

RECURSOS COMPUTACIONAIS DE APOIO À CONSTRUÇÃO DE MODELOS DE REGRESSÃO COM UTILIZAÇÃO DO SOFTWARE MICROSOFT EXCEL (Office 2007 – (Microsoft Co.).

Murolo, Afrânio Carlos

afraniamurolo@gmail.com

Professor do curso de Engenharia de Produção
Centro Universitário Padre Anchieta – Jundiáí- SP

Resumo

O objetivo deste trabalho é apresentar de forma clara e concisa a aplicação prática do tratamento estatístico e econométrico, tendo em vista a modelagem de regressão, por meio de caráter experimental para o aprimoramento da tendência de dados pesquisados em campo. Deve-se ressaltar que a construção dos modelos de regressão linear, exponencial, potência, polinomial e logarítmico, possibilitam através da ferramenta Excel a escolha de maneira rápida e eficaz do melhor modelo a ser incorporado por uma dada pesquisa, levando-se em conta o grau de ajustamento e a fidedignidade dos dados pesquisados.

Palavras-chaves: Regressão Linear, Exponencial, Polinomial, Potência, Logarítmica e coeficiente de Determinação.

ABSTRACT

The objective of this paper is to present in a concisely and clearly way the practical application of econometrical and statistical treatment, according to the modeling of regression, through experimental feature, leading to development of field surveying data. It should be noticed that the construction of linear regression, exponential, potency, polynomial and logarithmic models, allow through Excel tool a quick and effective choice of the best model to be incorporated at a given research, taking into account the adjustment degree and the reliability of researched data.

Key Words: Linear Regression, Exponential, Polynomial, Potential, Logarithmic, and coefficient of determination.

1- Introdução

Os modelos de regressões estendem-se em um horizonte de aplicações nas grandes áreas de concentração das pesquisas, possibilitam descrever e analisar uma grande gama de situações dentro do universo ou espaço amostral, dos quais os modelos de tendência poderão estar inseridos.

Dentre as várias aplicações de modelos de regressão nas engenharias, podemos ressaltar a análise de regressão para o consumo de energia elétrica frente a variáveis arquitetônicas para edifícios comerciais climatizados em 14 capitais brasileiras (Signor, 1999): proposta de um método para aplicação de gráficos de controle de regressão no

monitoramento de processos (Pedrini, 2009), previsão de carga multinodal utilizando redes neurais de regressão generalizada (Kenji,2011), uso da regressão linear para estimativa da relação entre a condutividade elétrica e a composição iônica da água de irrigação (Ribeiro; Maia; Medeiros, 2004), aplicação da regressão múltipla na identificação de variáveis relevantes no aumento dos custos operacionais em unidades operacionais de produção (Figueiras,2010) e desenvolvimento de gráficos de controle aplicados ao modelo funcional de regressão (Almeida,2003).

Neste trabalho, cabe salientar que a planilha Excel será de muita utilidade e nos facilitará substancialmente na construção dos modelos de regressão acima mencionados, como também trará um enorme ganho de tempo na construção gráfica e caracterização do modelo a ser adotado, dado a simplicidade de utilização de suas funções.

Nos estudos envolvendo modelos de regressão, temos como foco principal a construção gráfica de um sistema de dispersão das variáveis, x e y , oriundas de pesquisa, e sua melhor adaptação ao modelo matemático proposto, em função do gráfico do sistema de dispersão. Por outro lado, nesta etapa, a idéia principal é a busca ideal do melhor modelo, que mais se adapta ou se ajusta à variação dos dados observados da pesquisa, x e y , isto é, realizado usando o elemento ou dado estatístico denominado coeficiente de determinação, R^2 . Tal coeficiente nos indicará o grau de qualidade e ajustamento, quando R^2 tender a um ($R^2 \rightarrow 1$) o grau de qualidade e ajustamento será alto, em contrapartida, quando R^2 tende, ou fica cada vez mais próximo de zero ($R^2 \rightarrow 0$) o grau de qualidade e ajuste diminuí.

O cálculo deste coeficiente, o sistema de dispersão e a linha de regressão adotada, são gerados diretamente pelo Excel. É importante observar que, na prática, a comparação de vários modelos matemáticos de regressão, de forma simultânea, associados à mesma pesquisa de dados (x e y), nos permite através de R^2 encontrar e decidir pelo melhor ajuste (Murolo; Bonetto, 2011).

