



XI SILUBESA

Simpósio Luso-Brasileiro de Engenharia Sanitária e Ambiental

28/03 a 02/04/2004 - Natal - Rio Grande do Norte - Brasil

I-018 - PROGNÓSTICO DE DEMANDA DE ÁGUA MENSAL PARA EFEITO DE PLANEJAMENTO OPERACIONAL DA PRODUÇÃO DE ÁGUA

Marlene Alves de Campos Sachet⁽¹⁾

Bacharel em Estatística graduada em 1999 pela Universidade Federal do Paraná. Pós graduada em Administração de Empresa pela Faculdade de Administração e Economia - Centro de Desenvolvimento Empresarial - FAE/CDE em 2000, gestora da qualidade pela PUC/PR. Trabalha na Sanepar há 19 anos tendo atuado nas áreas de informática e suprimentos. Atualmente desempenha suas atividades no Grupo Específico de Consultoria Intercâmbio e Pesquisa - Sanepar como Pesquisadora.

(1)Endereço: Av. Iguaçu, 2920 apto 32, Bairro Água Verde, Curitiba - Pr. CEP 80230-041, Fone 55(41) 330-3129 -

 marleneacs@sanepar.com.br

RESUMO

O estudo apresentado ao longo deste trabalho, faz uma abordagem da forma, de como se desenvolve predição de um processo de um Sistema de Abastecimento de Água de pequeno porte. Neste caso ocupa-se principalmente, em apresentar uma metodologia para fazer estimativas de demanda mensal no sistema produtor de um Sistema de Abastecimento de Água, dando enfoque principalmente como essa água é consumida na forma do abastecimento público de uma cidade, em benefício da saúde pública, do saneamento ambiental e de suas atividades socioeconômica. Objetivando buscar um maior entendimento de como a água é demandada, a primeira preocupação foi a de quantificar a água consumida por categorias para este caso foi aplicado para categoria residencial, mas isso não implica que seja expandido para outras categorias, mas com algumas restrições. A técnica estatística - regressão múltipla foi utilizada neste estudo, visando contribuir na definição de uma metodologia adequada com um nível de confiabilidade razoável e para que se possa aplicar como subsídio ferramental para novos investimentos.

PALAVRAS-CHAVES: Prognóstico de Demanda, Regressão Linear Múltipla, Produção de Água, Planejamento, Operacional.

1. INTRODUÇÃO

A demanda de água para consumo nas atividades socioeconômica de uma população urbana, depende de fatores tais como: o clima, a temperatura, o crescimento populacional, umidade relativa do ar, atividades socioeconômico da população. A água é um ingrediente vital para a realização das atividades socioeconômica de uma população.

O prognóstico de demanda é uma ação importante para o planejamento estratégico, tático e operacional de um Sistema de Abastecimento de Água - SAA.

A quantidade de água necessária para uma cidade pode ser prevista ou simulada através de modelo estatístico e usando dados do passado.

Duas abordagens clássicas disponíveis para análise de tais dados são : A Regressão Múltipla e Análise de Série Temporal. Mais recentemente outros métodos passaram a ser pesquisados, desenvolvidos e testados como: o uso da Lógica Difusa e Redes Neurais.

No presente artigo usa-se a abordagem clássica de Regressão Múltipla, embora as outras abordagens citadas também

estejam sendo consideradas em outros estudos.

Regressão Múltipla é adequada para aplicação de dados anuais e para dados mensais em cidades que apresentam pequenas taxas de crescimento populacional.

A Sanepar opera 620 localidades(342 sedes municipais e 278 localidades menores), de Sistemas de Abastecimento de Água dos quais muitas são em cidades com esta característica de pequeno crescimento populacional anual.

2. OBJETIVO

O objetivo deste artigo é apresentar uma metodologia estatística, utilizando a Regressão Múltipla, para prognóstico mensal de demanda de água relativo ao sistema produtor de um Sistema de Abastecimento de Água. Esta metodologia será apresentada e explicada simultaneamente para a de Cidade de Paranavaí no Estado do Paraná.

3. ELEMENTOS SOBRE A REGRESSÃO

3.1 Regressão Linear Múltipla

O objetivo do estudo de ajustar um modelo de regressão ao prognóstico de demanda de água é o de prever a variável resposta através das variáveis independentes, em muitas situações se concluirá que previsões da variável resposta contendo uma única variável independente são muito imprecisas e modelos contendo mais variáveis independentes serão analisados para melhoria desta precisão. Dentre um conjunto possível de variáveis independentes pretende-se escolher um subconjunto destas que produza um " bom" modelo, isto é, um modelo que forneça estimativas precisas da variável resposta e que faça sentido para o objetivo proposto.

Formalmente, para o ajuste de um modelo de regressão linear múltipla, a variável resposta bem como as variáveis independentes devem ser contínuas.

Uma forma conveniente de análise é supor que cada observação é composta de duas partes: uma parte previsível e uma parte aleatória. A parte previsível é usualmente expressa por uma função matemática, e a parte aleatória impõe-se que obedeça uma distribuição de probabilidade.

A representação acima constitui um modelo matemático em que cada observação depende da parte previsível, conjuntamente com a parte aleatória. Tem-se assim o que se chama de variável dependente ou resposta, e a variável independente. Se uma relação linear entre a variável resposta e a independente existe, esta pode ser representada por uma função matemática.

No caso de duas ou mais variáveis quantitativas, bem como o coeficiente de correlação, podem ser de grande auxílio na escolha da referida função.

Este modelo, exibido a seguir, engloba a parte previsível e aleatória (erro) do fenômeno em uma relação linear entre a variável resposta e a variável independente:

$$Y = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_k X_{kj} + \varepsilon_j$$

onde $j=1,2, \dots, n$

Y = Variáveis dependente (resposta)

β_0 e β_k = Coeficientes de Regressão

X_{kj} = Variáveis independente(regressora)

Os dados, no entanto, não caem exatamente sobre a linha reta, ou seja, existe uma diferença entre o valor observado e o valor da linha reta, a qual é chamado "erro"(será representado por ε).

Este erro é assumido ser um "erro estatístico", isto é ele é uma variável aleatória que quantifica a falha do modelo em se ajustar aos dados exatamente. Tal erro pode ser devido ao efeito de outras variáveis.

Será apresentado no decorrer deste estudo o "modelo de regressão linear múltipla amostral", que é denotado por:

$$y_i = \beta_0 + \beta_1 x_i + \beta_k x_j \text{ em que } i=1,2,\dots,n$$

Assume-se que os erros têm média "zero" e variância desconhecida constante σ^2 . Assume-se ainda que os erros são não-correlacionados, isto é, o valor de um erro não depende do valor de qualquer outro erro.

