

## **Na busca pelo melhor método de combinação de previsões por meio do uso de análise de correlação e análise de cluster**

Liane Werner (Universidade Federal do Rio Grande do Sul) [liane@producao.ufrgs.br](mailto:liane@producao.ufrgs.br)

### **Resumo:**

Realizar previsões é uma atividade que permite a empresa dimensionar todos os setores da empresa para bem servir aos clientes, como minimizar gastos desnecessários. Sendo assim, os previsores buscam previsões acuradas. Este artigo se propõe apresentar uma forma de como realizar a investigação para comparar o desempenho de previsões de demanda. Para tanto, serão usados dados do estudo de caso realizado por Werner (2005), referentes a previsão de demanda de uma empresa de assistência técnicas de computadores pessoais. Serão comparados quatro métodos de combinação, com base no utilizando análise de correlação e análise de cluster. Ao realizar a comparação dos métodos de combinação da variância mínima considerando ou não a correlação, pela média e pelo mínimos quadrados ordinários, chegou-se a conclusão que a análise de correlação e a análise de cluster auxiliam na escolha do método mais acurado.

**Palavras chave:** combinação de previsões, análise de cluster, forecasting.

## **Searching for the best method of combining forecast through the use of correlation analysis and cluster analysis**

### **Abstract**

Make predictions is an activity that allows the company scale all sectors of the company to serve customers well, how to minimize unnecessary expenditures. Thus, forecasters seek accurate forecasts. This article intends to present a way of how to perform research to compare the performance of demand forecasts. Therefore, will be used data from a case study conducted by Werner (2005), concerning the demand forecast of a technical service company of personal computers. Four methods of combination are compared, using analysis based on correlation and cluster analysis. By performing a comparison of methods: combining the minimum variance with or without correlations, the mean and OLS, came to the conclusion that the correlation analysis and cluster analysis help to choose the most accurate method.

**Key-words:** Article, combining forecast, cluster analysis, forecasting

---

## 1. Introdução

As empresas são impulsionadas pela competitividade gerando o desenvolvimento de novas tecnologias e o aprimoramento das já existentes. Buscando expandir sua fatia de mercado e assim obter resultados mais promissores, frequentemente, as empresa realizam diagnósticos e análises de melhorias, visando atender às necessidades dos consumidores com o menor custo possível. Porém, não é suficiente atender às dimensões da qualidade percebidas pelo consumidor, para manter-se no mercado também se faz necessário o dimensionamento adequado da produção e da cadeia logística por meio de previsões.

Visto que o procedimento de realizar previsões envolve incerteza, é preciso que uma gama variada de técnicas seja utilizada, para que por meio das medidas de acurácia selecione-se aquela que melhor represente o conjunto de dados.

Como cada técnica de previsão (individual) pode captar diferentes aspectos dos dados, no momento de realizar a previsão, uma opção é incorporar as diversas previsões em uma nova previsão (ARMSTRONG, 2001; WERNER, 2005). Uma forma de agregar os mais diferentes aspectos dos dados, obtidos através de diferentes técnicas de previsão, é o método conhecido como combinação de previsões (WEBBY; O'CONNOR, 1996).

Desde o artigo seminal de combinação de previsões, de autoria de Bayes e Granger em 1969 até os dias atuais, muitos estudos buscam pela forma mais adequada de realizar a previsão, tanto em termos de acurácia, quanto em termos de complexidade (COSTANTINE; PAPPALARDO, 2010; CHAN et al, 2004; MENEZES et al, 2000; GRANGER; RAMANATHAN, 1984; BATES; GRANGER, 1969).

Pelo exposto anteriormente e pelo fato de que, conforme Abraham e Ledolter (2005), o desejo de previsor é sempre obter uma melhor acuracidade, este artigo se propõe apresentar uma forma de como realizar a investigação para comparar o desempenho de previsões de demanda. Para tanto, serão usados dados do estudo de caso realizado por Werner (2005), referentes a previsão de demanda de uma empresa de assistência técnicas de computadores pessoais. Serão comparados quatro métodos de combinação, com base no utilizando análise de correlação e análise de cluster.

## 2. Referencial teórico

Nesta seção serão apresentadas três subseções. A primeira se refere aos conteúdos relacionados com a combinação de previsões, focando os métodos a serem abordados neste estudo e também medidas de acurácia pertinentes. Na segunda subseção a análise de correlação e por fim a análise de cluster.