1.1- Método da construção e obtenção da Linha de Tendência, Sistema de dispersão, Modelo matemático de Regressão Linear e coeficiente de determinação.

Para exemplificar tal modelagem, faremos uso de uma pesquisa, denotada pela figura1, que relaciona uma série de tempo em meses, dada pela variável x , diretamente relacionada

com o suposto crescimento do valor de face de um título $y(x)$ em unidades monetárias. O problema sugere a utilização direta da planilha Excel para a construção do modelo $R.L.S.$, bastando para isso obedecer à seguinte sequência:

	A	B
1	x (meses)	y (u.m.)
2	1	1040,00
3	2	1071,20
4	3	1108,69
5	4	1086,52
6	5	1119,11
7	6	1169,47
8	7	1146,08
9	8	1134,62
10	9	1168,66
11	10	1192,04
12	11	1233,76
13	12	1264,60

Figura 1- Dados (x, y)

A fim de obtermos o sistema de dispersão correspondente, devemos proceder de acordo com a seleção mencionada abaixo, determinando a nuvem dos pontos pesquisados.

- Selecione as células **A2: B13** (de A2 até B13)
- Clique na guia **INSERIR** da faixa de opções e selecione a opção **DISPERSÃO** no grupo **Gráfico** (Figura 2).

Obs: Ao manter o mouse sobre a opção Dispersão sem clicar, aparecerá um comentário com uma rápida explicação.



Figura 2 – Selecionando a opção DISPERSÃO

- Selecione a opção de gráfico “Dispersão Somente com Marcadores”. (Figura 3)

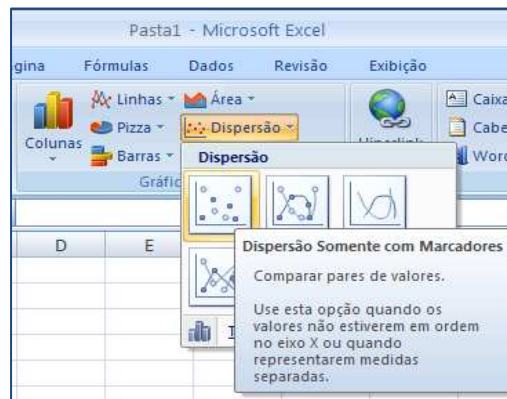


Figura 3 – Selecionando a opção Dispersão Somente com Marcadores

- Automaticamente aparecerá a área do gráfico na planilha. (Figura 4)

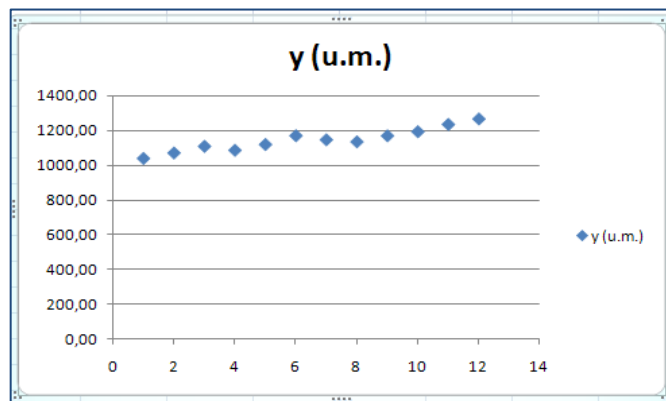


Figura 4 – Visualização do gráfico de dispersão

1.2 - Adicionando Linha de Tendência ao gráfico do tipo Linear

- Clique com o botão direito do mouse sobre qualquer ponto do gráfico e após escolha a opção *Adicionar Linha de Tendência...* (Figura 5)



Figura 5 – Adicionando Linha de Tendência

- Na caixa de diálogo Formatar Linha de Tendência, deixe selecionada a opção *Linear* e habilite a opção “*Exibir Equação no Gráfico*” e “*Exibir valor de R-quadrado no gráfico*”
- Clique no botão *Fechar*. (Figura 6)

