

# PREVISÃO DE VENDAS DE CERVEJA PARA UMA INDÚSTRIA DE RIBEIRÃO PRETO

José Gilberto S. Rinaldi (UNESP/Presidente Prudente)  
Randal Farago (Faculdades Integradas FAFIBE)

**Resumo:** Este trabalho aborda técnicas de previsão de vendas de cerveja para uma indústria do setor de bebidas de Ribeirão Preto, que possam auxiliar na tomada de decisão dos executivos da referida empresa. Em particular, as análises são feitas com a utilização de modelos estatísticos de previsão auto-regressivos e de regressão causal, os quais projetam para o futuro, observações feitas no passado. Os testes computacionais são apresentados e analisados com o objetivo de mostrar a importância de uma boa previsão.

**Palavras-chave:** Previsão de vendas; séries temporais; métodos causais; sazonalidade.

## 1. Introdução

A previsão de vendas desempenha um papel fundamental nas empresas que trabalham com grandes flutuações na demanda para vendas diretas. É, em geral, de responsabilidade dos departamentos de vendas e/ou marketing, e é um *input* para a decisão no planejamento e controle de capacidade, que é normalmente uma responsabilidade da gerência de produção. Uma previsão de demanda deve ser expressa em termos úteis para o planejamento e controle da capacidade, podendo haver a necessidade de alguma tradução em termos de hora-máquina, homem-hora, espaço etc. Além disso, deve ser tão exata quanto possível, pois enquanto a demanda muda instantaneamente, existe a defasagem entre decidir alterar e essa alteração surtir efeito. Por fim, deve dar uma indicação da incerteza relativa, ou seja, de igual importância é uma estimativa de quanto à demanda real pode diferir da média (Slack *et al*, 2002).

A variação da demanda com o tempo é um resultado de crescimento ou declínio em taxas de vendas, sazonalidade na demanda-padrão e flutuações gerais causadas por uma infinidade de variáveis controladas ou não. A maioria dos métodos de previsão de curto prazo lida com este tipo de variação temporal, freqüentemente chamado de série de tempo ou série temporal (Ballou, 2001).

As previsões de demanda são elaboradas através de métodos quantitativos, qualitativos ou da mistura dos dois tipos. Os métodos quantitativos utilizam dados históricos para prever a demanda em períodos futuros. A previsão futura da demanda requer a construção de modelos matemáticos e estatísticos que utilizam os dados disponíveis ao longo do tempo (série temporal). Na prática são menos usados, pois necessitam de um dispêndio de tempo maior para estudo do melhor modelo para o problema, assim como a necessidade de implementações computacionais ou uso de pacotes estatísticos.

Já os métodos qualitativos baseiam-se nas opiniões, julgamentos, intuição, pesquisas, métodos comparativos e depoimentos de especialistas na área, tais como gerentes de área, vendedores, fornecedores e clientes. Como as pessoas, em geral, têm diferentes preferências, o modelo pode ser tornar tendencioso, não retratando a realidade e, portanto, podendo não ser confiável. O terceiro modelo seria uma mistura que considerasse a opinião de especialistas e utilizasse alguma modelagem matemática para a resolução de um problema (Larson, 2004)

Os métodos quantitativos dividem-se ainda em: Séries Temporais (ou projeções de tendências) que podemos destacar os modelos de Suavização Exponencial, Método de Winters, Holt, Médias Móveis, Box-Jenkins (ARMA, ARIMA, SARIMA), Redes Neurais; e Métodos causais (ou explicativos) que compreendem as Correlações, Regressões e Modelos Econométricos. Já os métodos qualitativos dividem-se em: Avaliação subjetiva, que pode ser pesquisa de opiniões de vendedores, júri de executivos, pesquisa de mercado com consumidores finais; Exploratórios, que englobam métodos de construção de cenários, Método Delphi e métodos de analogia.

Uma classe de modelos de previsão que tem sido muito útil ao longo dos anos é a da decomposição de série de tempo. Esses métodos incluem a análise espectral, a análise clássica das séries de tempo e as análises de séries de Fourier. A previsão da decomposição clássica da série de tempo é construída sobre a filosofia de que o padrão de vendas históricas pode ser decomposto em quatro categorias: tendência, variação sazonal, variação cíclica e variação residual ou aleatória.