### 2.1. Métodos de Combinação e medidas de acurácia

No processo de previsão, a prática usual é determinar a melhor, entre algumas opções, por meio de alguma medida de acurácia. A melhor previsão é, então, aproveitada, e as outras, descartadas. Através do descarte de previsões classificadas como inferiores, muitas vezes, alguma informação útil é perdida. Logo, ao invés de escolher uma única técnica de previsão, parece razoável considerar informações provenientes de várias técnicas e combiná-las. Como ilustrado na figura 1, a combinação de previsões consiste em gerar, com base em dados históricos, um modelo e após obtém-se uma previsão objetiva. Paralelamente, aplica-se outra técnica de previsão, em geral uma análise subjetiva dos dados históricos, de onde obtém-se um previsão subjetiva. Estas previsões são, então, combinadas, gerando a previsão final (WEBBY; O'CONNOR, 1996).

---

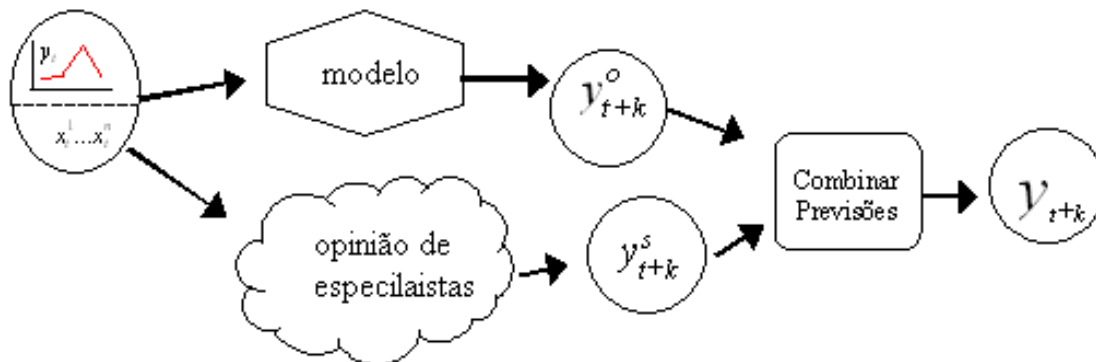


Figura 1 – Combinação de previsões [fonte: Adaptado de Webby; O'Connor, 1996]

Os precursores desse assunto foram Bates e Granger (1969). Eles propuseram que um método de combinar as previsões deveria constar de uma combinação linear de duas previsões objetivas, dando peso  $w$  para a primeira e peso  $(1-w)$  para a segunda, conforme a equação (1) que fornece a previsão combinação  $F_c$ .

$$F_c = wF_1 + (1 - w)F_2 \quad (1)$$

onde:  $w$  é o peso da previsão 1 e  $F_1$  e  $F_2$  são as previsões a serem combinadas.

A proposição destes autores é que, a previsão que possui os menores erros deveria dar um peso maior para a previsão (combinada). Sendo assim, propuseram a minimização da variância dos erros da previsão combinada, que ocorre quando  $w$  assume o valor dado pela equação (2). Por minimizar a variância, esse método ficou conhecido como método de variância mínima.

$$w = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \quad (2)$$

onde:  $\sigma_1^2$  e  $\sigma_2^2$  são as variâncias dos erros das previsão a serem combinadas;  $\rho$  é o coeficiente de correlação entre os erros das previsões;  $w$  é o peso dado a previsão 1.

Para o caso em que os erros não estão correlacionados ( $\rho = 0$ ),  $w$  fica reduzido ao valor dado pela equação (3).

$$w = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad (3)$$

Granger e Ramanathan (1984), chamaram a atenção para o fato de que os métodos convencionais de combinação de previsões poderiam ser vistos como uma forma estruturada de regressão. Eles argumentam que os métodos são equivalentes ao Método de Mínimo Quadrados Ordinários (MQO), tendo a previsão combinada como variável resposta e as previsões individuais como variáveis explicativas.

Face a muitas controvérsias sobre qual o melhor método, às dificuldades na estimação dos pesos e a complexidade destes e outros métodos (de combinação), a combinação por média aritmética, além de ser um dos métodos mais conhecidos é fácil de calcular. Para Menezes et al. (2000), uma possível resposta para o sucesso da média pode estar associada à instabilidade dos pesos que, frequentemente, resultam em mudanças que ocorrem ao longo do tempo na

matriz de covariância dos erros das previsões individuais. Devido a estas circunstâncias, a média, embora não tenha pesos ótimos, pode dar origem a resultados melhores que os de métodos mais sofisticados.