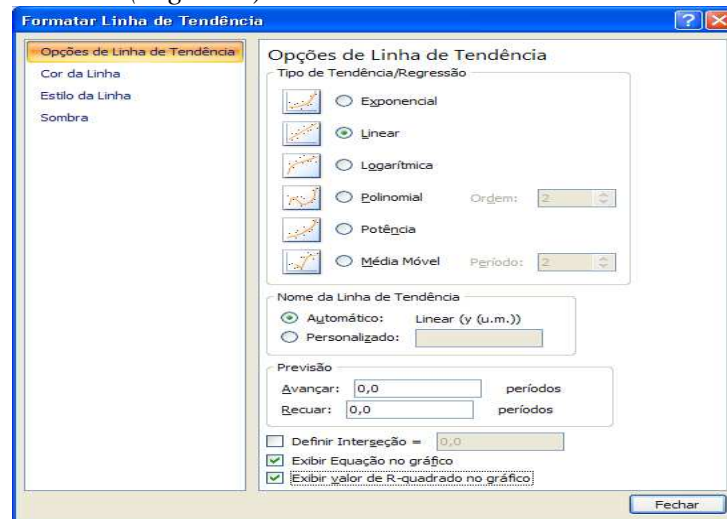


Figura 6 – Caixa de diálogo Formatar Linha de Tendência

- Observe na ilustração (Figura 7) o gráfico com a Linha de Tendência do tipo *Linear* e *R-quadrado*.

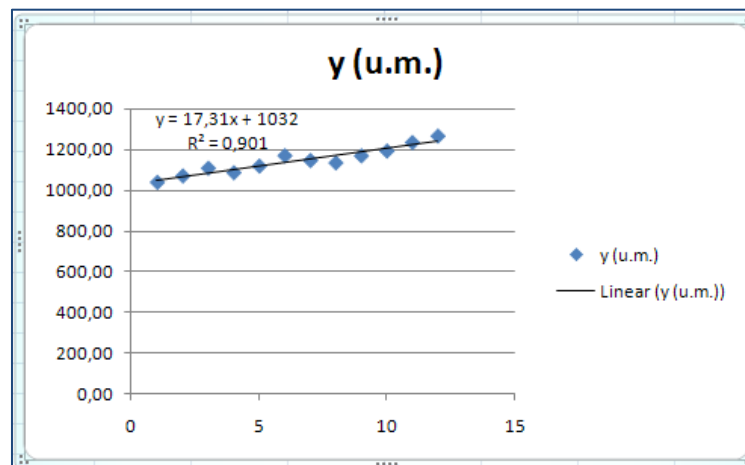


Figura 7 – Linha de Tendência Linear e R-quadrado

1.3 - Construção dos modelos de regressão, linhas de tendências e coeficientes de determinação.

Com o propósito de tomada de decisão para a escolha do modelo mais eficaz, utilizaremos a pesquisa tratando do modelo econômico da oferta de mercado (y) de um produto em 1.000 unidades, onde são reveladas as quantidades em que os produtores estariam dispostos a oferecer a vários níveis de preços (x) em unidades monetárias.

	A	B
1	Preço (x) em u.m.	Oferta (y) em 1.000 un.
2	10	100
3	12	120
4	14	135
5	15	167
6	17	198
7	19	220
8	21	268
9	23	310
10	26	390
11		

Figura 8 – Dados observados de Preço (x) e Oferta (y)

Nas ilustrações a seguir, faremos as construções e comparações entre os coeficientes de determinação R^2 dos modelos acima mencionados, mostrando: linha de tendência, diagrama de dispersão, equação de regressão e coeficiente de determinação para cada modelo adotado em referência à pesquisa adotada.

A ferramenta adotada possibilita construir os modelos citados no objetivo deste artigo, através do recurso “*adicionar linha de tendência*”, no entanto, como revelamos anteriormente, é importante refinar o processo de tomada de decisão, comparando vários modelos de regressão em função de seus respectivos coeficientes de determinação. Logo a decisão a ser tomada pelo melhor modelo, será determinada em função da regressão que apresentar maior coeficiente de determinação, R^2 .