A tendência representa o movimento em longo prazo nas vendas, no desempenho do marketing na empresa, e mudanças fundamentais na aceitação do mercado dos produtos e serviços da empresa. A variação sazonal refere-se aos picos e vales regulares nas séries temporais que repetem, geralmente, a cada 12 meses. As forças que causam esta variação regular incluem mudanças climáticas, padrões de compras vinculados às datas de calendário e disponibilidade dos bens. A variação cíclica é a ondulação de longo prazo (mais de um ano) no padrão de demanda. Variações residuais, ou aleatórias, são aquelas parcelas do total das vendas que não são responsáveis pelos componentes de tendência, sazonal ou cíclica. Se a série temporal for bem descrita pelos outros três componentes, a variação residual deverá ser aleatória (Ballou, 2001).

## 2. Modelagem e Métodos de Solução

A modelagem foi obtida, no contexto deste trabalho, quantitativamente, por meio de séries temporais (modelos de Box-Jenkins) e de métodos causais (regressão simples).

### 2.1 Modelos de Box-Jenkins

Uma série temporal, denotada por  $z_t$ , é um conjunto de valores sucessivos e dependentes de alguma variável (volume de vendas, temperatura, etc) registrada no decorrer do tempo.

Uma estatística importante na análise de séries temporais é o coeficiente de auto-correlação  $\rho$ , sendo este utilizado para descrever a correlação entre dois valores da mesma série temporal, em diferentes períodos de tempo (dias, semanas ou meses, por exemplo). De modo geral, o coeficiente de auto-correlação  $\rho_k$  mede a correlação entre observações distantes  $k$  períodos de tempo (ou seja, uma auto-correlação de *lag*  $k$ ).

A auto-correlação de *lag*  $k$  é medida pelo coeficiente  $\rho_k$ , definido por:

$$\rho_k = \frac{E[(z_t - \mu).(z_{t-k} - \mu)]}{\sqrt{E[(z_t - \mu)^2].E[(z_{t-k} - \mu)^2]}} \quad (1)$$

onde  $\mu$  é a média da série temporal (em equilíbrio).

Os modelos de Box-Jenkins (Box-Jenkins, 1976) assumem que os valores de uma série temporal são altamente dependentes, isto é, cada valor atual pode ser

explicado por valores anteriores da série. Os modelos ARIMA e SARIMA representam uma das classes mais gerais de modelos para a análise de séries temporais.

Os modelos, mais especificamente, são o AR( $p$ ) (auto-regressivo de ordem  $p$ ) e MA( $q$ ) (média-móvel de ordem  $q$ ), que podem ser integrados, gerando os modelos ARIMA( $p,d,q$ ), ou seja, composto de AR( $p$ ) e MA( $q$ ) com  $d$  diferenciações. Essas diferenciações são referentes a estacionariedade de  $z_t$ , ou seja, podem ser aplicadas diferenciações na série até que esta se estabilize no tempo. Em geral  $d$  assume os valores 0, 1 ou 2 no máximo, sendo que para  $d = 0$  nenhuma diferenciação foi aplicada. Existem ainda os modelos com sazonalidade. Esses são utilizados para analisar séries temporais que apresentam comportamento recursivo no tempo, por exemplo, a cada  $s$  períodos de tempo. Assim, se um comportamento se repete a cada ano, ou seja, com sazonalidade anual, então  $s = 12$  meses. São denotados por SARIMA ( $p,d,q$ )(P,D,Q)  $s$ , onde as letras maiúsculas denotam a parte sazonal do modelo e são análogas a  $p,d,q$  sendo que  $s$  funciona como descrito acima. Trata-se, portanto, de um modelo multiplicativo geral considerando conjuntamente as partes não sazonal e sazonal.

## 2.2 Análise de Regressão e Correlação

A análise de regressão causal é uma técnica estatística para investigar e modelar relações existentes entre variáveis. Trata-se de uma técnica bastante utilizada em várias áreas de conhecimento tais como engenharia, economia, ciências físicas, sociais e biológicas, entre outras.

É necessário tomar bastante cuidado para não extrapolar as limitações que se podem obter destas relações. Um comportamento revelado para os dados observados em uma determinada circunstância pode sofrer mudanças quando esta é alterada.

