

REGRESSÃO STEPWISE E BEST SUBSET PARA PREVISÃO DO CONSUMO DE ÁGUA EM EDIFÍCIOS DE ESCRITÓRIOS NA CIDADE DE JOINVILLE-SC¹

Maria Izabel do Nascimento², Andreza Kalbusch³, Elisa Henning⁴

¹ Vinculado ao projeto “Investigação de fatores relacionados ao consumo de água no ambiente construído”

² Acadêmico (a) do Curso de Engenharia Civil – CCT – Bolsista PIBIC/CNPq.

³ Orientador, Departamento de Engenharia Civil – CCT – andreza.kalbusch@udesc.br.

⁴ Coorientadora, Departamento de Matemática – CCT.

Há uma lacuna de estudos sobre o consumo de água em escritórios no Brasil. Compreender o uso de recursos hídricos é importante para uma maior conscientização e economia de água, visto que é um recurso limitado. Investigar o consumo de água de escritórios auxilia na identificação de desperdícios que podem ser evitados.

Para analisar o consumo de água de escritórios, diversos métodos estatísticos podem ser empregados, um deles é a regressão linear múltipla. A regressão linear múltipla investiga variáveis dependentes e independentes e define uma equação matemática, que nesta pesquisa foi usada para previsão do consumo de água. Assim, o consumo de água por área e per capita de 52 escritórios de advocacia, engenharia, imobiliária, contabilidade e outros (auditoria, recrutamento, seguros e eventos) da cidade de Joinville – SC foi estudado utilizando a regressão linear múltipla.

Dados do período de janeiro de 2018 a dezembro de 2019 constituídos de 24 variáveis independentes foram usados nesta pesquisa. Para criar um modelo significativo é importante realizar uma seleção de variáveis, para isso dois métodos foram empregados: regressão *best subset* e regressão *stepwise*. O primeiro cria modelos a partir de um subconjunto de variáveis explicativas e apresenta as possibilidades geradas, deixando o pesquisador optar pelo modelo desejado. Já o segundo adiciona e elimina variáveis até obter o modelo mais significativo estatisticamente e apresenta um modelo final.

Nesta pesquisa, a função logarítmica natural (LN) foi aplicada na variável dependente devido à ausência de normalidade. Para avaliação da validade do modelo foram feitos testes de modelagem, sendo eles: Teste *Shapiro-Wilk* para a verificação da normalidade, Teste *Outlier* para confirmar a ausência de outliers, Teste *Durbin-Watson* para verificar a independência dos resíduos, Teste *Breusch-Pagan* para observar a homoscedasticidade e VIF para testar a multicolinearidade. Para a acurácia do modelo, analisou-se o coeficiente de determinação (R^2) e o coeficiente de determinação ajustado (R^2 ajustado).

Os modelos de consumo hídrico per capita e por área gerados por meio de *regressão best subset* e *stepwise* atenderam aos testes de validade. Entretanto, ao analisar os valores de R^2 e R^2 ajustado, a modelagem de consumo por área obteve os melhores resultados. Em relação aos métodos utilizados, *best subset* e *stepwise* geraram modelos com as mesmas variáveis independentes. Sendo assim, a Tabela 1 apresenta os resultados da regressão linear múltipla *best subset/stepwise* para consumo de água por área ($m^3/m^2/ano$) em escritórios da cidade de Joinville-SC, juntamente com as estatísticas dos testes aos quais o modelo foi submetido.

Os resultados apontam que o consumo aumenta quando as reformas no sistema hidrossanitário ocorreram há mais de 5 anos e conforme aumenta a idade de imóvel. O consumo

também cresce de acordo com a quantidade de banheiros no escritório e a quantidade total de funcionários. Por sua vez, o consumo de água por área de pequenas, médias e grandes empresas é menor quando comparado ao consumo de água de microempresas. Os resultados também mostraram que as empresas que realizam atendimento ao público no escritório também possuem consumo menor, o que pode ser explicado pela realização de atendimento externo quando a empresa não faz atendimento ao público em sua sede, o que implica em menor tempo dos funcionários na edificação. Os resultados mostraram ainda que a área construída do escritório contribui para a diminuição do consumo por área da amostra, o que era esperado.

Tabela 1. Resultados da regressão linear múltipla (LN do consumo de água por área).

Variáveis	Estimativa	Erro Padrão	P-valor
Intercepto	-0,1753697	0,3448575	0,613744
Porte da empresa (faturamento) (0 para microempresa; 1 para pequena empresa)	-0,5151762	0,2221964	0,025358 *
Porte da empresa (faturamento) (0 para microempresa; 1 para média/grande empresa)	-1,4100558	0,3955860	0,000925 ***
Idade do imóvel (anos)	0,0100837	0,0041610	0,019764 *
Última reforma no sistema hidrossanitário: 0 (Menos de 5 anos) 1 (Mais de 5 anos)	0,3289602	0,1801229	0,074917 .
Realização de atendimento ao público: 0 (não) 1 (sim)	-0,5963581	0,2564884	0,024972 *
Área construída do escritório (m ²)	-0,0016505	0,0004776	0,001267 **
Número total de funcionários	0,0100844	0,0038222	0,011635 *
Número total de banheiros	0,1025841	0,0362511	0,007114 **
R ²	0,6063		
R ² ajustado	0,5313		
Teste Shapiro-Wilk	0,3207		
Teste Outlier	0,0014822		
Teste Durbin-Watson	0,558		
Teste Breusch-Pagan	0,7199		
VIF	All variables <2		

Significância dos códigos: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Palavras-chave: Consumo de água. Escritórios. Regressão linear múltipla.