

Alexandre Moreira Batista
Higor Fernando Manginelli

PROJETO DE TRABALHO DE CONCLUSÃO DE CURSO

Trabalho apresentado para a disciplina
Laboratório de Estatística II do curso
de graduação em Estatística da
Universidade Federal do Paraná.

Curitiba
2008

**PREVISÃO DE LUCRO BRUTO PARA UMA COMPANHIA BRASILEIRA
DE ALIMENTOS UTILIZANDO MODELOS DE SÉRIES TEMPORAIS**

Alexandre Moreira Batista
Higor Fernando Manginelli

Curso de Estatística
Universidade Federal do Paraná

2008

Resumo

As tomadas de decisão em estratégias comerciais têm impacto direto no resultado das empresas do ramo alimentício no Brasil. obter informações confiáveis e com antecedência auxiliam nos processos decisórios. Em cada demonstrativo de resultado, uma das principais linhas é o Lucro Bruto. O objetivo deste trabalho é gerar previsões com embasamento estatístico para os volumes de venda, Preço e Custos mensais de uma grande empresa do ramo alimentício, que será denominada Empresa ABC. A principal atividade da companhia é Frango, desta forma foi realizada a análise estatística necessária para aplicação de modelos de previsão de séries temporais lineares e não-lineares para esta atividade. Primeiramente serão testados modelos lineares da família ARIMA. Na seqüência serão gerados os modelos da família STAR e finalmente os modelos de RNA – Redes Neurais Artificiais. Após os modelos gerados, eles serão comparados à medida de Precisão na Previsão. Serão desconsideradas algumas premissas estatísticas, como de heterocedasticidade dos erros seriais, em prol de uma melhor previsão.

Palavras-Chave: Séries Temporais, Arima, Star, Rna.

1 Introdução

Neste trabalho foram consideradas as informações de vendas uma grande empresa do ramo alimentício, que será denotada por Empresa ABC, por motivos de confidencialidade.

Estas informações são:

- Volume de Vendas;
- Preço Bruto;
- CPV-Direto (Custo Direto com Produto Vendido);
- CPV-Indireto (Custo Indireto com Produto Vendido);

Os produtos da empresa ABC são separados em categorias, chamadas de atividades, conforme a matéria-prima que os compõem e o tipo de preparo destes alimentos. Desta forma, os dados foram separados em atividades: Frango, Industrializados, Margarinas, Peru e Suíno.

1.1 O Problema

O problema consiste em prever o Lucro Bruto da Atividade Frango Companhia ABC, considerando as vendas no Mercado Interno e Externo.

O Lucro Bruto consiste em:

$$LB = (ROB) - (CPV-Direto) - (CPV-Indireto)$$

Onde:

$$ROB = (\text{Volume de Vendas}) * (\text{Preço Bruto})$$

1.2 Hipóteses de Pesquisa

Prever resultados financeiros futuros da Companhia ABC, considerando os dados históricos das séries descritas acima.

2 Objetivos

Com base nos dados reais de vendas entre os meses de Janeiro de 2002 e Março de 2008, o objetivo deste estudo consiste em analisar o comportamento destes dados e realizar previsões para os próximos três meses, atualizando a cada novo resultado divulgado do mês, tendo sempre a previsão para os três próximos meses.

2.1 Objetivos Gerais

Desenvolver ferramentas de previsão com embasamento estatístico, para auxiliar o planejamento estratégico da Companhia ABC.

2.2 Objetivos Específicos

Agregar conhecimento técnico em modelos de previsão de Séries Temporais aos envolvidos no trabalho, bem como divulgar à comunidade científica as aplicações destes modelos em dados economico-financeiros.

3 Justificativa

A Empresa ABC dispõe hoje de modelos de previsão de vendas sem embasamento estatístico, com este trabalho busca-se preencher esta lacuna.

4 Metodologia

4.1 Metodologia Estatística de Séries Temporais

4.1.1 Definição de Séries Temporais

Uma série temporal pode ser considerada como um conjunto de observações ordenadas no tempo. Podem-se ter séries contínuas, como o registro de temperaturas em Curitiba ou o registro de marés em Paranaguá, e séries discretas como número anual de vestibulandos da UFPR ou os Índices Diários da Bolsa de São Paulo.