Porém para escolher qual a melhor combinação é preciso medir a acurácia. De acordo com Madridakis *et al.* (1998), em muitas situações a palavra acurácia refere-se à qualidade do ajuste, isto é, ao quão hábil é a técnica ou o modelo para reproduzir os dados que já são conhecidos. Entre as medidas de acurácia, mais conhecidas tem-se os autores abordam: ME (erro médio); MAE (erro absoluto médio); MSE (erro quadrático médio), sendo que as duas últimas consideram o erro absoluto de previsão. Outra medida muito utilizada é o MAPE (erro percentual absoluto médio) que é uma medida relativas. Cabe salientar que é proveitoso calcular mais que um critério de comparação, visto que, ocasionalmente, diferentes critérios fornecem indicações diferentes.

## 2.2 Análise de Correlação

Conforme Montgomery et al (2012) a análise de correlação resume o grau de relacionamento entre duas ou mais variáveis quantitativas. A forma mais básica de verificar esta relação é por meio do diagrama de dispersão. Outra forma de avaliar a correlação é através de um coeficiente, no caso de relação linear, utiliza-se o coeficiente linear de Pearson, conforme a equação (4).

$$\rho = \frac{Cov(x, y)}{\sigma_x \cdot \sigma_y} \quad (4)$$

Este coeficiente é útil para o caso que se deseja medir a intensidade da associação existente entre duas variáveis quantitativas independente da unidade de medida de cada variável.

O coeficiente de correlação fornece informação através do sinal, se  $\rho$  for positivo, existe uma relação direta entre as variáveis (valores altos de uma variável correspondem a valores altos de outra variável); se  $\rho$  for negativo a relação é inversa (valores altos de uma variável correspondem a valores baixos de outra variável). Agora se  $\rho$  for próximo de nulo, significa que não existe correlação linear.

## 2.3 Análise de Cluster

Também conhecida como análise de agrupamento, é uma técnicas para desenvolver subgrupos significativos de itens (indivíduos ou objetos). O objetivo específico é classificar (e agrupar) uma amostra de itens em um pequeno número de grupos mutuamente excludentes (HAIR et al, 2007). A finalidade desta técnica é reunir itens semelhantes dentro do grupo e diferentes de outros grupos (heterogeneidade)

Segundo Vicini (2005) na análise de agrupamento é importante escolher adequadamente o coeficiente que quantifique o quão parecidos são os itens. Esse coeficiente pode ser de similaridade ou dissimilaridade entre os itens a serem agrupados. Na medida de similaridade, quanto maior for seu valor, mais parecidos serão os itens. Já para a medida de dissimilaridade, quanto maior seu valor, menos parecidos são os itens.

A análise de agrupamentos realizada de forma hierárquica tem como resultado um dendograma como a figura 2 e o modo mais utilizado para decidir o número de grupos é a escolha do pesquisador.

