

## MODELO DE REGRESSÃO LINEAR MÚLTIPLA APLICADO À PREVISIBILIDADE DE DESEMPENHO EM HIDRÔMETROS

### **Felipe Vieira de Luca<sup>(1)</sup>**

Eng<sup>o</sup> Sanitarista e Ambiental – UFSC. Pós-Graduação em Gestão Pública – UFSC. Pós-Graduação Programa de Desenvolvimento de Dirigentes – Fundação Dom Cabral. Mestrando em Eng. Civil – UDESC. Coordenador de Micromedição e Faturamento, Eng<sup>o</sup> Sanitarista na Companhia Águas de Joinville – Santa Catarina, atuação em controle de perdas, micromedição, IoT e faturamento. <https://www.linkedin.com/in/felipe-de-luca-595b7aab/>

### **Hygor Bello da Silva<sup>(2)</sup>**

Graduando em Engenharia Civil – UDESC. Estagiário, atuação em micromedição e análise de dados, Companhia Águas de Joinville – Santa Catarina.

### **Edinara Fernanda Werner<sup>(3)</sup>**

Técnica em Saneamento – Centro Federal de Educacional Tecnológica de Santa Catarina. Técnica em Saneamento, atuação em controle de perdas, micromedição, IoT – Companhia Águas de Joinville – Santa Catarina. Graduação em Pedagogia – UFSC. Graduanda em Serviço Social – Faculdade Uniasselvi.

### **Edson Schwitzky<sup>(4)</sup>**

Graduação em Administração de Empresas – UNIVILLE. Pós-Graduação em Administração Pública e Gerenciamento de Cidades – UNINTER. Técnico em Saneamento – Associação Educacional e Tecnológica de Santa Catarina Assessoritec. Técnico em Saneamento, atuação em micromedição – Companhia Águas de Joinville – Santa Catarina.

### **Emilly Vitor Fritzen<sup>(5)</sup>**

Gerente de Faturamento e Captação de Clientes, Companhia Águas de Joinville – Santa Catarina. Pós-Graduação em Gerenciamento de Projetos – Anhanguera. Engenheira de Alimentos – UFSC.

**Endereço<sup>(1)</sup>:** Rua XV de Novembro, 3.950 - Glória - Joinville - Santa Catarina - CEP: 89216-202 - Brasil - Tel: +55 (47) 99286-0005 - e-mail: [felipe.luca.81@gmail.com](mailto:felipe.luca.81@gmail.com).

## RESUMO

Este trabalho tem como objetivo estabelecer um modelo de previsão para o Índice de Desempenho Metrológico em hidrômetros. Foram realizados ensaios laboratoriais durante 4 meses em diversos medidores retirados de campo, seguindo uma amostragem estratificada a partir do inventário do parque de medidores da cidade de Joinville. A escolha das variáveis preditoras foi submetida aos critérios da *Akaike Information Criterion* (AIC), com a utilização do método *stepwise*, bem como as adequações do modelo foram verificadas a partir das análises dos resíduos. Nas análises dos resíduos foram investigadas as suposições de normalidade, homocedasticidade, independência, linearidade e multicolinearidade, bem como a presença de outliers e pontos influentes. Por fim foi obtido um modelo de regressão linear múltipla que envolve a presença de variáveis quantitativas e qualitativas para descrever como ocorre a perda de desempenho em medidores de água do tipo taquimétrico, classe metrológica B. O modelo proposto explicou as variações em até 75,6% dos dados de laboratório, e quando aplicados à dados de validação, de modo a estimar a estabilidade do modelo, obteve-se o desempenho de 76%. A partir dos dados disponíveis não foi possível formular uma equação que envolvesse os tipos volumétrico e ultrassônico.

**PALAVRAS-CHAVE:** Regressão Linear Múltipla. Desempenho de Hidrômetros. Testes Estatísticos.

## INTRODUÇÃO

O foco deste trabalho são as perdas aparentes de água, conceituadas pela *International Water Association* (THORNTON, 2008) como “a perda comercial ou não-física, sendo o volume de água produzido pela companhia, consumida pelo cliente, mas não contabilizada devido a erros de medição dos hidrômetros, ligações clandestinas, violação dos hidrômetros e falhas no cadastro comercial”. Especificamente tratar-se-á das perdas aparentes ocasionadas por submedição, ou seja, erros de medição dos hidrômetros.

A submedição ocorre quando o hidrômetro não é capaz de registrar os fluxos de água e, por conseguinte, registros de volume consumidos, seja por problemas de instalação, erros de dimensionamento, desgastes devido a utilização, dentre outros fatores.

A submedição impacta diretamente o índice de perdas da concessionária de saneamento pelo fato de permitir a passagem de água até o consumidor final sem que a mesma seja contabilizada, comprometendo, portanto, o balanço hídrico e o faturamento da concessionária, uma vez que a receita de uma empresa desta natureza é proveniente da somatória do volume registrado em unidades consumidoras.

Segundo Tardelli Filho (2004), hidrômetros apresentam imprecisão natural que varia com os tipos de medidores disponíveis no mercado. Esta imprecisão pode ser incrementada devido a fatores tais como: instalação inadequada, descalibração do medidor, dimensionamento inadequado, operando com baixas velocidades, grande amplitude entre vazões máximas e mínimas, problemas na transmissão dos dados quando se utiliza telemetria, erros de paralaxe, dentre outros.

Como todos os instrumentos de medição, os hidrômetros possuem limites em relação a seus usos, sendo uns mais versáteis e outros menos. Apresentam imprecisão intrínseca ao equipamento. Este é um fato, e apresenta lógica coerente aos conceitos da área de metrologia: o instrumento de medição possui precisão dentro de um campo de aplicação considerando a incerteza de medição associada ao processo.

Neste trabalho tratar-se-á da imprecisão “de fábrica”, ou seja, da submedição que se encontra presente em função do projeto do medidor, de seu respectivo processo de fabricação pelas diferentes marcas e ainda da submedição proveniente de seu uso e da manutenção deficitária do parque de medidores.

## **OBJETIVOS**

O objetivo geral deste artigo é a determinação de um modelo de regressão linear múltipla aplicado à previsibilidade do Índice de Desempenho Metroológico (IDM) em hidrômetros. E os objetivos específicos são: efetuar a análise exploratória dos dados obtidos em laboratório para delimitar a amostra inicial do tratamento estatístico, estabelecer um modelo preditivo para o IDM em hidrômetros, verificar o modelo encontrado quanto aos métodos das análises dos resíduos e verificar a estabilidade de previsão em um conjunto de dados de validação.