3.2. Método Stepwise

O método de Stepwise é um procedimento que avalia todas as regressões possíveis sendo em algumas situações muito árduo. Vários outros métodos têm sido desenvolvidos para avaliação somente de um pequeno número de modelos de regressão, seja pela adição ou retirada das variáveis em cada tempo. Esses métodos são referidos como "procedimento do tipo stepwise", e podem ser classificados em 3 categorias:

Seleção de Forward - entra sem nenhuma variável no modelo e vai inserindo variável independente, uma de cada vez, no modelo até que o melhor (ou melhores) modelo é obtido.

Eliminação de Backward - começa com o modelo com todas as variáveis completas e vai retirando uma a uma, as que não forem significativas.

Regressão Stepwise - é uma modificação da seleção de Forward, onde, em cada passo todas as regressoras que entrarem no modelo são reaceessadas via sua estatística F parcial, e requer dois valores de corte $F_{in} > F_{out}$

3.3. Diagnóstico de Influência

Ocasionalmente encontramos um subconjunto de dados que podem exercer uma desproporcional influência no modelo de regressão estimado. Localizar estes pontos e acessar seu impacto no modelo, é um fator primordial para a decisão final na escolha do modelo.

Por fim, também serão usados intervalos de confiança para estimar os coeficientes de regressão e comparar resultados associados a quantidade de água a ser produzida, como por exemplo, intervalos de confiança ou intervalos da predição para produção mensal de água.

4. METODOLOGIA ESTATÍSTICA PARA PROGNÓSTICO DE DEMANDA MENSAL DE UM SUB-SISTEMA PRODUTOR DE ÁGUA

A seguir a metodologia é apresentada e exemplificada com a sua aplicação para a cidade de Paranavaí, que apresenta uma característica de pequeno índice de crescimento, atendendo o pressuposto para o uso da técnica de regressão múltipla.

4.1. Local da Aplicação

O estudo será aplicado na cidade de Paranavaí, que apresenta uma característica atendendo o pressuposto do uso da técnica de regressão múltipla.

A cidade de Paranavaí fica localizada ao noroeste do estado do Paraná.

Apresenta um clima subtropical úmido mesotérmico, verões quentes com tendência de concentração das chuvas (temperatura média superior a 22 °C), invernos com geadas pouco frequentes (temperatura média inferior a 18 °C), sem estações secas definidas.

A população atual da cidade é de 70.245 hab. na zona urbana e 5.417 hab. na zona rural, tem uma taxa anual de crescimento 0,91% (dados do censo 2000).

4.2. Variáveis do Estudo

4.2.1. Variável dependente

Volume de produção de água mensal ($\text{m}^3/\text{mês}$) demandada do sistema produtor para consumo da categoria residencial.

4.2.2. Variáveis Independentes para Categoria Residencial

- Número de economias;
- Temperatura($^{\circ}\text{C}$);
- Densidade pluviométrica(mm de precipitação de água);
- Preço de venda(R\$).

Na cidade de Paranavaí predominam as contas de consumo residencial, de taxa de consumo mínimo ($10\text{m}^3/\text{mês}$), portanto esta será a variável utilizada.

Acima estão as variáveis consideradas como as mais importantes e determinantes do consumo de água residencial.

4.3. Levantamento dos Dados

Os dados mensais de produção de água para consumo residencial foram retirados dos registros internos do Sistema de Abastecimento de Água - SAA da cidade de Paranavaí - Sanepar, referentes ao período de janeiro/1999 a dezembro/2001(36 meses).

Os dados da temperatura média mensal e densidade pluviométrica mensal foram fornecidos pela Sistema Meteorológico do Paraná – SIMEPAR.

Quanto aos registros destes dados, não foi feita validação simultânea, tendo sido adotado os dados registrados.

Tabela de Dados 1: Categoria Residencial de Paranavaí, Período de 01/1999 a 12/2001

Meses	Volume Produzido (m^3)	Número de Economias	Temperatura ($^{\circ}\text{C}$)	Densidade Pluviométrica (mm)	Preço(R\$)
jan-99	376572,2	19708	31	160,8	8,21
fev-99	322599,2	19731	31	139,6	8,21
mar-99	339079,1	19748	31	98,8	8,99
abr-99	334490,6	19839	29	91,2	8,99
mai-99	327097,6	19910	25	83,6	8,99
jun-99	302877,0	19940	23	129,4	8,99
jul-99	298461,1	20018	26	118,8	8,99
ago-99	315474,7	20046	28	0,0	8,99
set-99	305536,9	20109	30	51,8	8,99
out-99	330002,7	20184	30	71,0	8,99
nov-99	358894,3	20236	30	19,6	9,26
dez-99	346209,4	20288	32	213,2	9,26
jan-00	388106,3	20289	33	97,2	9,26
fev-00	328794,6	20366	30	404,4	9,26
mar-00	321269,5	20421	30	92,4	9,26

abr-00	326777,4	20542	30	59,6	9,26
mai-00	314467,8	20577	26	36,4	9,26
jun-00	298146,7	20633	26	89,4	9,26
jul-00	298045,7	20691	22	70,4	9,26
ago-00	295155,6	20736	27	175,8	9,26
set-00	326218,0	20813	26	208,2	9,26
out-00	338594,6	20872	32	94,6	9,26
nov-00	342391,7	20960	31	125,8	9,26
dez-00	319751,9	21062	31	129,6	9,26
jan-01	323532,8	21001	32	159,4	10,25
fev-01	315404,9	20982	31	173,4	10,25
mar-01	315111,1	20996	32	196,2	10,25
abr-01	323024,2	21003	31	43,8	10,25
mai-01	268310,5	21075	25	169,6	10,25
jun-01	258692,3	21119	23	112,2	10,25
jul-01	248301,1	21193	26	31,2	10,25
ago-01	301331,5	21242	29	62,6	10,25
set-01	335666,4	21248	28	84,0	10,25
out-01	291459,3	21269	30	63,0	10,25
nov-01	300059,2	21303	31	144,8	10,25
dez-01	311794,3	21299	30	226,80	11,55

Fonte: COMAPRO-SIS/SANEPAR/PARANVAI/2001

5. ANÁLISE DOS DADOS E COMENTÁRIOS

5.1. Estatística Descritiva Categoria Residencial

A seguir apresentamos a análise descritiva, tabela 2, figuras 1,2, 3 e 4, e principais estatísticas como média e desvio padrão.