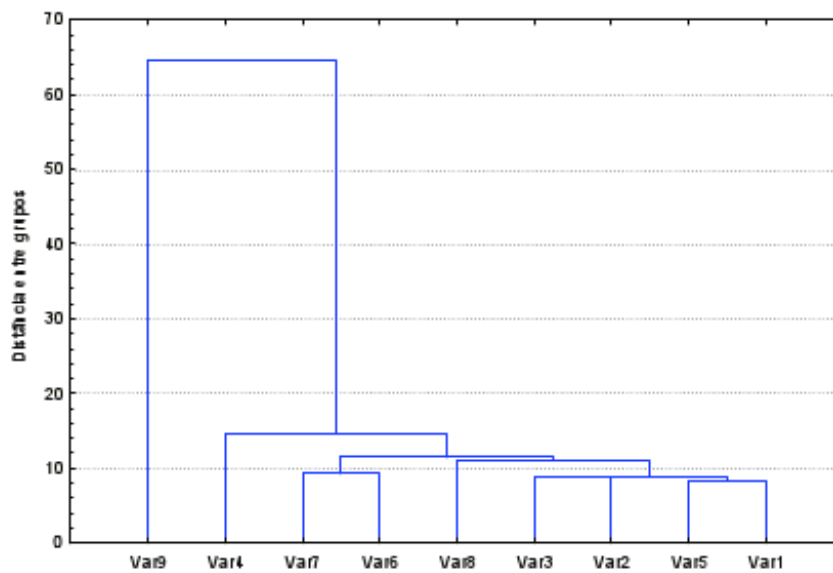


Figura 2 – Dendrograma genérico

### 3. Método proposto

Para realizar este estudo foi necessário primeiramente revisar os principais tópicos que compõe este artigo: combinação, correlação e análise de cluster. Uma vez, que ao obter previsões, o previsor almeja sempre uma alta acuracidade, a proposição de constituir combinação de previsões ganhou força como forma de realizar tal tarefa. Porém ainda se busca por uma resposta de qual é o método mais adequado. Posto isto, este artigo apresenta uma maneira de investigar qual o método de combinação a ser utilizado.

Utilizando os dados da demanda por serviços de assistência de computadores pessoais, quatro métodos de combinação serão comparados, a saber: média, variância mínima considerando ou não os erros correlacionados e por MQO, após obtidas as previsões e os respectivos erros por estes métodos, procede-se o cálculo do coeficiente de correlação de Pearson com os erros obtidos com cada método. Para corroborar com o que indicado na matriz de correlação, se constrói um dendrograma com os erros. Os resultados são comparados com a forma tradicional de escolher o método mais acurado, que é obter e escolher as medidas de acurácia mais altas.

### 4. Caso prático

Na área de assistência técnica de computadores, prever a demanda dos serviços está intimamente ligado às tarefas de disponibilizar recursos, tanto humanos como de equipamentos, para viabilizar o atendimento eficaz aos clientes. Tal fato indica que o planejamento da empresa precisa ser realizado em períodos curtos. Assim, a previsão de demanda será realizada em períodos mensais. A empresa, de pequeno porte, atua atendendo a três tipos diferenciados de clientes - contratos, garantia ou avulsos - cada tipo de cliente possui peculiaridades, e por esse motivo, foi necessário modelar e prever a demanda separadamente. Neste estudo será utilizado somente os dados dos clientes do tipo avulsos. Os clientes avulsos representam o maior desafio no esforço de previsão, pois o seu comportamento apresenta maior variabilidade, sendo influenciado de forma mais intensa pela conjuntura econômica e pelo cenário competitivo. Os valores observados encontram-se na figura 3.

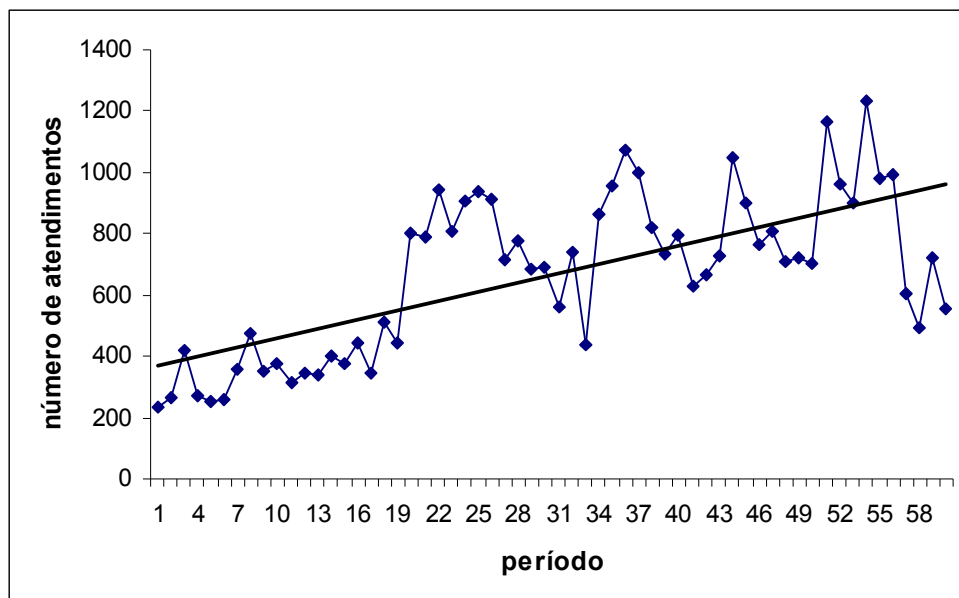


Figura 3 – Número de atendimentos para clientes do tipo avulsos e sua tendência [Fonte: dados fornecidos pela gerência da empresa em estudo]