Muitas vezes uma série discreta é obtida através da amostragem de uma série contínua. Por exemplo, o registro de temperaturas em Curitiba observada no tempo $[0, T]$, pode-se tomar N intervalos de tempo, onde:

$$N = \frac{T}{\Delta} \quad (1)$$

Há basicamente dois enfoques na análise de séries temporais, sendo que em ambos o objetivo é o mesmo, modelar a série temporal. No primeiro a análise é feita no *domínio temporal* e os modelos propostos são *modelos paramétricos* (com um número finito de parâmetros). No segundo, a série é conduzida no *domínio de frequências*, e os modelos são *não-paramétricos*.

4.1.2 Objetivos da Análise de Séries Temporais

Tendo então uma série $Z(t_1), \dots, Z(t_n)$, observada nos instantes t_1, \dots, t_n , pode-se estar interessado em:

- (a) investigar o processo gerador da série;
- (b) fazer previsões de valores futuros da série;
- (c) apenas descrever o comportamento da série;
- (d) procurar periodicidades relevantes nos dados;

Em todos os casos podem-se utilizar modelos probabilísticos no domínio temporal. Como citado em Morettin e Toloí (2004), os modelos devem ser parcimoniosos, isto é, o número de parâmetros no modelo de ser o menor possível e sua utilização não deve apresentar dificuldades.

4.1.3 Modelos de Previsão

A maioria dos métodos de previsão de séries temporais baseia-se na idéia que o comportamento futuro depende de seu comportamento passado, ou seja, seu padrão de comportamento deve ser estudado antes de formular previsões.

Os métodos de previsão utilizados na prática podem variar muito, desde modelos simples e intuitivos até mais complexos. No primeiro a análise de dados pode ser quase nula, mas no segundo, que é o utilizado neste trabalho, esta análise pode ser considerável.

Em Economia existem dois procedimentos predominantes: O Econométrico e o de Séries Temporais. O primeiro é baseado na teoria econômica, enquanto o segundo se baseia “no que os dados dizem”, não se limitando a teorias econômicas, porém produzindo melhores previsões.

Como já comentado, este trabalho foi focado em Análise Estatística de Séries Temporais, sem recorrer a alguma teoria econômica adjacente, e será utilizado as seguintes famílias de modelos:

- ARIMA;
- STAR: TAR, SETAR e LSTAR;
- RNA: Redes Neurais Artificiais;

4.1.4 Metodologia Box e Jenkins

Para as famílias de modelos ARIMA e STAR, será utilizada a abordagem de Box e Jenkins (1970).

A idéia original está exposta abaixo na figura 1.

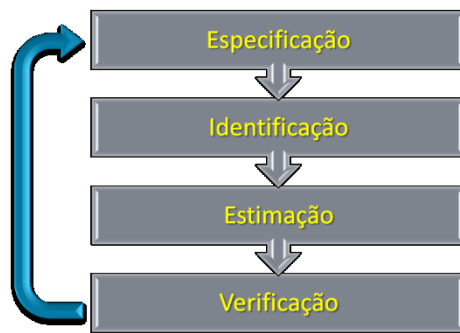


FIGURA 1 - CICLO BOX E JENKINS

Onde:

- (a) uma classe geral de modelos é especificada;
- (b) é realizada a identificação de um modelo com base na análise de gráficos de autocorrelação, gráficos de autocorrelação parcial e outros critérios;
- (c) na fase de estimação, os parâmetros do modelo identificado são estimados;
- (d) na última fase, a verificação do modelos é realizada através basicamente da análise de resíduos;

Com este procedimento, caso o modelo não seja adequado, repete-se o ciclo, até que o modelo ideal seja escolhido.

A fase crítica do procedimento acima é a identificação, que está sujeita a muita subjetividade. É bem possível então, que dois pesquisadores identifiquem modelos distintos para o mesmo conjunto de dados.

4.1.5 Modelos ARIMA

Se $W_t = \Delta^d Z_t$ for estacionária, podemos representar W_t por um modelo ARMA(p, q), ou seja, na equação 2 temos:

$$\phi(B)W_t = \theta(B)a_t \quad (2)$$

Se W_t é uma diferença de Z_t , então Z_t é um integral de W_t , onde Z_t segue um modelo auto-regressivo, *integrado*, de médias móveis, ou seja, ARIMA(p, d, q) pela equação 3:

$$\phi(B)\Delta Z_t = \theta(B)a_t$$

Pode-se representar vários modelos com modelos ARIMA, como mostra o quadro 1.