Como exemplo, seja  $y$  representando a pressão atmosférica e  $x$  representando o ponto de ebulição da água, então se pode aproximar a relação dessas duas variáveis por uma equação de reta, dada por

$$y = \beta_0 + \beta_1 x \quad (2)$$

onde,  $\beta_0$  é o intercepto e  $\beta_1$  é a inclinação. Contudo, evidentemente, os pontos não passam exatamente sobre a reta existindo então uma diferença, denotada por  $\varepsilon$  e denominada erro aleatório (ou resíduos). O modelo mais adequado para os dados é então

$$y = \beta_0 + \beta_1 x + \varepsilon. \quad (3)$$

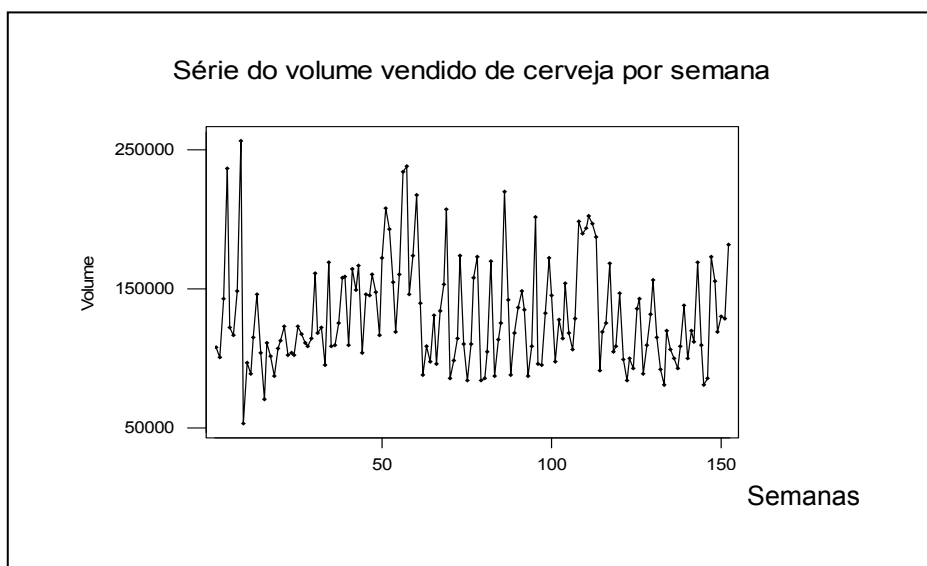
A equação acima é chamada de modelo de regressão linear, sendo  $x$  a variável independente e  $y$  a variável dependente, ou ainda,  $x$  como variável preditora e  $y$  como variável resposta.

Uma estatística, denotada por  $R^2$ , e freqüentemente denominada de coeficiente de determinação, pode ser entendida como uma “proporção de variação explicada pelo preditor  $x$ ”. Sabe-se que  $0 \leq R^2 \leq 1$  e ainda que valores perto de 1 implicam que boa parte de variabilidade é explicada pelo modelo de regressão e, ao contrário, quando tende a zero, que grande parte da variabilidade estará nos resíduos.

## 3. Ferramentas e Resultados Computacionais

A série apresentada para análise representa o volume de vendas por semana de cerveja em unidades específicas da empresa. Esta série compreende 156 semanas, sendo que, as últimas quatro foram retiradas e reservadas, para avaliar a qualidade das previsões fornecidas pelos modelos encontrados. A série pode ser contemplada

observando-se o Gráfico 1 abaixo. Por este gráfico pode-se observar que não parece haver tendenciosidade quanto à estacionariedade da série.



**Gráfico 1:** Gráfico da série semanal para volume vendido de cerveja, em unidades específicas da empresa.

### 3.1 Modelo obtido com Regressão Causal

Juntamente com a série de volume de vendas semanais de cerveja foi obtida a série do preço médio praticado para a respectiva semana. Com esta informação foi possível a modelagem do volume tendo como causa (explicação) seu preço médio semanal utilizando-se o *software* Minitab, versão 12. A equação de regressão obtida é

$$\text{Volume} = 261129 - 13181 \cdot \text{Preço Médio}.$$

A equação acima faz sentido, já que um aumento do preço médio praticado deve provocar uma diminuição do volume vendido. Os testes T-Student para constante e preço médio (preditores) do modelo estão na Tabela 1, com os respectivos P-valores (probabilidade do preditor não estar na equação de regressão ajustada).

Tabela 1: Teste para os preditores da regressão.