1.3.1- Construção do Modelo - Regressão Exponencial

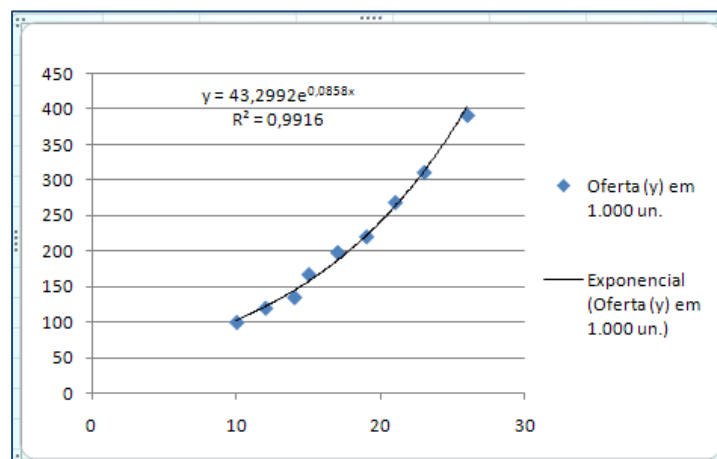


Figura 9 - Linha de Tendência Exponencial e R-quadrado

Ao exibir a Equação e R-quadrado no gráfico, na criação dos modelos, algumas vezes haverá a necessidade de aumentar casas decimais nos valores gerados. Para isso realize a seguinte sequência:

- Clique com o botão direito do mouse sobre os **valores da Equação e R-quadrado**. (Figura 10)

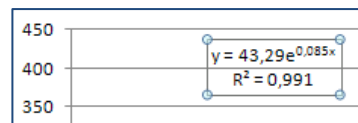


Figura 10 – Equação e R-quadrado

- Clique na opção **Formatar Rótulo de Linha de Tendência...** (Figura 11)

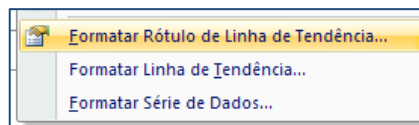


Figura 11 – Formatar Rótulo de Linha de Tendência

- Na caixa de diálogo **Formatar Rótulo de Linha de Tendência**, selecione a Categoria Número do item Número e digite o valor **4** (quatro) dentro da caixa **Casas decimais**.
- Clique no botão **Fechar**. (Figura 12)

