,88	279,10	37,27	144,28	37,81
73	224	112,77	149,40	70,32	317,45	32,48	152,13	78,77	293,84	34,90	144,87	28,46
74	226	105,24	141,31	67,07	297,72	34,28	142,23	74,19	272,66	35,16	142,78	35,68
75	233	159,80	121,09	57,32	255,79	24,23	120,07	62,41	231,00	24,86	122,33	23,45
76	234	140,32	111,76	52,77	236,72	20,35	115,21	59,83	221,86	17,89	115,18	17,92
77	235	165,42	168,72	79,76	356,88	1,99	169,15	87,94	325,38	2,26	168,00	1,56
78	236	88,44	111,83	52,95	236,17	26,45	116,04	60,23	223,58	31,21	113,48	28,31
79	237	119,43	122,87	58,05	260,04	2,88	122,18	63,12	236,48	2,30	119,63	0,16
80	243	123,81	127,21	60,43	267,81	2,75	126,26	65,83	242,16	1,98	129,30	4,44
81	246	118,08	107,25	50,30	228,71	9,17	113,03	58,38	218,86	4,28	109,82	7,00
82	247	86,37	157,41	74,69	331,77	82,26	156,37	81,50	300,02	81,05	156,70	81,43
83	253	161,27	111,03	52,57	234,46	31,16	115,10	59,74	221,74	28,63	109,78	31,93
84	255	164,44	129,70	61,34	274,25	21,13	129,90	67,16	251,26	21,00	126,81	22,88
85	259	135,73	130,14	60,92	278,00	4,12	129,37	66,25	252,64	4,68	128,80	5,11
86	260	50,94	122,99	58,31	259,45	141,46	124,93	65,10	239,75	145,26	125,99	147,35
87	262	140,93	130,46	61,70	275,86	7,43	130,77	67,61	252,93	7,21	130,29	7,55
88	269	217,03	115,56	54,71	244,08	46,76	113,87	59,19	219,09	47,53	113,72	47,60
89	270	172,28	132,66	62,77	280,37	23,00	136,11	70,86	261,45	21,00	132,79	22,93
90	272	134,44	141,45	66,59	300,49	5,21	142,99	74,02	276,23	6,36	142,98	6,35

N	Casa	Consumo medido CAJ (l/p/d)	Regressão linear múltipla				Regressão robusta				Regressão geográfica	
			Consumo previsto (l/p/d)	Intervalo de predição (95%)		Erro absoluto (%)	Consumo previsto (l/p/d)	Intervalo de predição (95%)		Erro absoluto (%)	Consumo previsto (l/p/d)	Erro absoluto (%)
91	273	153,33	147,52	69,90	311,33	3,79	149,34	77,82	286,59	2,60	148,82	2,94
92	277	128,23	122,82	58,17	259,32	4,22	125,54	65,09	242,15	2,10	121,03	5,61
93	281	217,29	133,62	63,43	281,49	38,50	129,87	67,59	249,55	40,23	132,06	39,22
94	296	92,74	98,64	46,51	209,19	6,36	98,47	50,91	190,48	6,18	99,68	7,48
95	298	150,99	145,86	69,32	306,93	3,39	143,43	74,80	275,03	5,01	146,81	2,77
96	299	226,83	95,30	44,98	201,91	57,98	97,35	50,38	188,12	57,08	97,26	57,12
97	301	128,78	118,48	56,12	250,12	8,00	120,52	62,49	232,42	6,41	116,51	9,53
98	306	217,23	118,09	55,41	251,68	45,64	114,28	58,69	222,50	47,39	117,50	45,91
99	312	115,29	121,67	57,73	256,43	5,54	120,05	62,50	230,60	4,13	118,95	3,17
100	318	109,29	106,69	50,51	225,34	2,38	110,02	57,10	211,96	0,67	105,06	3,86
101	324	217,16	125,63	59,26	266,31	42,15	132,40	68,45	256,08	39,03	126,94	41,55
102	327	225,24	137,69	65,09	291,27	38,87	139,01	71,83	269,02	38,29	136,36	39,46
103	328	189,24	111,82	52,83	236,69	40,91	109,10	56,52	210,60	42,35	112,98	40,30
104	332	327,10	190,65	89,35	406,82	41,71	189,79	97,96	367,69	41,98	189,89	41,95
105	337	184,25	122,32	57,94	258,26	33,61	124,96	64,79	241,02	32,18	120,59	34,55
106	339	268,69	158,97	75,06	336,71	40,83	149,64	77,33	289,60	44,31	154,92	42,34
107	348	94,87	131,38	62,06	278,13	38,47	124,63	64,60	240,46	31,37	130,60	37,65
108	351	163,90	161,12	76,44	339,62	1,69	156,19	81,26	300,20	4,71	158,57	3,25
109	352	320,39	186,31	87,97	394,60	41,85	189,00	97,92	364,81	41,01	187,61	41,44
110	353	125,65	168,54	79,89	355,57	34,14	164,36	85,49	316,00	30,81	166,46	32,48
111	362	150,16	124,00	58,87	261,19	17,42	122,66	63,91	235,43	18,31	122,04	18,73
112	364	121,45	153,39	72,42	324,87	26,29	148,55	77,01	286,55	22,31	152,18	25,30
113	367	204,15	113,43	53,69	239,62	44,44	114,71	59,46	221,33	43,81	114,85	43,74

			Regressão linear múltipla				Regressão robusta				Regressão geográfica	
N	Casa	Consumo medido CAJ (l/p/d)	Consumo previsto (l/p/d)	Intervalo de predição (95%)		Erro absoluto (%)	Consumo previsto (l/p/d)	Intervalo de predição (95%)		Erro absoluto (%)	Consumo previsto (l/p/d)	Erro absoluto (%)
114	370	91,71	119,59	56,57	252,79	30,39	115,17	59,79	221,82	25,57	117,74	28,37
115	375	99,88	104,17	49,14	220,80	4,29	107,68	55,64	208,38	7,80	104,93	5,05
116	377	220,72	140,35	66,40	296,64	36,41	133,57	69,20	257,81	39,48	142,27	35,54
117	378	114,53	152,48	71,79	323,87	33,14	151,68	78,47	293,20	32,44	151,25	32,06
118	380	117,79	132,06	62,09	280,86	12,11	133,33	68,54	259,39	13,20	129,99	10,36
119	383	183,23	165,60	78,18	350,79	9,62	165,38	85,51	319,83	9,74	165,55	9,64
120	395	174,56	111,93	53,00	236,36	35,88	112,27	58,38	215,89	35,69	115,55	33,81