## **METODOLOGIA**

Os cálculos para endossamento deste estudo foram realizados com dados da Companhia de Saneamento Básico Águas de Joinville. A Águas de Joinville é uma empresa pública do Município de Joinville, cuja finalidade é “explorar diretamente os serviços públicos de abastecimento de água e de esgotamento sanitário, compreendendo a captação de água bruta, o tratamento, a adução, a reservação e a distribuição para consumo público e a coleta de esgotos sanitários trazidos por meio de tubos e condutos, o transporte, o tratamento, o reaproveitamento e a disposição final, bem como outras soluções alternativas”, conforme Lei nº 5054/2004 e Estatuto Social da Companhia Águas de Joinville.

Inicialmente foi realizado um inventário do parque de medidores com o intuito de caracterizar o universo da pesquisa. Foram apurados no inventário o quantitativo e percentual de matrículas para os diferentes tipos de hidrômetros que constam no parque, bem como o que cada um dos respectivos tipos representa em termos de volume micromedido. Após isso, no inventário, foi apurado o ano de instalação dos medidores o que permite apurar a idade média do parque. Foi traçado posteriormente o histograma de ligações, volume micromedido e faturamento dentro de faixas de consumo. O inventário de medidores permitiu verificar como está a distribuição do parque em termos quantitativos e como o mesmo está composto qualitativamente por tipos, faixas de consumo, capacidades, dentre outras características, e, como o mesmo está constituído segundo diversas combinações destas características.

Vale ressaltar que com a metodologia supracitada a técnica de amostragem foi a probabilística estratificada, de modo que com a representatividade das variáveis analisadas pode-se selecionar as quantidades de hidrômetros para análise. Por exemplo a marca com maior representatividade no parque teve maior representatividade na população amostral.

Subsequentemente os hidrômetros foram ensaiados utilizando-se a metodologia presente na ABNT NBR 15.538 de 2014. Esta norma especifica os procedimentos para avaliação da eficiência em medidores de água fria em condições específicas e considerando um sistema de abastecimento misto. A avaliação dos medidores é quantificada por meio do Índice de Desempenho Metrológico (IDM), e trata-se de um valor numérico percentual que corresponde ao desempenho dos hidrômetros, calculado pela equação 1.

$$\text{IDM} = 100 + \text{EP} \quad \text{equação (1)}$$

Onde “EP” significa o erro ponderado, a parcela de água não registrada pelo medidor, um parâmetro de avaliação de desempenho obtido pela associação entre o perfil de consumo típico descrito na NBR 15.538 e o erro relativo apresentado pelo medidor de água, em faixas de vazões previamente definidas, representado pela equação 2. Os pesos e vazões podem ser consultados na tabela 1.

$$\text{EP} = \{ \sum [(\text{erro } Q_x) * (\text{peso } Q_x)] \} / 100 \quad \text{equação (2)}$$

$Q_x$  Trata-se das vazões do ensaio, a tabela a seguir representa as 10 faixas de vazões e seus respectivos pesos:

**Tabela 1: Vazões de Ensaio em hidrômetros – Adaptado ABNT NBR 15.538:2014.**

FAIXAS DE VAZÃO (L/h)	VAZÕES DE ENSAIO (L/h)	PERFIL DE CONSUMO (%)
0 a 5	2.5	4.56
5 a 15	10	6.99
15 a 30	22.5	6.83
30 a 50	40	7.34
50 a 150	100	23.21
150 a 350	250	23.92
350 a 550	450	12.27
550 a 850	700	7.29
850 a 1150	1000	5.86
1150 a 1500	1325	1.73

Os ensaios foram realizados no Laboratório de Hidrometria da Companhia Águas de Joinville (LabHidro), um centro dedicado a estudos do parque de medidores de Joinville. Atualmente a bancada volumétrica do LabHidro está validada pelo INMETRO. A figura 1 representa a linha de montagem para os ensaios de IDM.



**Figura 1: Bancada volumétrica para ensaios em hidrômetros da Companhia Águas de Joinville.**

A compilação dos dados de laboratório foi realizada em planilhas de EXCEL. E para o tratamento estatístico dos dados, com a finalidade de sintetizar e organizar as variáveis, foi realizada a análise exploratória usando métodos descritos e gráficos com auxílio da linguagem de programação R por meio do *software* RSTUDIO (R CORE TEAM, 2017).

Neste trabalho propõe-se analisar a correlação e regressão entre as variáveis que podem influenciar no decaimento do rendimento de hidrômetros. Desta forma, diversas características foram consideradas para estudo, como tipo de hidrômetro, marca, volumetria registrada e mediana do consumo dos últimos 12 meses.

A análise exploratória dos dados iniciou-se testando a normalidade dos valores de IDM, foram consideradas possíveis transformações matemáticas no banco de dados com o intuito de evitar problemas futuros em relação a variação não-constante e não-normalidade dos erros (TORMAN; COSTER; RIBOLDI, 2012). E a análise gráfica como ferramenta auxiliadora na exploração dos dados foi realizada com o auxílio de pacotes ggplot2 (WICKHAM, 2009).

Subsequentemente foi analisado o coeficiente de correlação. E são considerados significativos os coeficientes com p-valor menor que 0,05 (FARCOMENI, 2008).

Um modelo de regressão linear múltipla foi desenvolvido de modo a prever o valor de IDM em hidrômetros. A escolha das variáveis regressoras foi atrelada a uma análise dentre as quais que possuíam maior significância no modelo.

Para auxiliar na decisão das variáveis independentes do modelo que possuíam maior significância foi aplicado o método *Stepwise*. O método consiste em realizar a regressão por etapas e apoia-se em um algoritmo computacional que iterativamente testa uma série de modelos. Em cada iteração é considerado o critério de *Akaike Information Criterion* ((AIC), finalizando quando todas as variáveis apresentarem significância estatística, ou seja, p-valor menor que 0,05.

Com o modelo obtido foram avaliadas a normalidade, multicolinearidade, independência, linearidade e homocedasticidade dos resíduos para validar o modelo de regressão linear múltipla. A linearidade foi analisada com base no gráfico de Valores Residuais x Valores Ajustados, já a normalidade foi examinada com base que os valores de resíduos devem estar distribuídos com média igual a 0. E por fim com o pacote gvlma (PEÑA; SLATE, 2012), pode-se verificar a validação geral do modelo linear.

A validação *in loco* foi realizada com o cálculo de IDM de hidrômetros ensaiados no LabHidro e que não foram inseridos no banco de dados para a formulação do modelo de regressão. Deste modo pode-se calcular a eficiência do modelo em campo, que é uma alternativa para evitar modelos superajustados aos dados. E mais, as métricas do pacote *metrics* auxiliaram na determinação das incertezas associadas ao modelo.