Conforme os relatos do estudo apresentado por Werner (2005), as previsões (individuais), a serem utilizadas nas combinações, consistem de um modelo ARIMA(1,1,0), previsão  $F_1$ , e a previsão  $F_2$ , um modelo econométrico, suportado pelas variáveis: valor médio do dólar ( $X_1$ ); variação acumulada do PIB industrial ( $X_2$ ) e preço deflacionado pelo IPC-IEPE (divido por 30) ( $X_3$ ), sendo apresentado na equação (5)

$$Y = -509 - 854,2 X_1 + 480 X_2 + 618,1 X_3 \quad (5)$$

De posse dos valores observados e como os modelos estabelecidos, foi possível obter, ao longo do tempo, as previsões para estes dois modelos. O passo seguinte foi realizar a combinação das previsões pelos 4 métodos deste estudo. No quadro 1 encontram-se as equações de como foram obtidas as previsões das combinações.

Método	Equação da combinação
Média Aritmética	$F_c = 0,5F_1 + 0,5F_2$
Variância Mínima (sem correlação)	$F_c = 0,5194F_1 + 0,4806F_2$ (eq. 3)
Variância Mínima (com correlação)	$F_c = 0,5449F_1 + 0,4551F_2$ (eq. 2)
Mínimo Quadrados Ordinários	$F_c = 231,058 + 0,163F_1 + 0,497F_2$

Quadro 1 – Métodos de combinação e suas equações

Para cada um destes métodos, foi obtido o erro de previsão ao longo do tempo que a série foi analisada. De posse dos erros foi então realizado a análise de correlação, como pode ser observado na tabela 1, as correlações são todas significativas e somente o método MQO tem uma correlação fraca com os demais métodos indicando que o MQO deva ter um comportamento diferenciado. Para corroborar este resultado procedeu-se com a análise de cluster.

		média	VarMin_sem	Var Min_com	MQO
média	Corelação Pearson	1	1,000**	,999**	,578**
	p-value		,000	,000	,000
	N	59	59	59	59
VarMin_sem	Corelação Pearson	1,000**	1	1,000**	,578**
	p-value	,000		,000	,000
	N	59	59	59	59
VarMin_com	Corelação Pearson	,999**	1,000**	1	,579**
	p-value	,000	,000		,000
	N	59	59	59	59
MQO	Corelação Pearson	,578**	,578**	,579**	1
	p-value	,000	,000	,000	
	N	59	59	59	59

\*\* Correlação é significativa ao nível de 0,01 (bilateral).

Tabela 1 – Matriz de correlações lineares entre os métodos de previsão

A análise de cluster foi realizada no software SPSS utilizando a distância euclidiana como forma de dissimilaridade e obteve-se como resultado o dendodograma da figura 2. Ao observa esta figura confirma-se, pelo fato que do método MQO juntar-se muitos passos depois aos demais métodos que ele é diferenciado