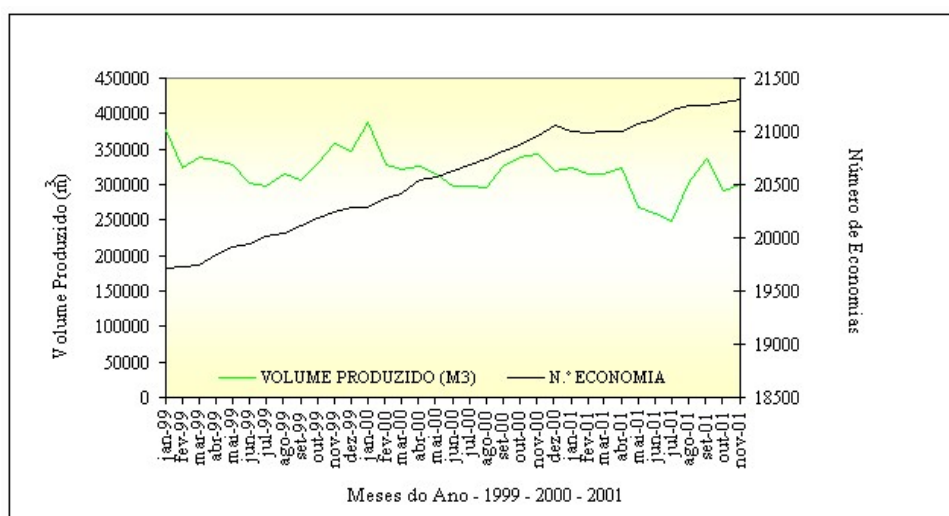
Através da estatística descritiva, constata-se que os valores extremos superiores de cada variável foram:

- Para número de economias : crescimento no ano de ano 2001 e respectiva média e desvio padrão.
- Para temperatura : crescimento no ano de 1999 para a média e ano 2000 desvio padrão.
- Para densidade Pluviométrica: crescimento no ano 2000 para a média e ano de 1999 para o desvio padrão.
- Para preço: crescimento no ano 2001 para a média e ano 2000 para o desvio padrão.

Tabela 2 - Médias, Desvio Padrões

Ano/Variáveis	Média			Desvio Padrão		
	1999	2000	2001	1999	2000	2001
N.º Economias	19979,75	20663,50	21144,17	200,14	239,18	128,49
Temperatura	28,93	28,57	28,84	2,78	3,10	2,83
Dens. Pluviométrica	98,15	131,98	122,25	59,72	98,22	64,82
Preço	8,9	9,26	10,36	0,34	0,00	0,37

Figura 1 - Categoria Residencial, Volume Produzido (m³) x Número de Economias

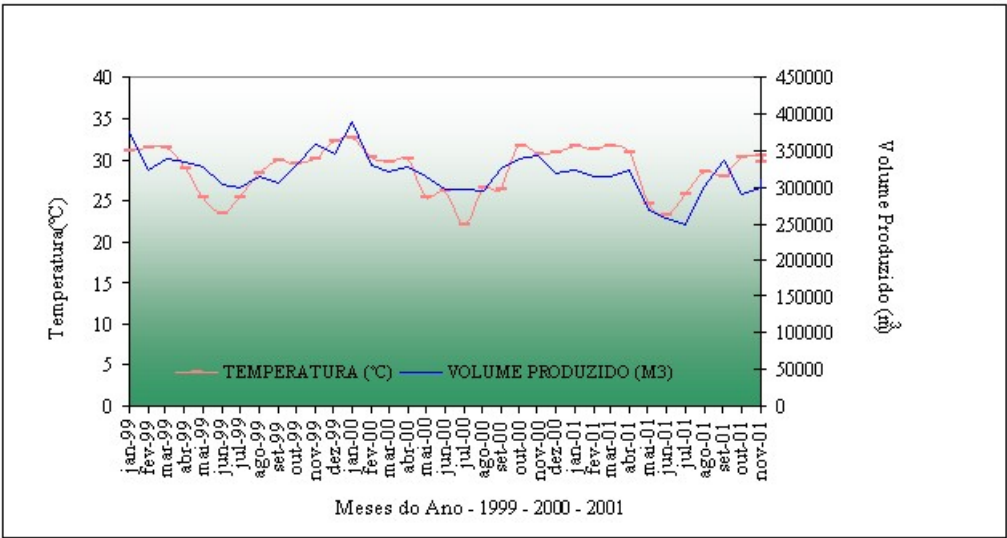


Fonte: Pesquisa/DMA- Sanepar

A figura 1, sugere que o crescimento do número de economias seja contínuo e de forma significativa no de 1999 a 2001, o consumo apresenta uma tendência decrescente.

Sugere-se que os atos de causa desse declínio do consumo de água, sejam a queda do poder aquisitivo da população e envelhecimento do parque de hidrômetros.

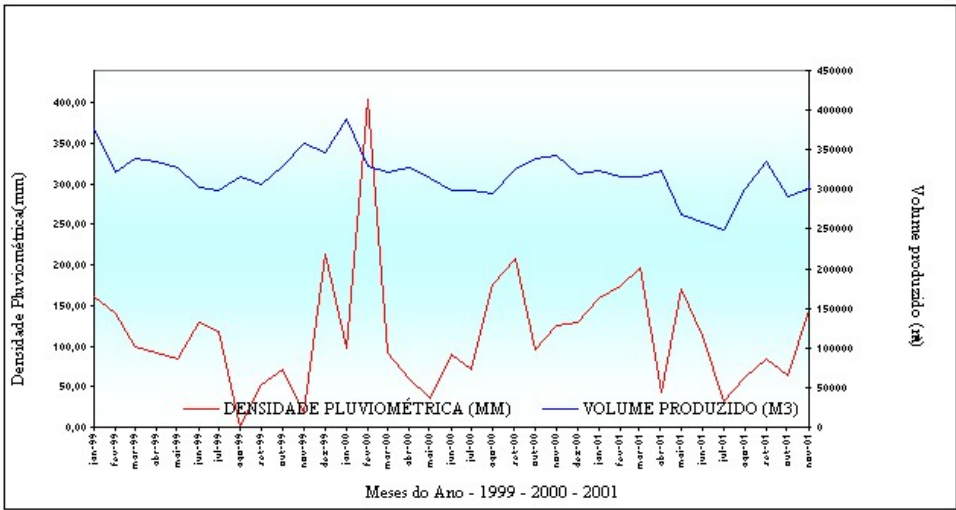
Figura 2 - Categoria Residencial, Volume Produzido (m³) x Temperatura(°C)



Fonte: Pesquisa/DMA- Sanepar

A figura 2, mostra que a temperatura é um fator muito significativo na produção de água da cidade de Paranavai. Altas temperaturas resultam em mais consumo de água pela população.