## RESULTADOS OBTIDOS

Inicialmente foi realizado gráficos de distribuição de modo a entender como a variável dependente se comportava. A saber os gráficos de distribuição são: o histograma na figura 2 e o BoxPlot na figura 3. A partir do histograma pode-se observa que os dados não apresentam uma distribuição normal, e sim trimodal. A hipótese de não normalidade dos dados foi confirmada quando aplicado o teste de *Shapiro-Wilk*. O valor obtido para o conjunto de 306 amostras foi menor que o nível de significância de 5%, indicando que os dados não apresentam uma distribuição normal.

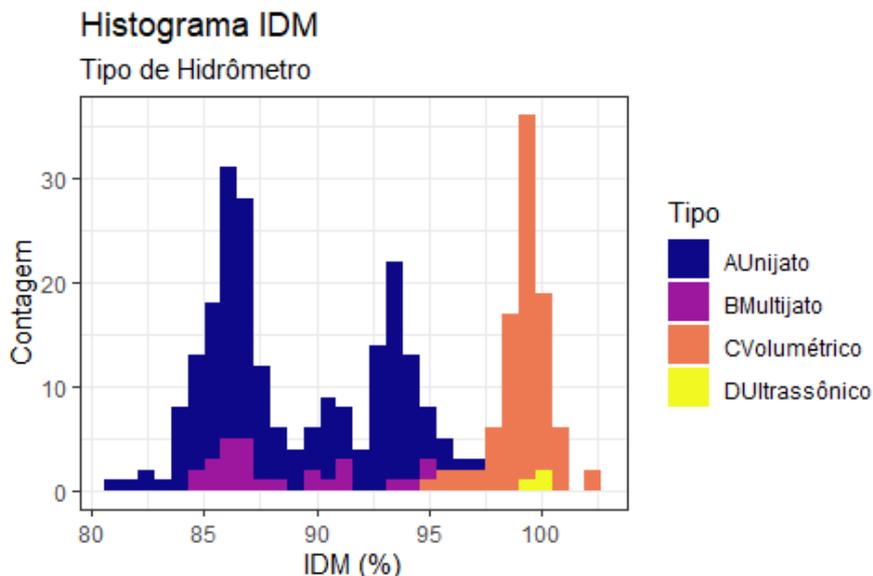


Figura 2: Histograma dos valores de IDM – todos os tipos.

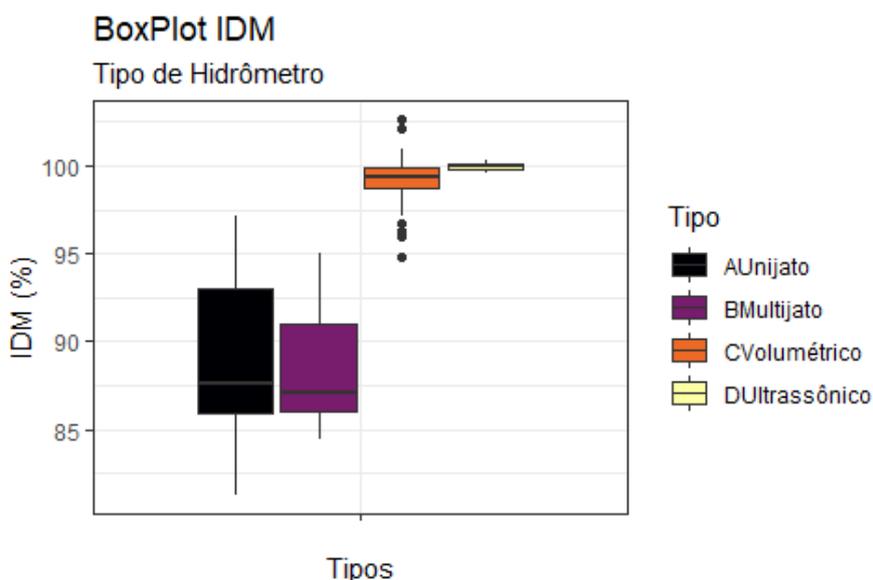


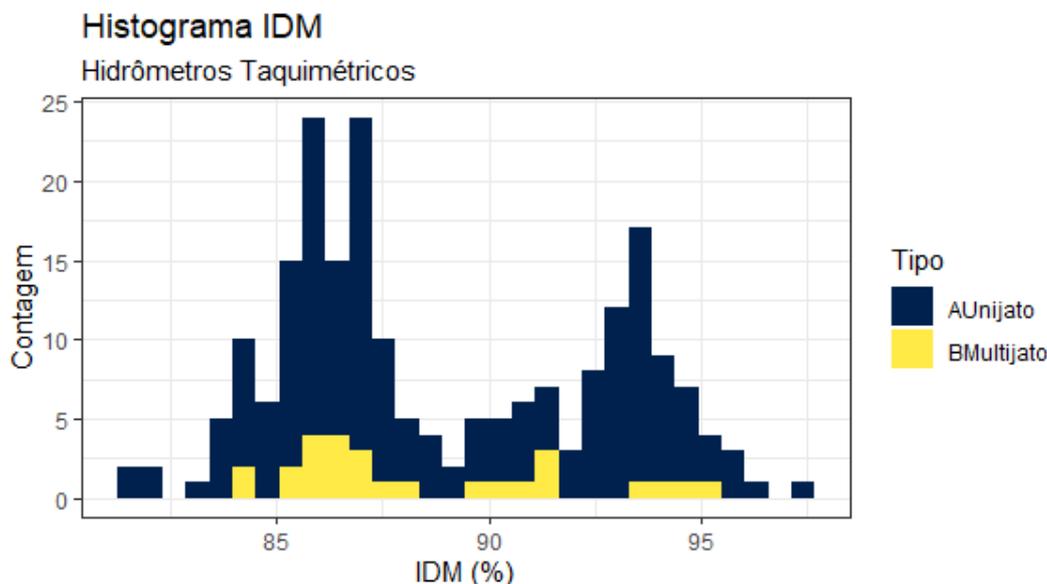
Figura 3: BoxPlot dos valores de IDM – todos os tipos.