$ARIMA(p,0,0)$	=	$AR(p)$
$ARIMA(0,0,q)$	=	$MA(q)$
$ARIMA(p,0,q)$	=	$ARMA(p,q)$

QUADRO 1 - RESUMO DE MODELOS ARIMA

4.1.5 Modelos STAR

Nas últimas décadas, os modelos não-lineares começaram a ser mais desenvolvidos e surgiram os modelos de troca de regime tais como o modelo autorregressivo com limiar (TAR), modelos autorregressivo com transição suave (STAR).

Estes modelos encontraram um grande número de aplicações de sucesso na literatura, a idéia básica dos mesmos é atribuir modelos lineares distintos em regiões distintas onde se encontram os valores das séries em estudos. Nos casos do STAR as transições entre cada modelo são suaves.

O modelo TAR, como uma idéia de aproximar uma função não-linear geral por um modelo linear por partes. Em Chan & Tong (1986) é descrito uma generalização dos modelos TAR com uma transição mais suave, tal generalização ficou conhecida como modelos STAR (Smooth Transition Autoregressive). Posteriormente, Teräsvirta (1994) desenvolveu ainda mais este tipo de modelo. Em 1997, Liu & Li propõem um modelo auto-regressivo com um limiar que considere a heterocedasticidade da variância condicional da série temporal.

A sua popularidade se deve ao fato de sua facilidade de especificação, estimação e interpretação quando comparado com outros modelos não-lineares. Todavia apesar do desenvolvimento de inferência por Hansen (1997) pouco se sabe a respeito das propriedades estatísticas de seus estimadores. A idéia principal dos modelos TAR é descrever o seu processo por partes, ou seja, alterar os parâmetros de um modelo auto-regressivo linear dependendo da região em que se

encontrem valores de uma determinada variável, isto é, onde a quebra de cada modelo depende da região a que pertence o valor da variável.

Os modelos LSTAR podem seguir o mesmo ciclo de modelagem da metodologia Box & Jenkins. Ou seja, especificação, estimação dos parâmetros e validação do modelo. É importante dizer que para muitos outros modelos não lineares isso não seria possível.

4.1.5 Modelos RNA

De modo geral, uma rede neural é uma máquina construída para modelar a forma pela qual o cérebro realiza determinada tarefa ou função (Haykin, 1999). Pode ser implementada por meio de componentes eletrônicos ou de software.

Como o cérebro humano, as redes neurais são constituídas de um conjunto de unidades de processamento conectadas entre si, chamadas de neurônios artificiais ou, simplesmente, de neurônios.

A Figura 2 representa um neurônio artificial à direita, em que X1, X2 e X3 correspondem aos valores de entrada e Y equivale ao valor de saída do neurônio. Observa-se que, após a multiplicação dos valores de entrada pelos respectivos pesos W1, W2 e W3, os produtos obtidos são somados, resultando no potencial de ativação n. Posteriormente, o valor de n é submetido a uma função de ativação, cujo resultado Y é o valor de saída do neurônio. A função de ativação tem como objetivo limitar, dentro de uma escala de valores, a amplitude do sinal de saída. Portanto, matematicamente, tem-se:

$$v = \sum_{i=1}^n W_i \times X_i \text{ e } Y = \varphi(v) \Rightarrow Y = \varphi\left(\sum_{i=1}^n W_i \times X_i\right)$$

A função de ativação, representada por $j(\cdot)$, é aplicada ao potencial de ativação n para a produção do valor de saída Y do neurônio artificial. Segundo Haykin (1999), a função mais utilizada na construção de redes neurais artificiais é a sigmoideal.