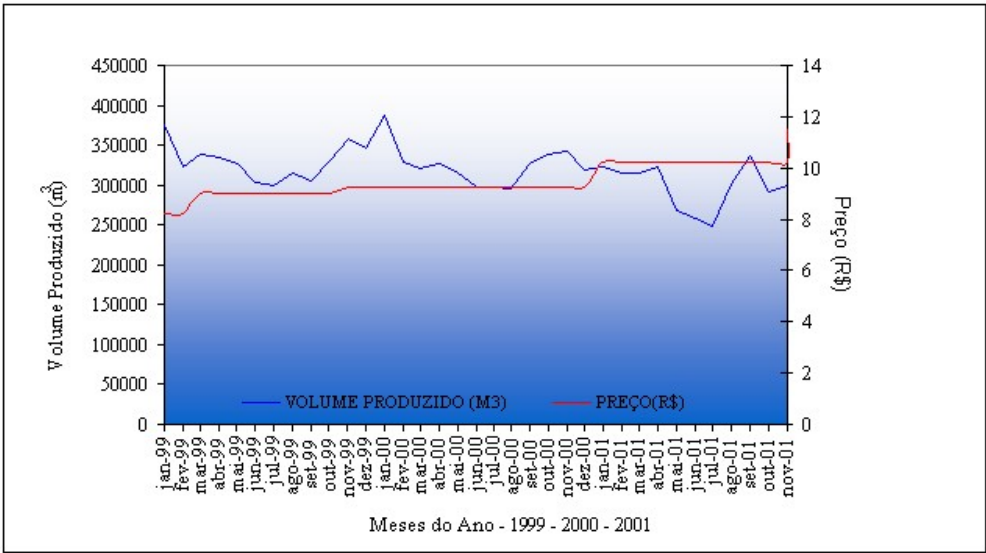
Figura 3 - Categoria Residencial, Volume Produzido (m³) Densidade Pluviométrica



Fonte: Pesquisa/DMA- Sanepar

Demonstra-se na figura 3 que há indícios de alto consumo de água logo após as estiagens seguidas de períodos chuvosos.

Figura 4 - Categoria Residencial, Volume Produzido (m³) x Preço (R\$)



Fonte: Pesquisa/DMA- Sanepar

A variável preço demonstra, conforme figura 04, que temporariamente é significativa, principalmente quando a taxa mínima de água tem reajuste. No figura está bem claro, que houve um aumento da tarifa em dezembro/01, porém o consumo de água teve uma queda relevante.

A estatística descritiva demonstra como as variáveis estão se comportando no período analisado, mas para o profissional tirar conclusões mais relevantes, outras técnicas inferenciais adequadas devem ser aplicadas. Para este caso será aplicado a Regressão Linear Múltipla, devido ao experimento apresentar diversas variáveis independentes.

5.2. Análise de Regressão para a Categoria Residencial

A tabela 1- Dados, apresenta 4 variáveis independentes e portanto $2^4=16$ possíveis modelos de equações de regressão, isso considerando todas com o intercepto e sem a presença de interações. Os resultados referente a R^2_p , R^2_a , e $C(p)$ Mallows encontram-se na tabela 3 - Possíveis modelos para categoria residencial.

Tabela 3 - Possíveis Modelos para Categoria Residencial

p	Variável do Modelo	R^2_p	R^2_a	Cp-Mallows
1	Temperatura	0,37	0,36	16,26
1	N.º de Economias	0,20	0,20	28,89
1	Preço	0,20	0,17	30,16
1	Densidade Pluviométrica	0,08	-0,21	44,43
2	N.º Economias Temperatura	0,59	0,56	2,15
2	Temperatura Preço	0,59	0,55	2,66
2	Temperatura Densidade Pluviométrica	0,37	0,33	18,26
2	N.º Economias Densidade Pluviométrica	0,23	0,17	29,77

2	N.º Economias Preço	0,22	0,17	30,16
2	Densidade Pluviométrica Preço	0,27	0,17	30,70
3	N.º Economia Temperatura Preço	0,59	0,56	3,07
3	N.º Economia Temperatura Densidade Pluviométrica	0,59	0,55	4,11
3	Temperatura Densidade Pluviométrica Preço	0,57	0,53	4,55
3	N.º Economias Densidade Pluviométrica Preço	0,24	0,16	30,84
4	N.º Economias Temperatura Densidade Pluviométrica Preço	0,59	0,54	5,00

A partir dos resultados obtidos na tabela 3, verifica-se que pelos critérios R^2_p , R^2_a e $C(p)$, alguns conjuntos de variáveis são apresentados como candidatos ao melhor modelo final, tais como:

- Número de economias e temperatura, $R^2_p=0,59$, $R^2_a=0,56$, $C(p)=2,14$
- Temperatura e preço, $R^2_p=0,59$, $R^2_a=0,55$, $C(p)=2,66$
- Número de economias, temperatura e preço, $R^2_p=0,59$, $R^2_a=0,55$, $C(p)=3,07$
- Número de economias, temperatura e densidade pluviométrica, $R^2_p=0,59$, $R^2_a=0,55$, $C(p)=4,11$
- Temperatura, densidade pluviométrica e preço, $R^2_p=0,57$, $R^2_a=0,53$, $C(p)=4,55$
- Número de economias, temperatura, densidade pluviométrica e preço, $R^2_p=0,59$, $R^2_a=0,54$, $C(p)=5,00$.

Analisando seus $C(p)$ Mallows todos indicam valores abaixo do número de seus parâmetros ($p=5$), isso indica que todos são fortemente candidatos ao modelo final.

A análise de escolha ao melhor candidato a um modelo final pelo critério de análise R^2_p , R^2_a e seus $C(p)$'s, não indica uma escolha definitiva. Outras técnicas estatísticas deverão ser aplicadas para que a melhor escolha seja selecionada, a fim de que o modelo final apresentado seja tão significativo em sua aplicabilidade.

Inicialmente será verificado o grau de associação entre a variável resposta volume produzido(m^3) de água e as variáveis independentes número de economias, temperatura, densidade pluviométrica e preço, utilizando o coeficiente de correlação de Pearson, conforme mostrado na tabela 4.

Tabela 4 - Matriz de Correlação

	Número de Economia	Temperatura	Densidade Pluviométrica	Preço	Volume Produzido
Número de Economia	1.00000	-0,044223	0,073793	0,846297	-0,457307
Temperatura	0,044223	1,0000000	0,178026	-0,015650	0,628006
Densidade Pluviométrica	0,073793	0,178026	1,000000	0,115542	0,086356
Preço	0,846297	-0,015650	0,115542	1,000000	-0,438859
Volume Produzido	-0,457307	0,628005	0,086357	-0,438859	1,000000

Quanto mais próximo de 1 o coeficiente estiver, mais forte será o grau de relação linear existente entre as variáveis em estudo. Desta forma, observamos que a variável volume produzido de água apresenta associação linear positiva nas variáveis temperatura e densidade pluviométrica, e uma relação linear negativa nas variáveis economia e preço, vide Figuras 1 e 2.