Preditor	Coefficiente	Teste t-student	P-valor
Constante	261129	6,40	0,000
Preço Médio	-13181	-3,20	0,002

Apesar dos testes apontarem para que os preditores sejam altamente significativos (Tabela 2) e que isto também ocorre com a análise de variância da equação de regressão (com P-valor 0,002), o coeficiente de determinação ( $R^2$ ) é bastante baixo, ficando apenas em 0,058 (ou 5,8%). Assim, os resultados não podem negar que haja uma relação entre volume e preço médio, contudo este modelo não é apto a prever esta relação.

Para uma melhor avaliação faremos a previsão para os quatro pontos retirados da série. Os resultados podem ser observados na Tabela 2.

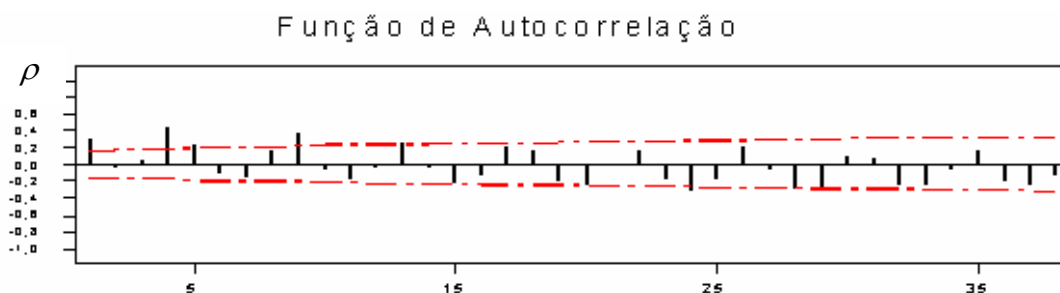
Tabela 2: Predição para os quatro pontos futuros da série.

Observação	Real	Predito	Erro
153	120603	120112	-0,4%
154	120938	121247	0,3%
155	125143	122569	-2,1%
156	158327	123439	22%

A predição para a semana 156 revela a fragilidade do modelo após erros irrelevantes nas anteriores. Assim pode-se esperar que, para algumas semanas, haja grandes distorções para as previsões, se o modelo for utilizado.

### 3.2 Modelo obtido com Box-Jenkins

Como observado anteriormente, pelo Gráfico 1, na série de volume não parece haver tendenciosidade quanto à estacionariedade da série. Desta forma, não se devem fazer necessárias diferenciações para um bom ajuste do modelo a ser considerado. Para começo de análise, em qual modelo deve-se considerar, são calculadas as autocorrelações, presentes no Gráfico 2 abaixo.



**Gráfico 2:** Gráfico da função de autocorrelação para a série de volume. *lag*

Para a função de autocorrelação observa-se que, para os lags 4 e 9, se apresentam respectivamente o maior e o segundo maior valor calculado até o lag 35. Isto ocorre provavelmente devido ao mês ter aproximadamente 4,3 semanas, extensão de tempo entre os prováveis planejamentos de compras. Se considerados dois meses, ou seja, aproximadamente nove semanas, explica-se a alta autocorrelação para o lag 9. Esta análise leva a avaliar uma sazonalidade de período 4. Após algumas tentativas considerou-se o modelo  $(p, d, q).(P, D, Q)_s \equiv (1, 0, 1).(1, 0, 0)_4$ , com resultados na Tabela 3 (onde S, em SAR(1), denota sazonalidade).

Tabela 3: Teste para estimativas de parâmetros:  $(p, d, q).(P, D, Q)_s \equiv (1,0,1).(1,0,0)_4$

Tipo	Coefficiente	Teste T-Student	P-valor
AR(1)	1,0000	4307,69	0,000
SAR(1)	0,4629	6,30	0,000
MA(1)	1,0026	15259,74	0,000

Pela Tabela 3 pode-se observar que os tipos de componentes do modelo de série temporal não foram rejeitados pelo teste T-Student, de forma muito clara, se forem observados também os P-valores. É importante ressaltar que todas as suposições para o modelo proposto, tais como, o teste F para validá-lo, normalidade, não tendenciosidade e autocorrelação dos resíduos e sua função de autocorrelação foram checados, não se