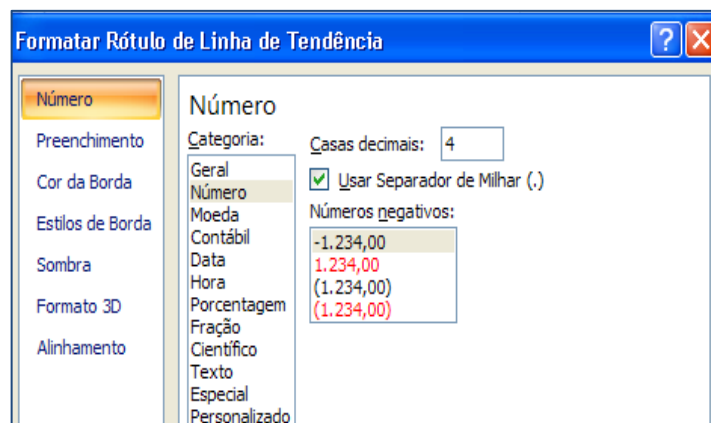


Figura 12 – Alterando o número de casas decimais

1.3.2- Construção Modelo - Regressão Potência

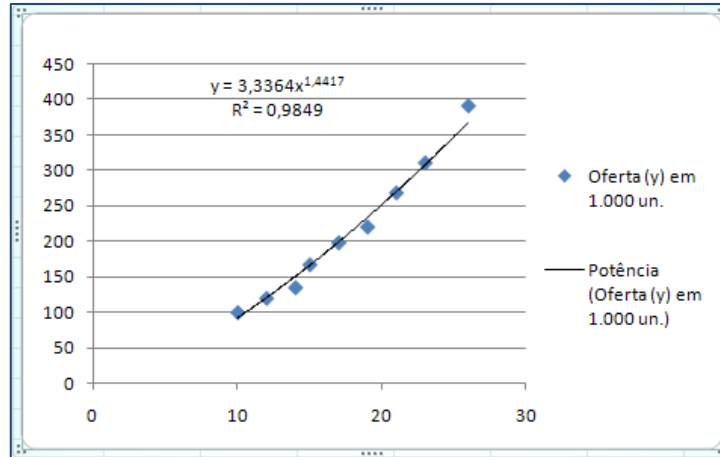


Figura 13 - Linha de Tendência da Potência e R-quadrado

1.3.3 - Construção Modelo - Regressão Polinomial

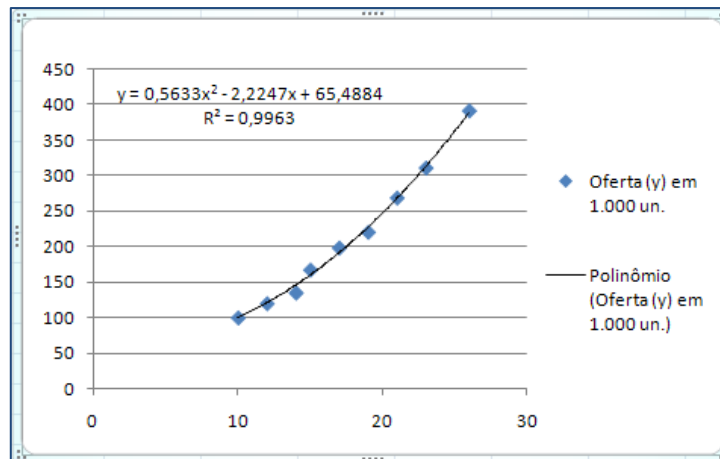


Figura 14 - Linha de Tendência da Polinomial e R-quadrado

1.3.4 - Construção Modelo - Regressão Linear

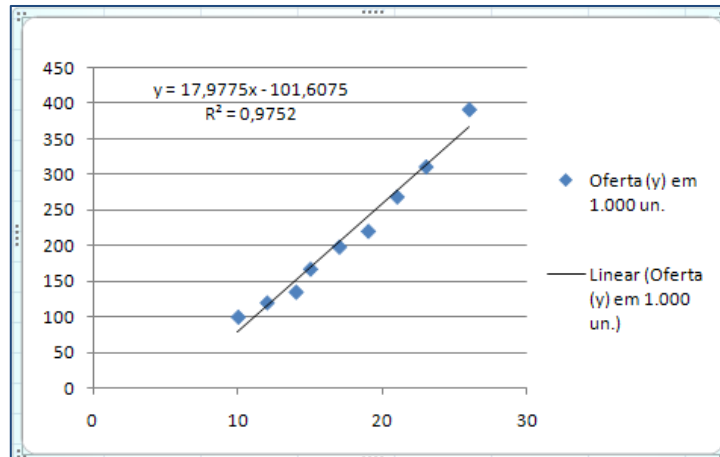


Figura 15 - Linha de Figura Tendência Linear e R-quadrado

1.3.5 - Construção Modelo - Regressão Logarítmica

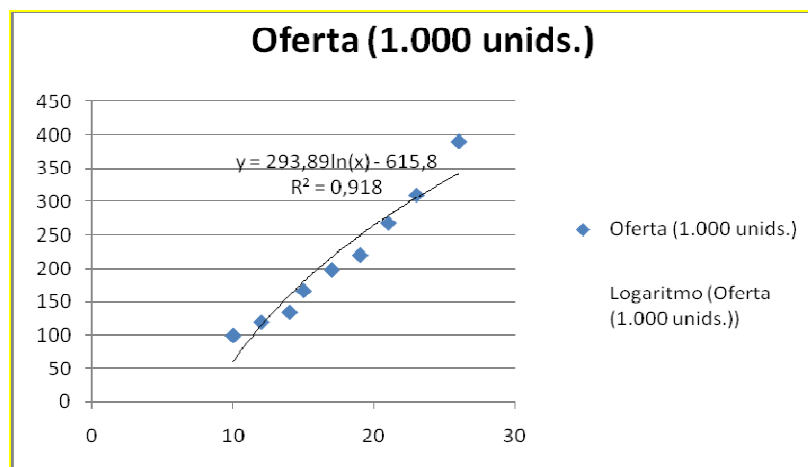


Figura 16 - Linha de Tendência Logarítmica e R-quadrado

2- Material e Métodos

O presente experimento foi conduzido por meio de dados levantados em parceria com a empresa de consultoria Sinapse Consultoria Tecnológica, obtidos diretamente da fonte produtora, no segmento de papel e celulose. Assim, foi determinado modelos estatísticos de regressão, cujo tratamento originou em tendências para o produto fabricado em relação ao horizonte da pesquisa.

A pesquisa está baseada em valores observados da oferta do produto, prevista e observada, relacionada com os preços observados, tendo ainda como foco a expectativa caracterizada

pela modelagem matemática que mais se adapta à variação dos dados levantados frente ao mercado consumidor.