Demonstra uma forte relação linear positiva entre as variáveis independentes número de economias e preço, ou seja que as duas variáveis podem estar explicando a mesma situação. Diante disso outra técnica será aplicada. O mais indicado é o teste de multicolinearidade. Ver tabela 5.

Tabela 5 - Teste de Multicolinearidade - Fator de Inflação(VIF)

	Tolerância	Fator de Inflação
Número de Economias	0,282467	3,5402366
Temperatura	0,964478	1,0368303
Densidade Pluviométrica	0,953103	1,0492045
Preço	0,280796	3,561317

Como os resultados apresentados na tabela 5 teste de multicolinearidade das variáveis, apresentam-se com todos os valores inferiores a 5, isso indica que apesar de existir uma correlação linear entre as variáveis destacadas na tabela 4, as estimativas dos coeficientes de regressão não estão sendo fortemente afetadas por esta correlação.

Para escolha do melhor modelo final que seja tão significativo em sua aplicabilidade, o método de stepwise será aplicado conforme tabela 6 a seguir.

Tabela 6 - Método de Stepwise - Seleção do melhor modelo

Método Aplicado	Modelo Final	R ²	F	Prob>f	p-valor significativo	C(p)
Backward	Densidade Pluviométrica e Preço	0,59	0,07 0,10	0,7864 0,3010	0,10	3.07
Stepwise	Temperatura e Economia	0,59	20.23 16.54	0.0001 0.0003	0.15	2.14
Forward	Temperatura, número de economias e preço	0,59	20.23 16.54 1.10	0.0001 0.0003 0.3010	0,05	3.07

Analisando a tabela 6 para todos os modelos apresentados, o que mais foi significativo foi o método de Stepwise, nota-se que os valores apresentados prob>f são inferiores ao p-valor significativo que é de 0,15 seu R² permaneceu igual aos demais, e C(p) Mallows é o menor valor apresentado.

Diante desses resultados parece muito favorável à candidato ao modelo final contendo as variáveis temperatura e número de economias.

5.2.1. Indicação a um Provável Modelo Final com as Variáveis Temperatura(°C) e Número de Economias.

$$\text{Vol. Produzido} = \beta_0 + \beta_1 \text{Economia} + \beta_2 \text{Temperatura}$$

Tabela 7 - Sumário da Regressão para o modelo (Vol. Prod = $\beta_0 + \beta_1 \text{Economia} + \beta_2 \text{Temperatura}$)

	Beta	Erro Padrão de Beta	B	Erro Padrão de B	T(32)	p-valor
Intercepto			623553,0	131494,4	4,74	0,000039
N.º Economia	-0,430377	0,113026	-23,3	6,1	-3,80	0,000579
Temperatura	0,608973	0,113026	6052,9	1123,4	5,38789	0,000006

$$R = 0,76 \quad R^2 = 0,59 \quad R^2_{\text{ajustado}} = 0,56$$

O valor do coeficiente de determinação múltiplo $R^2 = 59\%$, indica o quanto o volume produzido (m³) está sendo explicado pelas variáveis número de economias e temperatura.

Apresentamos a seguir o ajuste do modelo final :

$$\text{Vol. Produzido} = \beta_0 + \beta_1 \text{Economia} + \beta_2 \text{Temperatura},$$

$$\text{Vol. Produzido} = 623553 - 23,3(\text{Economia}) + 6052,9(\text{Temperatura}),$$

A interpretação dos coeficientes dos parâmetros é dada por: β_0 , como zero não está presente em nenhuma das regressoras, β_0 não tem qualquer significado na prática, $\hat{\beta}_1$, indica um decréscimo de 23,3 do volume produzido (m³), mediante a mudança de uma unidade no número de economias, quando a variável temperatura se mantiver constante, $\hat{\beta}_2$ indica um acréscimo de 6052,9, na variável volume produzido (m³) quando ocorre o acréscimo de uma unidade na variável temperatura e quando número de economia se mantiver constante.

Adiante será apresentado o teste da Anova.

Tabela 8 - Anova, para Categoria Residencial

	Soma dos Quadrados	Gl	Quadrado Médio	F	p-valor
Regressão	160689304,00	2	8034465225,0	22,71	0,000001
Residual	116718860,00	33	353693516		
Total	277408164,00				

Na tabela 8, verifica-se que a redução na Sqres foi de 116718860,00 unidades ao quadrado, ou ainda, o ganho em usar o modelo $\text{Vol. Produzido} = \beta_0 + \beta_1 \text{Economia} + \beta_2 \text{Temperatura}$ foi de 59%.

Houve uma substancial redução na Sqres (de 277408164 para 116718860,0). Assim, conclui-se que número de economias e temperatura estão explicando 59% da variação total ocorrida no volume produzido (m^3) nas estações de tratamento de água da cidade de Paranavaí. A variação restante, 41% é devida a outros fatores (alguns possíveis de serem controlados outros não).

O teste F apresentado na tabela 8, também apresenta uma redução na Sqres, como verifica-se através do seu valor $F=22,71$ com p-valor 0,000001, concluímos que há evidências estatísticas de que número de economias e temperatura ajudam a explicar o volume produzido (m^3) de água na estação de tratamento de água da cidade de Paranavaí.

A análise de correlação e multicolinearidade (Fator de Inflação) entre as variáveis será apresentada a seguir conforme resultados nas tabela 9 e 10, respectivamente.

Tabela 9 - Matriz de Correlação

N.º Economia Temperatura Volume Produzido

N.º Economia 1,000000 -0,044223 -0,457307

Temperatura -0,044223 1,000000 0,628005

Tabela 10 - Teste de Multicolinearidade - Fator de Inflação(VIF)

	Tolerância	Fator de Inflação
Economia	0,998044	1,0019598
Temperatura	0,998044	1,0019598

Analisando a matriz correlação e o teste de multicolinearidade das variáveis número de economias e temperatura, nota-se que não existe problemas que afetam o andamento da análise.

As figuras 5 e 6 trazem os gráficos dos resíduos padronizados versus valores preditos e o figura da normalidade dos erros, os quais indicam que não há indícios graves de violação dos pressupostos da homogeneidade das variâncias e da normalidade dos erros.

A seguir apresentamos o teste de Durbin Watson para verificar se os erros estão ou não autocorrelacionados, conforme tabela 10.