Analisando o gráfico da figura 2 as distribuições dos tipos Unijato e Multijato são semelhantes, assim como, também são semelhantes as distribuições dos tipos volumétrico e ultrasônicos. O teste estatístico que permite verificar se há diferença entre grupos é o teste de *Kruskal-Wallis*, que retornou o valor menor que 0,05, indicando que sim, há diferença estatística em pelo menos uma das populações. De modo a determinar quais dos tipos se diferenciavam um dos outros foi realizado o teste *post-hoc Kruskal-Nemenyi*. Quando a relação entre as marcas for menor que 0,05, ou 5%, pode-se afirmar estatisticamente que há diferença entre as marcas. A tabela 2 apresenta o resumo dos valores obtidos:

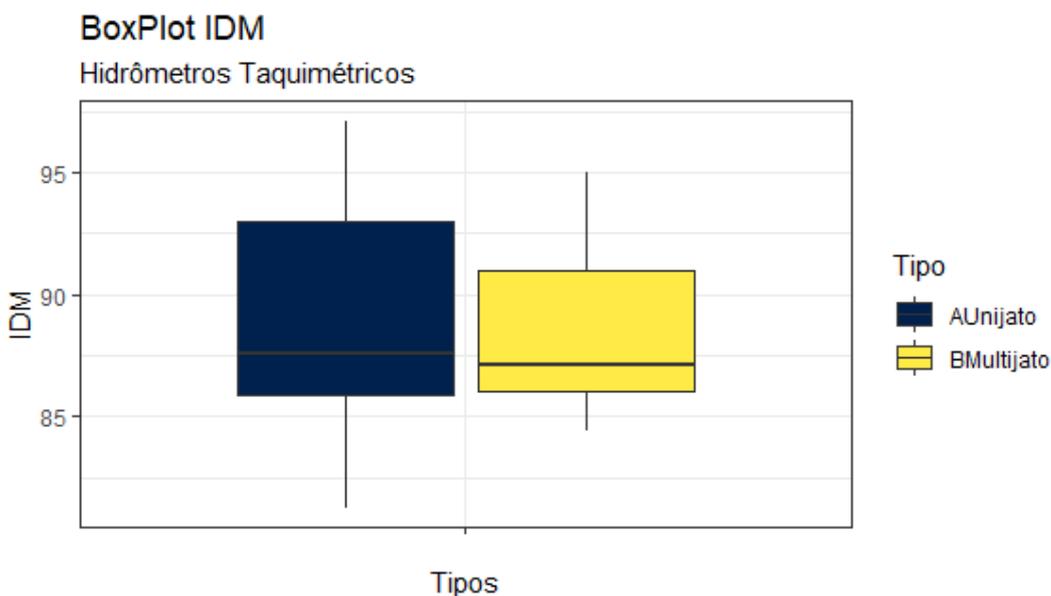
Tabela 2: Resumo dos valores do teste *Post-hoc Kruskal-Nemenyi* – todos os tipos.

	UNIJATO	MULTIJATO	VOLUMÉTRICO
MULTIJATO	0,9844	-----	-----
VOLUMÉTRICO	< 2e-16	5,2e-14	-----
ULTRASSÔNICO	0,0039	0,0043	0,9700

A partir dos resultados obtidos no teste *post-hoc* *Kruskal-Nemenyi* optou-se por trabalhar separadamente os tipos unijato e multijato em um grupo, e volumétrico com ultrassônico em outro grupo. Iniciando as análises para os conjuntos de dados que continham os tipos unijato e multijato. As análises tiveram como início a elaboração de gráficos de distribuição para os tipos unijato e multijato, nas figuras 4 e 5.



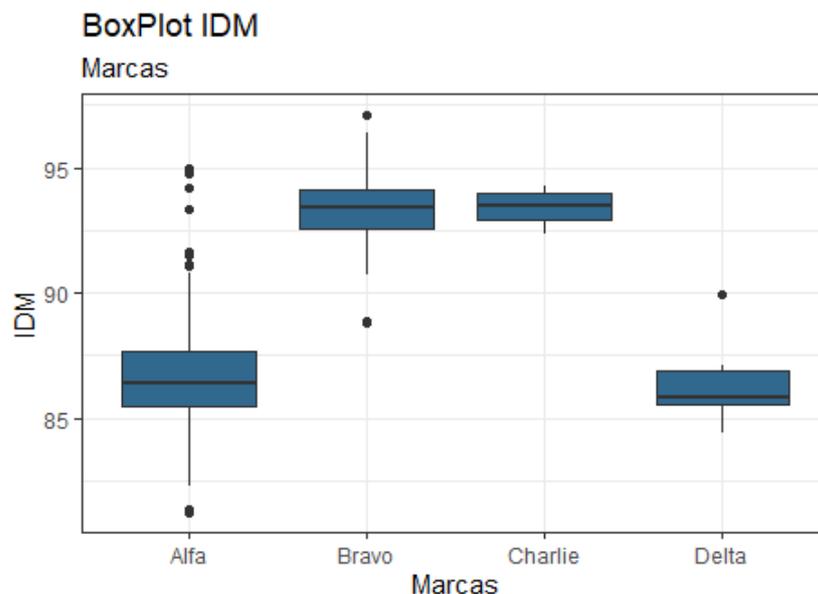
**Figura 4: Histograma dos valores de IDM – tipos unijato e multijato.**



**Figura 5: BoxPlot dos valores de IDM – tipos unijato e multijato.**

Por meio dos gráficos de dispersão BoxPlot da figura 5 e do Histograma da figura 4 pode observar que aparentemente não havia diferença na distribuição para os tipos. No gráfico BoxPlot o tamanho da “caixa”, bem como o quartil inferior e mediana são bem próximos, diferenciando apenas na dispersão total dos dados. A hipótese visual estava correta, o teste de *Mann-Whitney* confirmou a premissa, para um nível de significância a 5%. O valor obtido no teste foi de 0,61, deste modo rejeita-se a hipótese que as os tipos se diferem.

Após essas constatações foi aplicado uma análise semelhante, mas com o viés das marcas de hidrômetros taquimétricos que compõem o parque de hidrômetros da Companhia Águas de Joinville. O gráfico da figura 4 apresenta as distribuições de IDM por Marca:



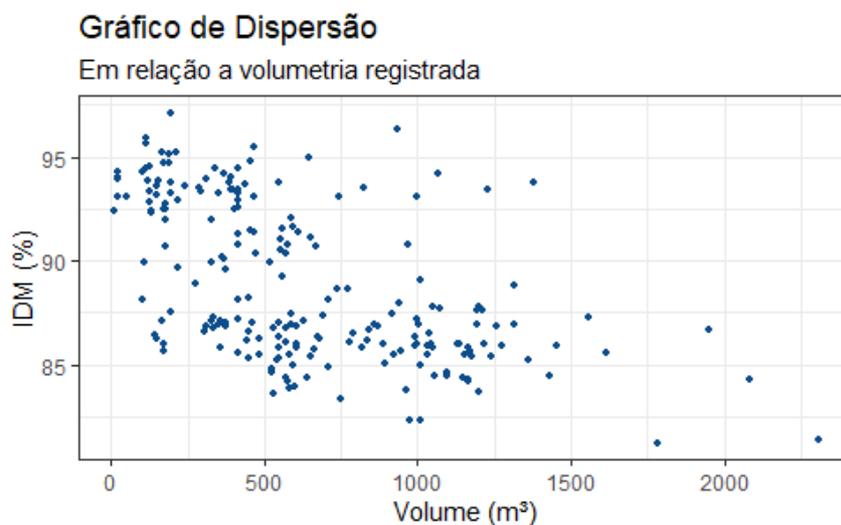
**Figura 6: BoxPlot dos valores de IDM por Marca – tipos unijato e multijato.**

Com uma análise visual percebeu-se que há diferença entre as marcas. Para confirmar essa hipótese foi aplicado o teste de *Kruskal-Wallis*. O teste retornou um valor menor que o nível de significância de 0,05. Para determinar quais marcas se diferem de quais marcas estatisticamente utilizou-se o teste *post-hoc Kruskal-Nemenyi*. A tabela 3 apresenta os resultados para esse teste.