Diante dos modelos teóricos de previsão baseados em Wooldridge (WOOLDRIDGE, 2010), Moreira (MOREIRA, 2008), Medeiros et al. (MEDEIROS, 2011) e DRAPER (DRAPER, 1981), foram construídos Modelos Estatísticos de Tendência de Mercado para análise e interpretação entre oferta projetada e observada e o sistema de dispersão de pontos pesquisados.

A coleta de material, ou dados estatísticos, foi obtida diretamente da Empresa Produtora tendo-se como base o histórico e a evolução da demanda do período considerado (MEDEIROS, 2011; GONÇALVES 2011; MUROLO, 2011). As técnicas utilizadas foram fundamentadas na Análise de Regressão, tomando como base modelos estatísticos e econométricos, testados estatisticamente (MCGUIGAN, 2010). A representatividade das amostras de informações coletadas, obtidas diretamente da fonte de dados reproduziu de forma altamente confiável as características importantes dos dados populacionais da variável demanda enfocada. Pode-se afirmar ainda, que o grau de confiabilidade em relação aos Modelos Estatísticos de Tendência, gerados a partir das amostras, além de passarem pelo crivo dos critérios de escolha e decisão sob o ponto de vista científico, tem grande qualidade no que tange à representatividade das amostras coletadas diante da abordagem do processo de estimação.

3- Resultados e Discussão

Verificando o quadro abaixo, temos uma noção geral do resumo da análise de tendência e sistemas de dispersão de todos os modelos gerados de acordo com pesquisa dos dados levantados:

Modelos	Equação	Coefficiente de Determinação, R^2
Linear	$y = 17,9775 x - 101,6075$	0,9752
Logarítmica	$y = 293,89 \ln(x) - 615,8$	0,918
Exponencial	$y = 43,2992 e^{0,0858x}$	0,9916
Potência	$y = 3,3364 x^{1,4417}$	0,9849
Polinomial	$y = 0,5633 x^2 - 2,2247 x + 65,4884$	0,9963

Fig. 17- Resumo da análise de tendência

Analisando os dados e resultados, resumido na figura 17, identifica-se o modelo matemático da R.L.S., determinado por $y = 17,9775 x - 101,6075$ e $R^2 = 0,9752$. Deste modo, a equação de regressão explica cerca de 97,52% da variação total de y (oferta em 1.000 unidades). Os restantes 2,48% são atribuídos aos fatores incluídos no termo erro. Poderemos ainda calcular o coeficiente de correlação linear $r_{x,y}$, sendo: $r = \sqrt{R^2}$ e $r = \sqrt{0,901} \cong 0,9492$ ou 94,92%. É importante salientar que $r_{x,y}$ é positivo, porque $a = 17,9775$ também é positivo, lembrando que a equação de regressão linear é caracterizada, por: $y = ax + b$.

Observa-se que o modelo linear se apresentou como um bom ajuste, indicando haver uma relação linear para a pesquisa, indicando R^2 próximo de 1. O modelo logarítmico apresentou bom ajuste, indicando $R^2 = 0,9180$, próximo de 1, com margem aproximada de 8,2%. Já os modelos Exponenciais, Potência e Polinomial apresentaram, respectivamente, coeficientes de determinação, R^2 , iguais a **0,9916; 0,9849 e 0,9963** com margens de erro de respectivamente: **0,84%, 1,51% e 0,37%**. Verific-se por meio dos resultados obtidos que o modelo **Polinomial** apresenta notadamente com maior coeficiente de determinação, indicando maior proximidade de um, e ainda se caracterizando por uma margem de erro substancialmente próxima de zero.

4- Conclusão

Analisando os coeficientes de determinação dos modelos matemáticos gerados para as curvas Exponencial, Logarítmica, Linear Potência e Polinomial, verificamos que $R^2(\text{logarítmica}) = 0,918 < R^2(\text{Linear}) = 0,9752 < R^2(\text{Potência}) = 0,9849 < R^2(\text{exp}) = 0,9916 < R^2(\text{polinomial}) = 0,9963$.