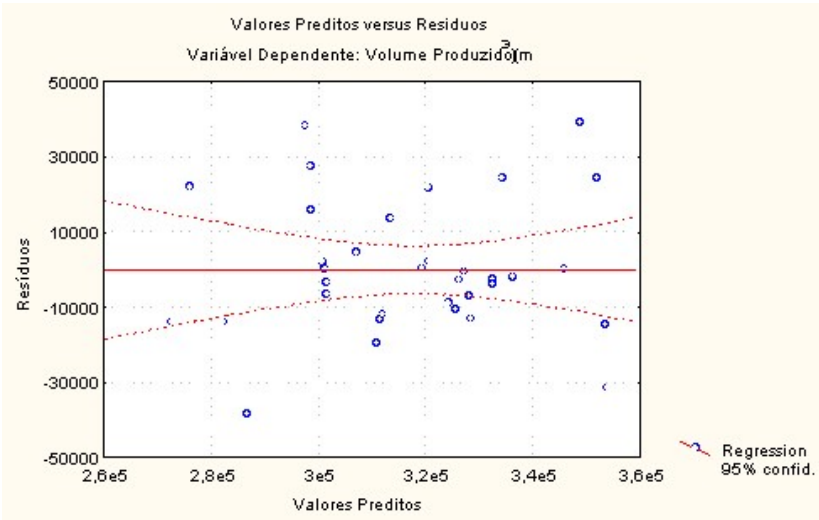
Tabela 10 - Estatística de Durbin Watson -

Correlação dos Erros

	Durbin Watson	Correlação Serial
	d	
Estimado	1,92	-0,70466

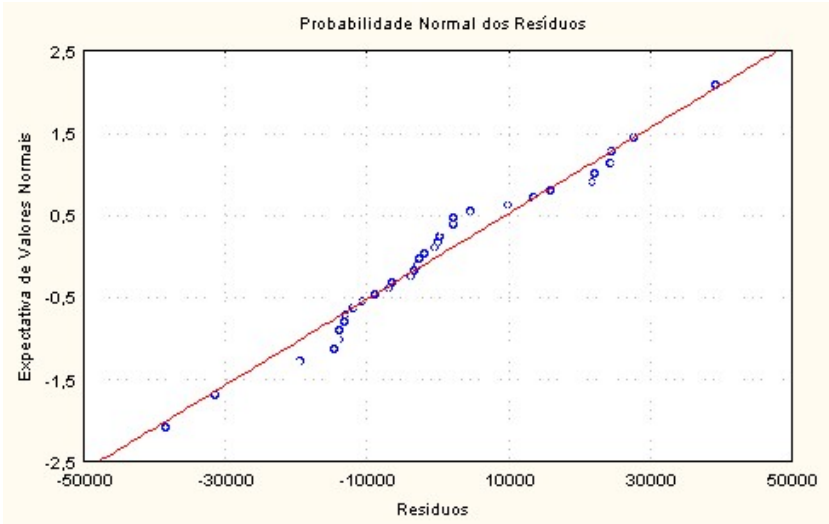
O teste de Durbin Watson, indica que os pressupostos da autocorrelação dos erros estão atendidos, ou seja não existe autocorrelação séria entre os resíduos, o valor tabela $d_l=1,39$ e $d_u=1,60$ e valor encontrado é de 1,92.

Grafico 5 -Valores Preditos versus Resíduos



Fonte: Gecip/DMA/Sanepar

Figura 6 - Probabilidade Normal dos Resíduos



Fonte: Gecip/DMA/Sanepar

Ocasionalmente encontramos um subconjunto de dados que exercem uma desproporcional influência no modelo de

regressão estimado. O diagnóstico apresentado abaixo irá demonstrar se estes pontos tem algum impacto no modelo apresentado.

5.2.2. Diagnóstico de Influência

Tabela 11 - Análise de Influência.

			Hat Diag	Covratio		Intercepto	Economias	Temperatura
Obs	Residuo	Rstudent	H	Ratio	Dffits	Dfbetas	Dfbetas	Dfbetas
1	23754.6	13.753	0.1270	1.0574	0.5244	0.3726	-0.4247	0.1859
2	-29648.8	-17.409	0.1227	0.9529	-0.6511	-0.4570	0.5224	-0.2347
3	-12747.8	-0.7199	0.1196	1.1871	-0.2654	-0.1846	0.2114	-0.0969
4	-3231.9	-0.1780	0.0884	1.1994	-0.0554	-0.0448	0.0459	-0.0018
5	14834.8	0.8439	0.1272	1.1763	0.3222	0.2478	-0.2012	-0.2015
6	3207.9	0.1873	0.1885	1.3470	0.0903	0.0609	-0.0443	-0.0706
7	-17052.3	-0.9530	0.0902	1.1084	-0.3001	-0.2259	0.1875	0.1647
8	-11195.9	-0.6112	0.0620	1.1293	-0.1572	-0.1193	0.1127	0.0306
9	-31424.1	-17.839	0.0574	0.8755	-0.4403	-0.2590	0.2907	-0.1249
10	-5100.8	-0.2755	0.0503	1.1466	-0.0634	-0.0332	0.0378	-0.0192
11	25078.6	13.898	0.0461	0.9643	0.3054	0.1436	-0.1664	0.0967
12	1830.9	0.0999	0.0718	1.1804	0.0278	0.0058	-0.0104	0.0191
13	37827.2	22.515	0.0966	0.7809	0.7362	0.0990	-0.2364	0.5748
14	-1801.4	-0.0966	0.0380	1.1390	-0.0192	-0.0059	0.0074	-0.0067
15	-7964.4	-0.4275	0.0356	1.1180	-0.0822	-0.0186	0.0247	-0.0296
16	540.2	0.0289	0.0327	1.1337	0.0053	0.0001	-0.0005	0.0020
17	12798.8	0.6973	0.0550	1.1092	0.1683	0.0359	-0.0043	-0.1183
18	-2135.4	-0.1155	0.0551	1.1593	-0.0279	-0.0038	-0.0014	0.0196
19	22901.4	13.739	0.1870	1.1359	0.6590	0.1076	0.0474	-0.6062
20	-8500.9	-0.4578	0.0413	1.1217	-0.0950	0.0067	-0.0213	0.0499
21	30393.8	17.225	0.0600	0.8942	0.4351	-0.0433	0.1255	-0.2931
22	8679.6	0.4747	0.0698	1.1545	0.1300	-0.0629	0.0441	0.0907
23	20581.5	11.367	0.0577	1.0335	0.2812	-0.1655	0.1385	0.1477
24	467.8	0.0255	0.0666	1.1749	0.0068	-0.0046	0.0040	0.0033
25	-3187.3	-0.1747	0.0791	1.1875	-0.0512	0.0307	-0.0240	-0.0335
26	-5860.5	-0.3181	0.0594	1.1550	-0.0800	0.0486	-0.0411	-0.0414
27	-11732.9	-0.6467	0.0787	1.1449	-0.1890	0.1124	-0.0876	-0.1242
28	2278.9	0.1237	0.0612	1.1665	0.0316	-0.0198	0.0169	0.0161
29	-15099.6	-0.8468	0.1019	1.1426	-0.2852	0.0829	-0.1392	0.1995
30	-11777.4	-0.6853	0.1721	1.2680	-0.3124	0.0590	-0.1282	0.2559
31	-38111.9	-22.655	0.0927	0.7736	-0.7241	0.3427	-0.4617	0.3925
32	-1643.9	-0.0897	0.0720	1.1809	-0.0250	0.0188	-0.0196	-0.0009
33	38764.9	22.850	0.0751	0.7535	0.6509	-0.4507	0.5036	-0.1155
34	-16772.8	-0.9317	0.0802	1.1004	-0.2752	0.2182	-0.2125	-0.0659