**Tabela 3: Resumo dos valores do teste *Post-hoc Kruskal-Nemenyi*.**

	ALFA	BRAVO	CHARLIE
BRAVO	3,9e-14	-----	-----
CHARLIE	0,00051	1,0000	-----
DELTA	0,9585	8,6e-06	0,00424

As variáveis quantitativas estudadas como possíveis influenciadoras na perda de desempenho foram a mediana de consumo dos últimos 12 meses e a volumetria total registrada nos hidrômetros. Os gráficos das figuras 5 e 6 apresentam os gráficos para se detectar visualmente se há alguma hipótese a se testar acerca da influência dessas variáveis na perda de desempenho dos hidrômetros taquimétricos, classe metrológica B:



**Figura 7: Gráfico de Dispersão IDM x Volumetria.**



### Gráfico de Dispersão

Em relação a mediana do consumo

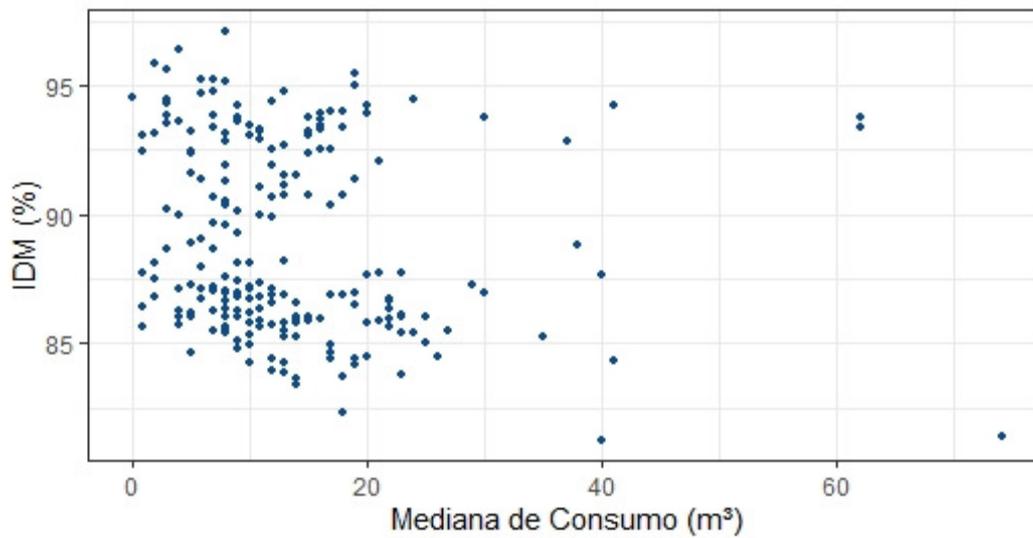


Figura 8: Gráfico de Dispersão IDM x Mediana de Consumo.

Foi observado, também, essa mesma relação entre o IDM e as variáveis volumetria e mediana de consumo, mas figuras 9 e 10, respectivamente, com o viés da diferenciação entre as marcas.

### Gráfico de Dispersão

Em relação a volumetria registrada

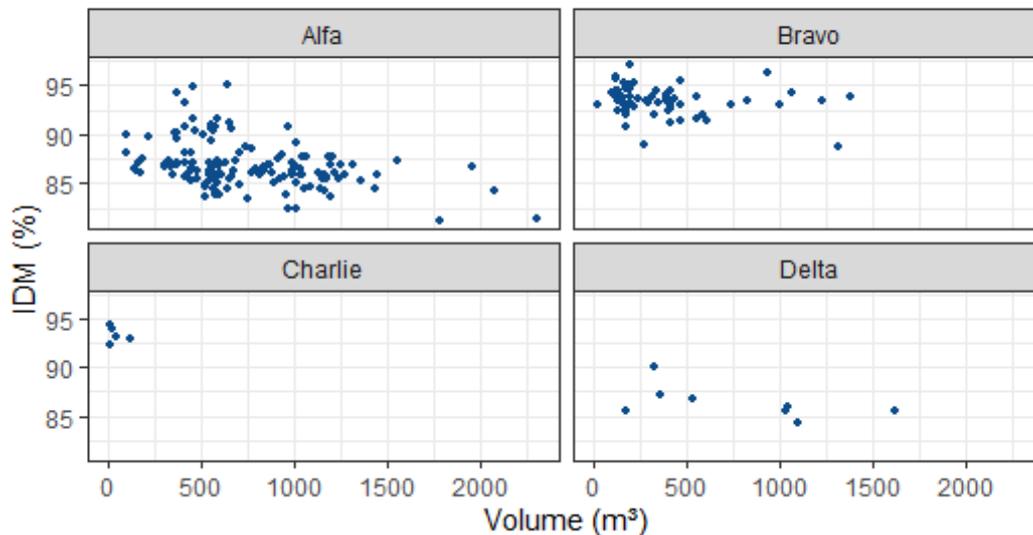
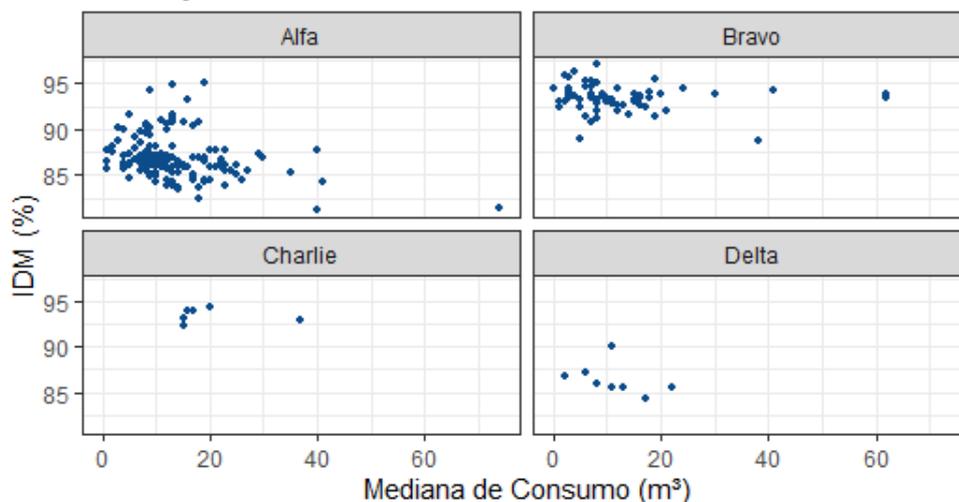


Figura 9: Gráfico de Dispersão IDM x Volumetria - com o viés das Marcas.