Os resultados obtidos acima nos permite concluir que o modelo **“regressão polinomial”** é o que mais se ajusta à variação dos dados pesquisados, por apresentar maior coeficiente de determinação, $R^2 = 0,9963$, cuja equação de regressão explica cerca de 99,63% da variação total de y (oferta em milhares de unidades) e os restantes 0,37% são atribuídos aos fatores incluídos no termo erro, caracterizando-se em uma excelente modelagem para os dados pesquisados. Fica claro, que mediante a escolha do melhor modelo matemático, frente à pesquisa adotada, que ocorre o aprimoramento do grau de confiabilidade e minimização da margem de erro de futuras projeções da função oferta em unidades diante das possíveis

variações de preços em unidades monetárias. Somente usando a análise de tendência, pode-se aceitar o melhor modelo para ajustar os dados pesquisados, encontrando a melhor a tendência e menor margem de erro nas futuras previsões. Com este caráter simples, mas substancial é que se propõe a junção da teoria e a prática da modelagem, tornando a tomada de decisão cada vez mais confiável, dentro do universo e espaço amostral em que se insere a referida pesquisa.

Agradecimentos

A Sinapse Consultoria Tecnológica e aos colegas colaboradores Rafael A. Murolo (FEA-USP/RP) e ao Engenheiro Adênio C. Junior (Sinapse consultoria/informática).

Referências Bibliográficas

ALMEIDA, S.S. Desenvolvimento De Gráficos De Controle Aplicados Ao Modelo Funcional De Regressão. 2003; Tese (Doutorado em Engenharia de Produção) - Escola de Engenharia, Universidade Federal de Santa Catarina; Florianópolis; SC; **2003**.

CHAMBERS, J.C.; MULLICK, S. K.; SMITH, D.D.; How To Choose The Right Forecasting Technique; Harvard Business Review; Jul. / Ago **1971**.

DRAPER, N.; SMITH, H.; Applied Regression Analysis; 2.ed.; Wiley; Nova York; **1981**.

DA SILVA, E.M.; GONÇALVES, V.; MUROLO, A. C.; Estatística Para Os Cursos De Administração E Ciências Contábeis; 4 ed. v2; Atlas, São Paulo; SP; **2011**.

FILGUEIRAS, V M.; Aplicação Da Regressão Múltipla Na Identificação De Variáveis Relevantes No Aumento Dos Custos Operacionais Em Unidades Estacionárias De Produção; Monografia em Engenharia de Produção; Faculdade Salesiana Maria Auxiliadora; Macaé; RJ; **2010**.

KENJI, N.F.; Previsão De Carga Multinodal Utilizando Redes Neurais De Regressão Generalizada; Dissertação (Mestrado em Engenharia Elétrica)-Universidade Estadual Paulista, Ilha Solteira, SP, **2010**.

KOENKER, R.; Quantile Regression. Cambridge; Cambridge University Press, **2005**.

MOREIRA, D.A.; Administração Da Produção E Operações; 2. ed.; Cengage Learning; São Paulo; SP; **2008**.

MCGUIGAN, J.R.; MOYER, R.C.; HARRIS, F.B.; Economia De Empresas-Aplicações, Estratégia E Táticas; Cengage; ; São Paulo; SP; **2010**.

MUROLO, A.C.; BONETTO, G.A; Matemática Aplicada A Administração, Economia E Ciências Contábeis. 2 ed. Cengage Learning; São Paulo; SP; **2011**.

PEDRINI, D.C.; Proposta De Um Método Para Aplicação De Gráficos De Controle De Regressão No Monitoramento De Processos; Dissertação (Mestrado em Engenharia de Produção)- Universidade Federal do Rio Grande do Sul, Porto Alegre, RS; **2009**.

RGSDALE, C.T.; Spreadsheet Modeling E Decision Analysis. Cincinnati ed.; South-Western Thomson Learning; **2011**.

SIGNOR, R.; Análise De Regressão Para O Consumo De Energia Elétrica Frente A Variáveis Arquitetônicas Para Edifícios Comerciais Climatizados Em 14 Capitais Brasileiras; Dissertação (Mestrado em Engenharia Civil) – Universidade Federal de Santa Catarina; SC; **1999**.

STEVENSON, W.J.; Administração Das Operações De Produção; 6. ed.; LTC; RJ; **2002**.

WOODRIDGE, J.M.; Introdução À Econometria Uma Abordagem Moderna; 4 ed; Cengage Learning; São Paulo: **2010**.