35	-13256.2	-0.7393	0.0965	1.1537	-0.2416	0.1940	-0.1787	-0.0980
36	4305.2	0.2367	0.0846	1.1918	0.0720	-0.0579	0.0565	0.0168

Soma do Resíduos 1.4842954E-9

Soma do Quadrado dos Resíduos 11582052455

Resíduo Preditos (Press) 14032047365

Pelo diagnóstico de influência, observa-se que, os elementos da diagonal da matriz H, os h_{ii} , mostram que os pontos 6, 19 e 30 possuem $h_{ii} > 2p/n = 0,1666$, o que evidencia que tais pontos devam ser investigados pois são possíveis pontos influentes. Os $dfbetas$ e $dfits$ apresentados, não indicam nenhuma observação discrepante, seu ponto de corte é igual a 1. O covratio é $1 \pm 3.p/n$, neste caso (0,75 ; 1,25), as observações 6 e 30 merecem uma atenção especial.

Através do diagnóstico de influência os pontos 6 e 30 é o que merece mais atenção nesta análise. Vejamos a seguir análise com e sem os pontos.

Tabela 12 - Retiradas de Observações do Modelo

Observação	β_0	β_1	β_2	QMres	R^2
Com obs. 6 e 30	623553,0	-23,3	6052,9	353693516,0	0,59
Sem obs. 6	617905,1	-23,1	6105,5	364565317,0	0,57
Sem obs. 30	615837,5	22,5	5.736,5	357527009,6	0,52
Sem obs. 6 e 30	582593,9	-20,9	5652,0	324261782,7	0,55

Retirar observação 30, produz uma mudança em $\beta_0 = 1,23\%$, $R^2 = 11,86\%$, $\beta_1 = 3,43\%$, $\beta_2 = 5,22\%$, QMres=1,08%, portanto concluímos que esta observação demonstra uma influência fraca nos coeficiente de regressão. Já a retirada da obs. 6, produz uma mudança relativamente menor nos coeficientes de regressão e a retirada de ambas não produzem mudanças similares àquelas observadas quando da retirada somente da observação 30.

Em investigações feitas posteriormente, notou-se que as observações 6 e 30 correspondem aos meses de junho/99 e junho/2001, respectivamente.

No mês de junho/99 e junho/01, obteve-se uma redução na produção de água na ordem de 7,4% e 3,58%, respectivamente. Acompanhando os meses posteriores, está redução continua gradativamente, isso demonstra uma indicação que é início de inverno, onde as pessoas, consomem menos água, tomam menos banho, lavam menos roupas etc.

Diante desses fatos, a retirada dessas duas observações indicam que não tem indícios suficientes para serem retirada do modelo.

O modelo final fica: **Vol. Produzido** = $\beta_0 + \beta_1 \text{Economia} + \beta_2 \text{Temperatura}$

5.2.3. Inferência Sobre a Predição

Seria interessante dar um intervalo de confiança que descrevesse a precisão dessa predição. Para isso, devemos responder às seguintes questões antes de começar a calcular: Deseja-se prever o volume médio produzido de água (m^3) para todos os meses que tenha um número x de número de economias e a um temperatura indicada. Ou deseja-se prever o volume produzido de água (m^3) em um mês específico que tenha um determinado número de economias e/ou uma temperatura indicada.

Vimos que a reta de mínimos quadrados para prever o volume produzido de água ($m^3/mês$) com base nos dados

apresentados da cidade de Paranavaí é:

$$\text{Vol. Produzido} = \beta_0 + \beta_1 \text{Economia} + \beta_2 \text{Temperatura}$$

- Portanto para estimar a **resposta média**, utilizamos um intervalo de confiança. É um intervalo de confiança comum para o parâmetro

$$\mu_{\text{Vol. Produz.}} = \beta_0 + \beta_x^*$$

O modelo de regressão nos diz que $\mu_{\text{Vol. Produz.}}$ é a média das respostas do volume produzido quando o número de economias ou temperatura toma um valor denominado de X^* . É um número fixo, cujo valor não conhecemos.

- Para estimar uma **resposta individual**, ou seja, mensal, utiliza-se um intervalo da predição. Um intervalo de predição estima uma única resposta aleatória (volume produzido(m^3)), em vez de um parâmetro como $\mu_{\text{Vol. Produz.}}$. A resposta (volume produzido) não é um número fixo. Se fizer mais observações com $x=x^*$, obter-se respostas diferentes.

Para ambos os casos, o significado de um intervalo de predição é muito semelhante ao significado de um intervalo de confiança. Um intervalo de confiança de predição de 95%, assim como um intervalo de confiança de 95%, está correto 95% das vezes em uma aplicação reiterada. "Aplicação reiterada", agora significa que fazemos uma observação sobre volume produzido de água para cada um dos meses utilizando as variáveis independentes encontradas no modelo para este estudo (número de economias e temperatura), e fazemos então mais uma observação sobre o volume produzido nas variáveis independentes simuladas (valores simulados). Formando um intervalo de predição com base nas n observações (neste caso são os meses), devemos ver se ele abrange esse volume produzido de água adicional.

Isso ocorrerá em 95% de todas as repetições.

Eis os detalhes das fórmulas a serem utilizadas.

Para este estudo será deixada esta parte do conteúdo a título de ilustração de como simular uma predição.

A seguir as fórmulas para cada caso.

I.C. 95% para **resposta média**, volume produzido, quanto as variáveis independente (número de economias e temperatura) tem-se um valor denotado por x^* , é

$$\text{vol. prod.} \pm t^* \cdot EP_{\mu}$$

$$\text{O erro padrão } EP_{\mu} \text{ é } EP_{\mu} = s \sqrt{\frac{1}{n} + \frac{(x^* - \bar{X})^2}{\sum (x - \bar{X})^2}}$$

O somatório abrange todas as observações sobre a variável exploratória x .