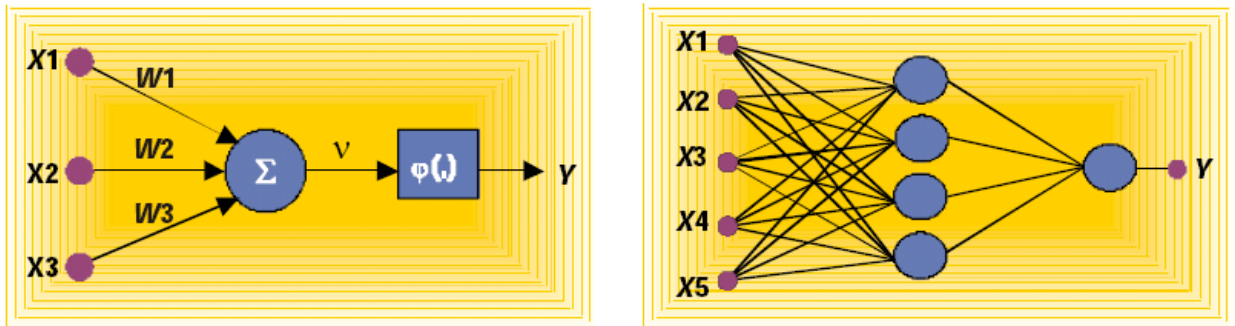


Figura 02

Definir a arquitetura de uma rede neural consiste em determinar sua organização estrutural: número de camadas da rede e de neurônios em cada camada, e tipo de conexão entre os neurônios e a topologia da rede (Braga, Ludermir e Carvalho, 2000). Normalmente, as redes neurais são estruturadas em camadas, contendo um ou mais neurônios. Todas elas possuem camada de entrada, que recebe os sinais de entrada, e camada de saída, que armazena o resultado final da rede. As camadas intermediárias localizam-se entre estas duas.

Quanto ao tipo de conexão, as redes neurais podem ser feedforward (acíclica) ou recorrente (cíclica). A recorrente diferencia-se da acíclica por possuir pelo menos um link de realimentação (feedback), pelo qual a saída de algum neurônio de uma camada i é utilizada como entrada para um neurônio de camada de ordem menor ou igual a i .

As redes neurais podem ser completamente conectadas, quando todos os neurônios estão conectados entre si, ou fracamente conectadas, quando pelo menos dois neurônios não apresentam ligações entre si. A Figura 2 à esquerda, ilustra uma rede neural acíclica completamente conectada.

4.2 Recursos Computacionais

4.2.1 Obtenção dos Dados

Na empresa ABC, o banco de dados do qual serão retiradas as informações, usa tecnologia *Microsoft SQL Server 2005*¹.

Este banco de dados tem informações desde o volume de vendas, passando pelo preço dos produtos, até as despesas operacionais da empresa.

Com o apoio da equipe de informática da empresa, será realizada uma consulta ao banco de dados para retornar os dados necessários.

4.2.2 Tratamentos dos Dados

Serão trabalhados no *Microsoft Excel 2007*¹, onde serão geradas partes das tabelas e gráficos apresentados no trabalho.

Para todo tratamento estatístico será utilizado o software R (versão 2.6) com o os pacotes: Tseries, TsDyn, específicos para séries temporais, tanto na parte descritiva da série quanto para realizar as previsões e comparativos entre modelos.

6 Referências Bibliográficas

MORETIN, P. A.; TOLOI M. C. **Análise de Séries Temporais** ABE – Projeto Fischer. 1. ed. São Paulo: Edgard Blücher, 2004.

MATOS, O. C. de; **Econometria Básica:** Teoria e Aplicações. 3.ed. rev. e ampl.. São Paulo: Atlas, 2000.

BRESSAN, A.A.; Tomada de Decisão em Futuros Agropecuários com Modelos de Previsão de Séries Temporais. **RAE-eletrônica**, v. 3, n. 1, Art. 9, jan./jun. 2004 . Disponível em <<http://www.rae.com.br/electronica/index.cfm?FuseAction=Artigo&ID=1914&Secao=FINANÇAS&Volume=3&Numero=1&Ano=2004>>. Acesso em: 13/08/2007.

EHLERS, R. S.; **Análise de Séries Temporais:** Apostila. 3.ed. Curitiba: 2005. Disponível em <<http://leg.ufpr.br/~ehlers/notas/stemp.pdf>>. Acesso em: 12/08/2007.

CHAVES NETO, A.; **Análise de Séries Temporais:** Notas de Aula. Curitiba 2006. Recebido via email em julho de 2006