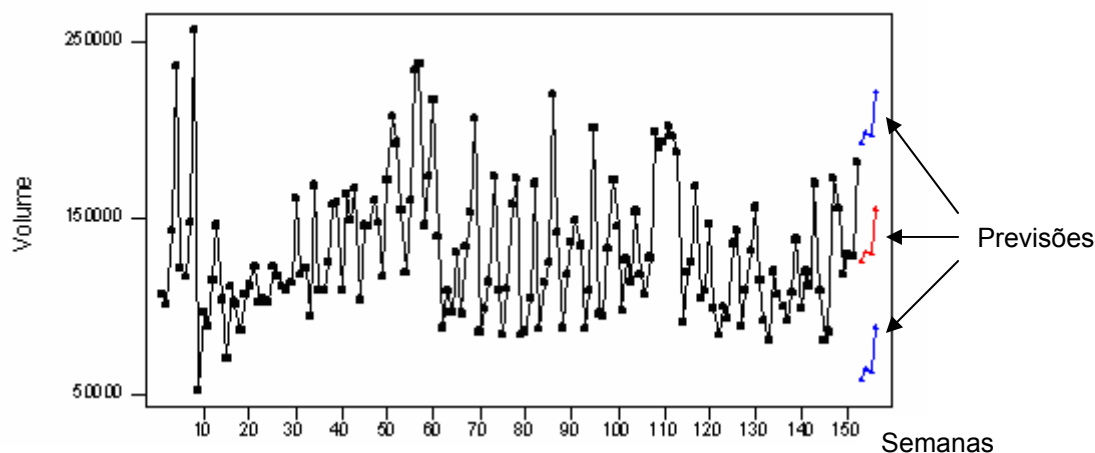
verificando maiores problemas. Somente a função de autocorrelação não teve a estabilização desejada.

Isto provavelmente ocorre devido ao descompasso entre a periodicidade do modelo e o suposto comportamento mensal das vendas. Quanto a normalidade dos resíduos a situação parece bastante favorável ao modelo. Os resíduos têm bom comportamento, com média em torno de zero e dispersão aparentemente aleatória, não revelando tendenciosidade ou autocorrelacionamento.

Para avaliar melhor o modelo deve-se realizar as previsões das quatro semanas futuras e contrastar estas com os valores reais. As previsões comparativas aos valores reais podem ser vistas na Tabela 4 (onde L. denota limite). No Gráfico 3, pode-se ver as previsões das últimas quatro semanas com seus respectivos intervalos de confiança

Tabela 4: Comparação dos valores previstos (95% de confiança) com valores reais.

Período	Previsão	L. Inferior	L. Superior	Real	Erro
153	125881	58841	192920	120603	4,4%
154	131267	64227	198307	120938	8,5%
155	130544	63505	197584	125143	4,3%
156	155167	88127	222208	158327	-2%



**Gráfico 3:** Gráfico com as previsões para as quatro semanas futuras com seus intervalos de confiança.

Apesar dos intervalos de confiança possuírem grande amplitude, as previsões pontuais apresentam, em geral, erros suportáveis se considerada a natureza da série estudada.

#### 4. Considerações Finais

Os dois modelos considerados têm características bastante distintas em seus ajustes. O modelo de regressão causal não é capaz de explicar os dados da série de volume de vendas de cerveja, contudo não nega que exista uma relação entre este e o preço médio praticado. O modelo de Box-Jenkins proposto pode ser utilizado para as previsões, pois os erros parecem ser suportáveis pela complexidade natural do problema. Este modelo pode ser ainda melhor avaliado e caracterizado para sua utilização.

Pode-se ainda pensar em compor algum modelo de Box-Jenkins conjuntamente com variáveis explicativas, incorporando mais informações ao modelo e supostamente melhorando suas previsões.

## 5. Referências Bibliográficas

BALLOU, R. H. “**Gerenciamento da Cadeia de Suprimentos: Planejamento, Organização e Logística Empresarial**”. 4 ed. Porto Alegre: Bookman, 2001.

BOX, G. E. P.; JENKINS, G. M., “**Time series analysis – Forecasting and Control**”. Holden-Day: San Francisco, 1976.

LARSON, R.; FARBER, B., “**Estatística Aplicada**”. 2 ed. São Paulo: Prentice Hall, 2004.

SLACK, N.; CHAMBERS, S.; JOHNSTON, R., “**Administração da Produção**”. 2 ed. São Paulo: Atlas, 2002.