### Gráfico de Dispersão

Em relação a mediana do consumo



**Figura 10: Gráfico de Dispersão IDM x Mediana de Consumo - com o viés das Marcas.**

Observou-se que existe uma correlação entre o IDM x Volumetria e IDM x Mediana de Consumo. Para Quantificar esse valor e confirma as hipóteses foi utilizado o teste de correlação não paramétrico de *Spearman*, a utilização deste teste deve-se ao fato de que o IDM não apresenta uma distribuição normal, como supracitado. O resultado obtido pode ser visualizado na tabela 4.

**Tabela 4: Resultado das correlações de *Spearman*.**

RELAÇÃO	RESULTADO	TIPO DE CORRELAÇÃO
IDM x Volumetria	-0,64	Moderada
IDM x Mediana de Consumo	-0,23	Fraca

O resultado do teste de *Spearman* para a correlação entre as variáveis IDM x Volumetria foi de -0,64, indicando que há uma correlação negativa moderada e muito próxima de uma correlação forte, este quando ultrapassa o valor de -0,70. Já o resultado para a relação entre IDM x Mediana de Consumo foi de -0,23, uma correlação negativa fraca.

Após as análises entre as relações da variável dependente de estudo formulou-se o modelo de regressão linear múltipla envolvendo a variável dependente IDM e as variáveis explicativas: Tipo, Marca e Volumetria. A expressão é:

$$\text{IDM} = \text{Intercepto} + \text{Tipo} + \text{Marca} + \text{Volumetria} + \text{Mediana de Consumo} \quad \text{equação (3)}$$

O coeficiente de correlação do modelo atingiu um valor de  $R^2 = 75,7\%$ . E os coeficientes da equação, bem como o p-valor pode ser consultado na tabela 5.

**Tabela 5: Coeficientes do Modelo de Regressão Linear Múltipla**

VARIÁVEL	COEFICIENTE	P-VALOR
Intercepto	87,96	< 2e-16
Tipo Unijato	0	-----
Tipo Multijato	2,81	3,19 e-09
Marca Alfa	0	-----
Marca Bravo	6,20	< 2e-16
Marca Charlie	5,61	1,57 e-09
Marca Delta	-2,86	0,000298
Volumetria	-0,001928	0,000104
Mediana de Consumo	-0,003271	0,85634

Aplicou-se, então, o método *stepwise* de seleção de variáveis com significância estatística para o modelo. E o resultado foi um modelo de seguinte expressão:

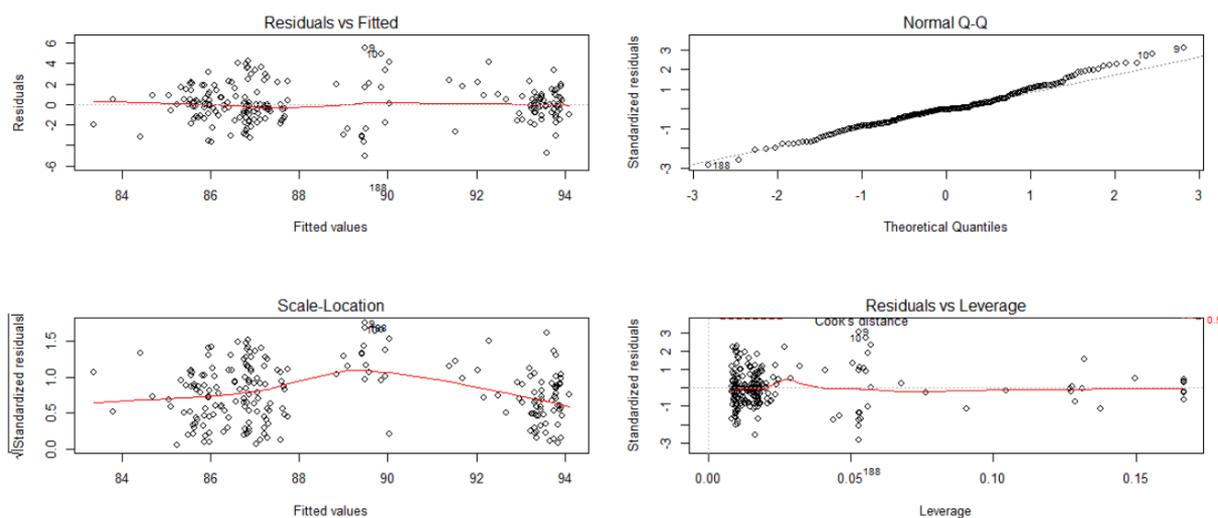
$$\text{IDM} = \text{Intercepto} + \text{Tipo} + \text{Marca} + \text{Volumetria} \quad \text{equação (4)}$$

O coeficiente de correlação do modelo atingiu um valor de  $R^2 = 75,9\%$ . E os coeficientes da equação, bem como o p-valor pode ser consultado na tabela 6.

**Tabela 6: Coeficientes do Modelo de Regressão Linear Múltipla**

VARIÁVEL	COEFICIENTE	P-VALOR
Intercepto	87,96	< 2e-16
Tipo Unijato	0	-----
Tipo Multijato	2,81	3,19 e-09
Marca Alfa	0	-----
Marca Bravo	6,18	< 2e-16
Marca Charlie	5,54	7,14 e-11
Marca Delta	-2,85	0,000289
Volumetria	-0,001991	1,99 e-08

O diagnóstico final do modelo pode ser consultado nos gráficos da figura 9 e a avaliação do modelo na tabela 7.



**Figura 11: Diagnóstico do Modelo.**

**Tabela 7: Análise Global do Modelo de Regressão Linear Múltipla**

GVLMA TEST	P-VALOR	DECISÃO
Global	0,21	Aceito
Assimetria	0,13	Aceito
Curtose	0,29	Aceito
Linearidade	0,23	Aceito
Heterocedasticidade	0,32	Aceito

O modelo atendeu todas as suposições necessárias a um modelo de regressão linear. O NVC reforçou o diagrama *Residuals vs Fitted*, ao apresentar o p-valor de 0,46. E o teste de *Shapiro-Wilk* para a normalidade dos resíduos retornou o valor de 0,07 ou seja pode-se confirmar que há normalidade.