I.C. de predição de 95% com um **única observação sobre volume produzido**, quando as variáveis independentes (número de economias e temperatura) toma o valor x^* , é

$$\text{vol. prod.} \pm t^* \cdot EP_{\mu}$$

$$\text{O erro padrão para a predição } EP_{\text{vol. prod.}} \text{ é: } EP_{\text{vol. produz}} = s \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{X})^2}{\sum (x - \bar{X})^2}}$$

Em ambos os casos a fórmula, t^* é o valor crítico $(1-0,95)/2$ superior da distribuição t com $n-2$ graus de liberdade, obtém-se esse valor através de tabelas estatísticas apropriadas.

Nota: A interpretação dos intervalos da predição é de importância secundária. O principal é que é mais difícil prever uma resposta do que prever uma resposta média.

Neste caso do estudo é muito mais indicado usar o intervalo de predição.

6. CONSIDERAÇÕES GERAIS

Embora o tratamento estatístico indicasse a relação **Vol. Produzido** = $\beta_0 + \beta_1 \text{Economia} + \beta_2 \text{Temperatura}$, como melhor modelo, isto poderia mudar completamente se fossem acrescentadas outras variáveis significativas na determinação da demanda de água.

Ao tratar do assunto em questão, o gestor deve ter uma visão global de um Sistema de Abastecimento de Água, tendo sempre como foco não apenas produzir uma determinada quantidade de água, mas também de identificar como essa água é demandada.

O modelo final obtido por qualquer um dos procedimentos deve ser analisado quanto ao seu sentido prático. Os profissionais que não tem experiência podem concluir por um modelo não realístico. Sugere-se portanto que o analista tenha sempre a frente alguns questionamentos a fazer tais como: o modelo obtido é razoável e faz sentido à luz do problema real? Vai ser usado para seu propósito pretendido, devido a altos custos para coleta de dados?

As recomendações para o profissional interessado em aplicar a presente metodologia, alguns procedimentos deverão ser seguindo a fim de que o modelo tenha grande eficiência.

Após análise efetuada e a escolha do modelo final ajustado deve-se verificar como este funciona na prática.

A validação fornece uma proteção tanto para o modelo desenvolvido quanto para o usuário deste. Alguns procedimentos são úteis para verificar a validade do modelo, os mais significativos são:

Análise dos coeficientes do modelo e dos valores preditos através de: comparações com experimentos anteriores (quando existirem), proceder algumas simulação, coleta de novos dados para verificar o desempenho preditivo do modelo e/ou a partição (split) dos dados. A partição consiste em deixar parte dos dados originais fora da análise para investigar o desempenho preditivo do modelo com a parte não utilizada no ajuste.

Nota: O autor acredita que se introduzisse uma variável relativa a renda familiar e outra relativa ao rendimento da medição, a equação para previsão de demanda fosse do tipo :

$$\text{Vol. Prod.} = \beta_0 + \beta_1 \text{Economia} + \beta_2 \text{Temperatura} + \beta_3 \text{Preço de Água} + \beta_4 \text{Renda Familiar} + \beta_5 \text{rendimento da Medição}$$

7. CONCLUSÕES

O estudo apresentado ao longo deste trabalho, fez uma abordagem da forma, de como se faz predição de um processo de um Sistema de Abastecimento de Água de pequeno porte.

Preocupou-se principalmente, em apresentar uma metodologia referente a maneira de, como a água é consumida na forma do abastecimento público de uma cidade, em benefício da saúde pública, do saneamento ambiental e da eficiência dos serviços prestados, propiciando a melhor produtividade dos ativos existentes, o planejamento e parte dos investimentos para ampliação dos sistemas. Definir e implementar um conjunto de ações e instrumentos tecnológicos, normativos, econômicos e institucionais, concorrentes para uma efetiva economia dos volumes de água demandados para consumo na áreas urbanas.

Para se buscar um maior entendimento de como a água é demandada, a primeira preocupação foi a de quantificar a água consumida por categorias e grupos de consumidores, que neste trabalho foi só apresentado para a categoria residencial, sendo esta a predominante em uma cidade, mas isso não implica que a mesma metodologia seja aplicada para outras categorias, porém com algumas restrições.

O segundo passo, foi identificar as principais variáveis independentes que foram determinantes no consumo de água de cada grupo de consumidores de água, quantificar esta demanda do sistema produtor e elaborar uma metodologia para se fazer prognóstico de demanda de um Sistema de Abastecimento de Água de pequeno porte.

A um nível de confiabilidade de 95% e baseado nas variáveis apresentadas, o modelo final para prognóstico de demanda de água futura para a cidade de Paranavai é: **Vol. Produzido** = $\beta_0 + \beta_1 \text{Economia} + \beta_2 \text{Temperatura}$),

A metodologia apresentada foi do ponto de vista sugestivo, sugestões de métodos e análises estatísticas, critérios estes, que são chaves de um bom prognóstico de previsões de demanda de água futura de um Sistema de Abastecimento de Água.

Nota: Foi tomado os somatórios da micromedição de água como consumo, mas na realidade esta medição pode estar contemplado de um erro decorrente do envelhecimento natural dos medidores de água e uma evolução negativa do rendimento da medição da cidade.

8. REFERÊNCIAS BIBLIOGRÁFICAS

1. AMERICAN WATER WORKS ASSOCIATION, Proceedings, Computer Conference 1992, Nashville, Tennessee.
2. BERENHAUSER, C.J.B; Pulici, CLOVIS, A Demanda Urbana de Água, Revista DAE, São Paulo, n.º 135, pg. 118, 1983
3. BILLINGS, Bruce; Jones, Vaughan, Forecasting Urban Water Demand, AWWA, Denver Colorado, 1996
4. BUSSAB, W. de Oliveira – Análise de Variância e Regressão, São Paulo, Editora Atual 2.Ed., São Paulo.
5. DOWNING, D, CLARK, Jeffrei, Estatística Aplicada, 2.ª Ed. Edit. Saraiva, São Paulo, 2002.
6. HOFFMANN, Rodolfo, Vieira, Sônia, Análise de Regressão Uma Introdução à Econometria, 3.º ed, Edit Hucitec, São Paulo, 1998
7. MONTGOMERY, Douglas C; Peck, Elizabeth A . Introduction To Linear Regression Analysis, 2ª Edition
8. SILVA, R.T; ROCHA, W.S., Caracterização da Demanda Urbana de Água-A3, PNCA, Brasília, 1999