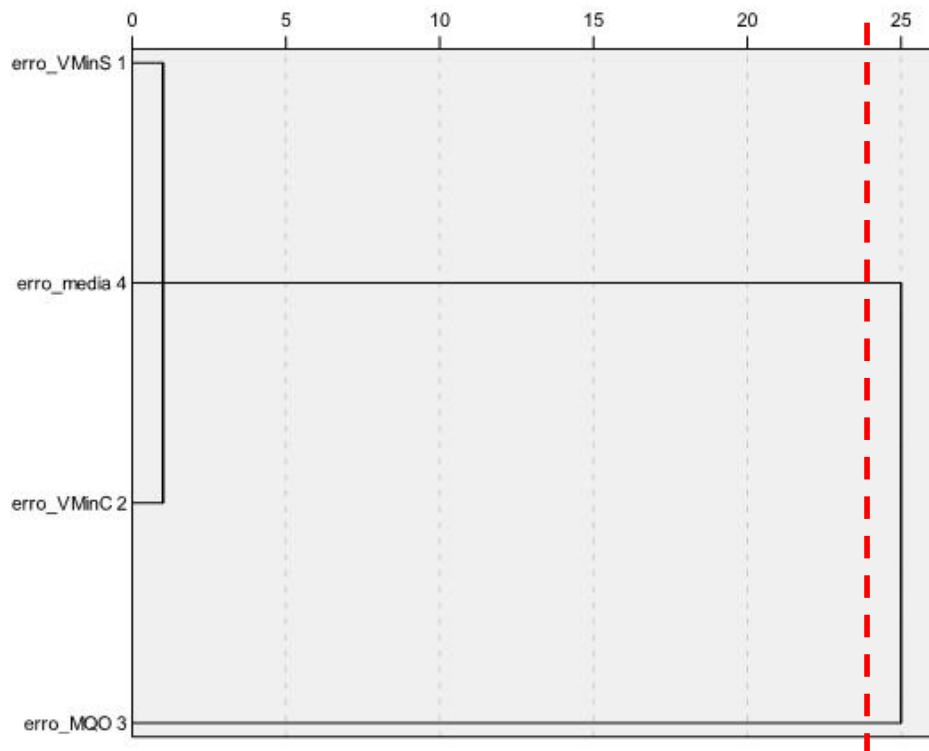


Figura 4 – Dendograma para os 4 métodos de combinação



Por fim, calculam-se as medidas de acurácia: ME, MSE, MAE, MAPE para cada um dos métodos de combinação, que se encontram na tabela 2. Analisando estas medidas verifica-se que o método MQO é o mais acurado e deve ser escolhido para realizar as previsões futuras, para o número de atendimentos avulsos.

	ME	MSE	MAPE	MAE
média	2,03	20064,39	0,182	114,54
Var Min_sem	1,95	20033,75	0,181	114,08
Var Min_com	1,84	20018,97	0,180	113,46
MQO	0,24	6727,99	0,133	68,28

Tabela 2 – Medidas de acurácia para os métodos de combinação

## 5. Considerações finais

Utilizar os resultados da análise de correlação e da análise de cluster é a forma de investigar proposta, pois visa corroborar com os resultados das medidas de acurácia. Neste estudo ficou claramente exemplificado que o método MQO é o mais acurado, tendo em vista que para este método todas as medidas foram menores e na correlação este método apresentou correlação fraca com os demais métodos. Além disto, na análise de cluster o MQO foi o último a se agrupar.

Esta proposição poderá auxiliar em situações em que as medidas de acurácia não são todas melhores em um único método, como neste estudo, Sendo assim fica a contribuição para buscar o método mais acurado, a fim de que se possam gerar boas previsões no futuro.

## Referências

- ABRAHAM, B.; LEDOLTER, J. *Statistical Methods for Forecasting*. New York: John Wiley & Sons, 2005.
- ARMSTRONG, J. S. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers, 2001.
- BATES, J. M. & GRANGER, C. W. J. *The Combining of Forecasts*. *Operational Research Quarterly*, v.20, n.4., p. 451-468, 1969.
- CHAN, C. K.; KINGSMAN, B. G.; WONG, H. *Determining when to update the weights in combined forecasts for product demand: an application of the CUSUM technique*. *European Journal of Operational Research*. v. 153, p. 757-768, 2004.
- COSTANTINE, C.; PAPPALARDO, C. *A hierarchical procedure for combination of forecasts*. *International journal of forecasting*. v. 26, p. 725-743, 2010.
- GRANGER, C. W.J. & RAMANATHAN, R. *Improved Methods of Forecasting*. *Journal of Forecasting*, v.3, p.197-204, 1984.
- MAKRIDAKIS, S., WHEELWRIGHT, S. C. & HYNDMAN, R. J. *Forecasting. Methods and Applications*. Third Edition. New York John Wiley & Sons, 1998.
- MENEZES, L. M.; BUNN, D. W.; TAYLOR, J. W. *Review of guidelines for the use combined forecast*. *European Journal of Operational Research*. v. 120, p. 190-204, 2000.
- MONTGOMERY, D.; PECK, E.; VINING, G.G. *Introduction to Linear Regression Analysis*. New York: John Wiley & Sons., 2012.
- VICINI, L. *Análise multivariada da teoria à prática*. Monografia de Especialização. Universidade Federal de Santa Maria, 2005.
- WEBBY, R. & O'CONNOR, M. *Judgement and Statistical Time Series Forecasting: a Review of the Literature*. *International Journal of Forecasting*, 12, p.91-118, 1996,



**WERNER, L.** *Um Modelo Composto para Realizar Previsão de Demanda através da Integração da Combinação e de Previsões e Ajuste Baseado na Opinião.* Tese de Doutorado. Universidade Federal do Rio Grande do Sul, 2005.

---