As métricas de erro do modelo, a partir do pacote *metrics* podem ser consultadas na tabela.

**Tabela 8: Métricas de Erro do Modelo de Regressão Linear.**

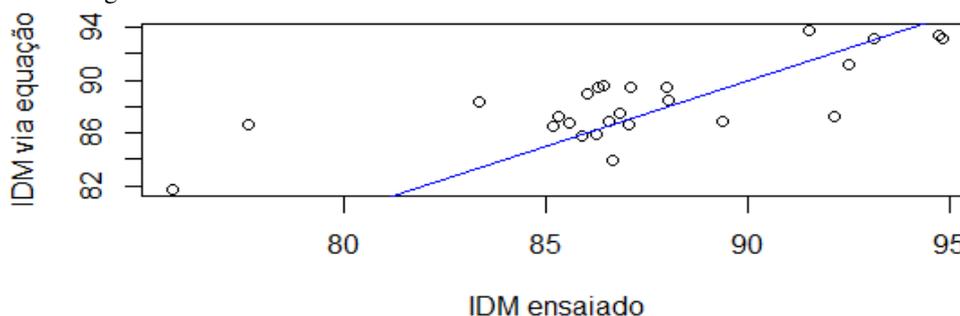
MÉTRICA	VALOR
MAPE	0,015 %
MAE	1,36 %
RMSE	1,81 %

Finalizado a análise dos resíduos aplicou-se o modelo a 26 novos dados de IDM de hidrômetros taquimétricos ensaiados após as análises e determinação do modelo, de modo a validar o modelo. O resultado da correlação entre os valores obtidos em laboratório para as novas amostras e os valores obtidos via equação foi de 76,3 %. As métricas de erro podem ser consultadas na tabela 9.

**Tabela 9: Métricas de Erro do Modelo de Regressão Linear aplicado aos dados de validação.**

MÉTRICA	VALOR
MAPE	0,026 %
MAE	2,23 %
RMSE	3,01 %

O gráfico da distribuição dos dados de IDM de campo comparado ao IDM obtido pela equação pode ser consultado na figura 10.



**Figura 12: Diagnóstico do Modelo.**

Subsequentemente realizou-se os mesmos tratamentos supracitados para os tipos volumétrico e ultrassônico, entretanto o modelo de regressão não apresentou validade estatística para utilização. Deste modo, optou-se por apresentar os coeficientes do modelo de regressão linear múltipla para esse conjunto bem como a análise geral do modelo pelo teste *Global Validation of Linear Model Assumptions* (gvlma), respectivamente nas tabelas 10 e 11.

**Tabela 10: Coeficientes do Modelo de Regressão Linear Múltipla – volumétrico e ultrassônico.**

VARIÁVEL	COEFICIENTE	P-VALOR
Intercepto	99,16	< 2e-16
Tipo Volumétrico	0	-----
Tipo Ultrassônico	1,21	0,4350
Marca Alfa	0	-----
Marca Bravo	0,13	0,6541
Marca Echo	1,16	0,0248
Volume	-0,000152	0,2597

**Tabela 11: Análise Global do Modelo de Regressão Linear Múltipla**

GVLMA TEST	P-VALOR	DECISÃO
Global	1,958e-09	Não Aceito
Assimetria	1,697e-03	Não Aceito
Curtose	4,634e-08	Não Aceito
Linearidade	0,0602	Aceito
Heterocedasticidade	0,072	Aceito

Os gráficos de distribuição para os tipos volumétrico e ultrassônicos estão representados no histograma da figura 13 e no BoxPlot da figura 14.

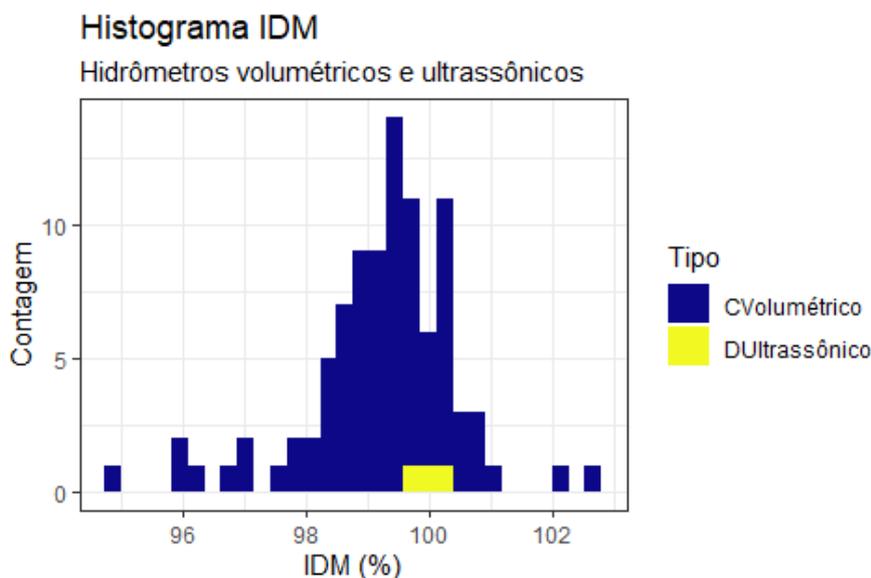


Figura 13: Histograma dos valores de IDM.

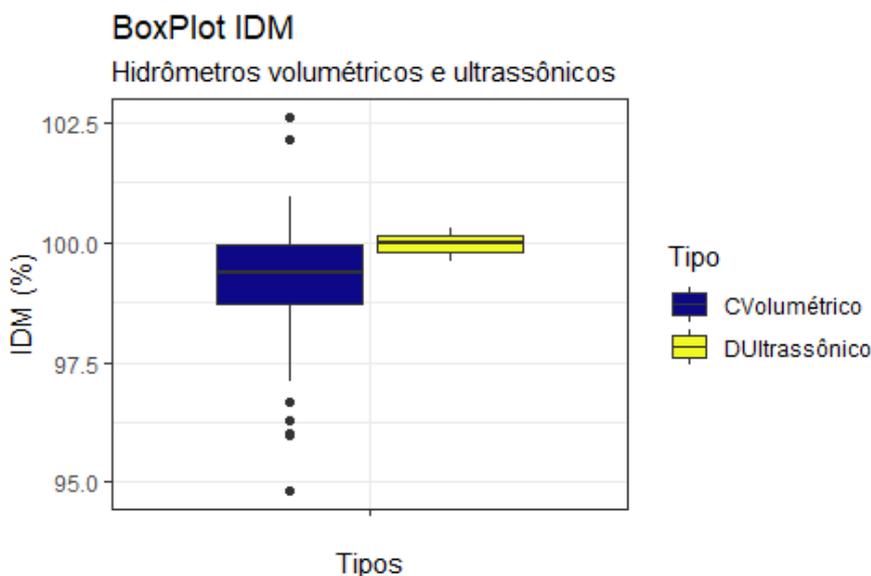


Figura 14: Histograma dos valores de IDM.

## ANÁLISE E DISCUSSÃO DOS RESULTADOS

Referente a parcela do Modelo de Regressão Linear Múltipla que envolve a classe do tipo, unijato ou multijato, esta não apresentou diferença estatística na distribuição. Este fato foi confirmado visualmente e com o teste de *Mann-Whitney*. Ainda assim optou-se por separar a obtenção de modelos matemáticos para os tipos uni e multijato devido ao início de funcionamento desses hidrômetros ser diferentes. As especificações indicam que os modelos unijato apresentam início de funcionamento em torno de 15 L/h, já o modelo multijato em torno de 30 L/h.

Os verdadeiros nomes das marcas foram ocultados por uma questão de exposição e respeito. Na estatística os testes são ferramentas para verificar se uma hipótese observada experimentalmente possui validade ou pode ser sustentada. Há no meio comercial discussões acerca da eficiência de certas marcas em determinados

modelos. Essas hipóteses puderam ser confirmadas, em alguns casos, e em outros, refutadas. Os coeficientes da marca hora são positivos ora são negativos, indicando que realmente algumas marcas são mais eficientes, ou até mesmo mais resistentes as condições de trabalho. Isso não desmerece nenhuma marca, pelo contrário, deve servir como ferramenta motivadora para cada vez mais os medidores serem eficientes e auxiliarem as companhias de saneamento na diminuição dos índices de perda de água no Brasil.

O coeficiente da volumetria é negativo, indicando que quanto vai ser utilizado em campo, pela passagem de água, maior será a queda de desempenho. E isso faz todo o sentido físico, haja vista da metodologia construtiva e de funcionamento dessa classe de medidores de água.

O modelo como um todo atendeu todas as suposições necessárias a um modelo de regressão linear. As métricas de erro apresentaram indicadores aceitáveis para a variável em estudo. E mais, quando aplicado o modelo à dados ensaiados em laboratório, em que estes valores a equação nunca havia “visto”, o desempenho foi de 76,3 %. Valor este muito próximo do desempenho do poder explicativo com os dados de treino do modelo.

Por fim, o modelo que procurou prever o decaimento da eficiência na medição de hidrômetros dos tipos volumétrico e ultrassônico não apresentou validade estatística devido à não significância das variáveis adotadas, de modo que a análise geral do modelo nas premissas de assimetria e curtose não foram atendidas. Destaca-se que pelo método construtivo de funcionamento destes medidores, o desgaste é menor, devido à passagem de água – volumetria – em comparação aos hidrômetros taquimétricos, classe metrológica B. Portanto, a análise e estudos de outras variáveis que possam influenciar no IDM é válida, para que possa ser encontrado modelo de previsibilidade do desempenho para estes tipos.

## CONCLUSÕES

Com base nas análises estatísticas, premissas, hipóteses e testes estatísticos pode-se encontrar um modelo de previsão do Índice de Desempenho Metrológico para hidrômetros do tipo taquimétrico, classe metrológica B. O modelo encontrado pode explicar a variabilidade dos dados em 75,6 %. Entre as variáveis analisadas o tipo de hidrômetro taquimétrico, metrologia B, unijato ou multijato, bem como a marca e a volumetria apresentaram significância estatística para compor o modelo.

Durante a pesquisa, observou-se que as análises exploratórias dos dados, bem como entendimento das hipóteses dos profissionais ligados ao setor, são de suma importância ao entendimento do modelo e dos valores obtidos de IDM, seja em laboratório ou na avaliação de resíduos estatísticos.

A sugestão de trabalhos futuros inserida nesse contexto seriam projetos objetivando encontrar equação que envolvesse os tipos de hidrômetros ultrassônicos e volumétricos, cujas classes metrológicas se apresentam superiores aos taquimétricos estudados. Uma vez que com as variáveis analisadas e os dados disponíveis o resultado apresentou-se inconclusivo para estes tipos.

A equação confirmou a hipótese de que ocorre um desgaste no medidor com o aumento da volumetria registrada e hipóteses acerca das diferenças de desempenho entre marcas. Através da modelagem matemática apurada na pesquisa, viabilizou-se a eficientização da gestão do parque de hidrômetros da Companhia Águas de Joinville, auxiliando na tomada de decisão das campanhas de substituições preventivas, ao passo que a previsibilidade do IDM permitiu levantar as melhores oportunidades de recuperação de receita. A concessionária apresenta índice de submedição de 4,64% (abril de 2020), número abaixo do valor de referência defendido pela International Water Association (5%).

## REFERÊNCIAS

1. ABNT. ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 15538/2007: Hidrômetros para água fria - Ensaio para avaliação de desempenho de hidrômetros em altas e baixas vazões em hidrômetro até 2,5 m<sup>3</sup>/h de vazão nominal para água fria. Rio de Janeiro, 2007.

2. FARCOMENI, A. *A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. Statistical Methods in Medical Research, Roma*, v. 17, n. 4, p. 347 – 388, 2008. Disponível em: <<https://doi.org/10.1177/0962280206079046>>. Acesso em: 02 maio 2020.
3. Lei Municipal nº 5054/2004, Município de Joinville.
4. PEÑA, E. A.; SLATE, E. H. *Gvlma: Global Validation of Linear Models Assumptions. Journal of the American Statistical Association*, [S.I.], v. 101, n. 473, p. 341 – 354, 01 jan. 2012. Disponível em:<<http://dx.doi.org/10.1198/016214505000000637>>. Acesso em: 10 mai. 2020.
5. R CORE TEAM (2017). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. Disponível em: <<https://www.Rproject.org/>>.
6. TARDELLI FILHO, J. Controle e redução de perdas, In: Tsutiya, M. T. Abastecimento de Água, Departamento de Engenharia Hidráulica e Sanitária da Escola Politécnica da Universidade de São Paulo, São Paulo, 2004.
7. THORNTON, J., STURM, R., KUNKEL, G., *Water Loss Control, Second Edition. New York: McGraw-Hill*, 2008.
8. TORMAN, V. B. L.; COSTER, R.; RIBOLDI, J. Normalidade de variáveis: Métodos de verificação e comparação de alguns testes não-paramétricos por simulação. *Revista HCPA*, v. 32, n.2, p. 227-234, 2012. Disponível em: < <http://hdl.handle.net/10183/158102>>. Acesso em: 02 maio 2020.
9. WICKHAM, H. *Elegant Graphics for Data Analysis. New York: Springer-Verlag*, 2009. 277p