

## METODOLOGIA BOX & JENKINS: UMA APLICAÇÃO EM DADOS DE PRODUÇÃO DE LEITE CRU DO ESTADO DE MINAS GERAIS

### Box & Jenkins methodology: an application in raw milk data from the state of Minas Gerais

*Eduardo Campana Barbosa<sup>1\*</sup>; Thelma Sáfadi<sup>2</sup>; Carlos Henrique Osório Silva<sup>3</sup>; Rômulo César Manuli<sup>1</sup>*

#### RESUMO

Utilizou-se a metodologia Box & Jenkins para obter um modelo estatístico que estimasse a produção de litros de leite dos seis primeiros meses de 2013 no estado de Minas Gerais, ajustando modelos SARIMA  $(p, d, q) \times (P, D, Q)_s$ , no qual  $d$  e  $D$  são o número de diferenças necessárias para remover a tendência e sazonalidade da série,  $p$  e  $q$  a ordem dos operadores autoregressivos e de médias móveis,  $P$  e  $Q$  a ordem dos operadores autoregressivos e de médias móveis sazonais e  $s$  a periodicidade sazonal. Por meio do Critério de Informação de Akaike (AIC) selecionou-se os seis modelos mais parcimoniosos e para encontrar o melhor foram analisados os indicadores Erro Quadrático Médio (EQM) e Erro Percentual Médio Absoluto (MAPE), além das pressuposições de resíduos ruído branco. O modelo Autoregressivo Integrado e de Médias Móveis Sazonal SARIMA  $(3,1,2) \times (0,1,2)_{12}$  foi superior, pois atendeu ao princípio da parcimônia, obteve estimativas de produção de leite mais ajustadas e consequentemente menores valores para os indicadores de erro EQM e MAPE. Os resíduos estimados por este modelo foram ruído branco.

**Palavras-chave:** previsão; modelagem; tendência; sazonalidade.

#### ABSTRACT

The Box & Jenkins methodology was used to obtain a statistical model for estimate the production in liters of milk of the 6 first months of 2013 in Minas Gerais

---

1 Universidade Federal de Viçosa, Departamento de Estatística (UFV-DET), Av. P.H. Rolfs, Campus Universitário, 36570-000, Viçosa, MG, Brasil. E-mail: duducampana@hotmail.com

2 Universidade Federal de Lavras, Departamento de Ciências Exatas (UFLA-DEX).

\* Autor para correspondência.

**Recebido / Received: 27/01/2014**

**Aprovado / Approved: 13/03/2014**

state, adjusting SARIMA  $(p, d, q) \times (P, D, Q)_s$  models, where  $d$  and  $D$  are the number of differences to remove the trend and seasonality of time series,  $p$  and  $q$  are the order of the autoregressive and moving average operators,  $P$  and  $Q$  are the order of the autoregressive and moving average seasonal operators and  $s$  is the seasonal periodicity. The Akaike Criterion Information (AIC) procedure was used to select the 6 most parsimonious models and to find the best one the error indicators Mean Squared Error (EQM) and Mean Absolute Percent Error (MAPE) were analyzed, in addition to the assumptions of residues white noise. The Seasonal Autoregressive Integrated Moving Average SARIMA  $(3,1,2) \times (0,1,2)_{12}$  was upper, view of the principle of parsimony and with more precise estimates. The forecast was more adjusted to the real values of milk production in Minas Gerais state and the model had smaller error indicators. The residues estimated were by this model white noise.

**Keywords:** forecasting; modeling; trend; seasonality.

## INTRODUÇÃO

Segundo Silva et al. (2007), quando o processo de ordenha é completo e ininterrupto, sob condições adequadas de higiene e com vacas saudáveis, isto é, bem alimentadas e descansadas, obtêm-se como retorno o leite cru, produto característico e uma das principais matérias-primas brasileiras. Hemme; Otte (2010) complementam afirmando que o leite de vaca cru, além de possuir grande importância no cenário agropecuário brasileiro, tem contribuído para a evolução do setor de laticínios. Conseqüentemente, com a melhoria do aspecto econômico e social do país, devido à geração de renda, empregos e a composição de suprimentos alimentícios para a população, providos de derivados e subprodutos do leite.

Conforme Fao (2009), o Brasil é responsável por 5,9% do leite produzido mundialmente, ocupa a quinta colocação no ranking, com produtividade inferior aos Estados Unidos, Índia, China e Rússia, respectivamente. Uma característica marcante da produção de leite brasileira é a heterogeneidade entre os estados do país, fato explicado por influências econômicas, ambientais e, principalmente, investimentos em tecnologia (TEIXEIRA et al., 2003). De acordo com Carneiro et al. (2008), a organização dos processos visando o aumento

de serviços e a redução de custos operacionais, além dos investimentos em tecnologias, são fatores fundamentais para explicar essa heterogeneidade. Quanto mais avançado o processo produtivo melhor será o produto em termos de qualidade e confiabilidade, o que implicará no aumento de demanda.

Conforme Álvares (2005), o aumento da produção de leite no Brasil e, principalmente em Minas Gerais, pode ser explicado pela reestruturação do mercado lácteo a partir da década de 1990, além da recente busca por mercados externos. Esses eventos priorizaram um aspecto fundamental: o planejamento do processo produtivo do leite. Logo, sob esse enfoque, percebeu-se o quanto é importante o estudo das tendências desse mercado para conseguir atender a alta demanda.

Neste sentido, a análise de séries temporais é uma metodologia estatística frequentemente utilizada para estabelecer projeções futuras, embasadas em teorias de que o comportamento do mercado, ao longo do tempo, consiste na repetição de tendências, sazonalidades e variações aleatórias, de forma semelhante como ocorrido no passado (MORETTIN; TOLOI, 2004). No presente artigo ajustaram-se modelos Box & Jenkins do tipo Seasonal Autoregressive Integrated Moving Average (SARIMA), que em português significa Autoregressivo Integrado e de Médias Móveis Sazonal, para previsão

da demanda de leite cru (resfriado ou não) no estado de Minas Gerais. O principal objetivo foi estimar os níveis de produção de leite para os seis primeiros meses de 2013 e, através de comparações com os respectivos valores reais, buscou-se verificar a aplicabilidade e qualidade do ajuste dos modelos SARIMA a séries temporais de produção de leite. Uma justificativa para o desenvolvimento deste artigo é a escassez de trabalhos sobre previsão de demanda para a produção de leite cru no Brasil, pois o que muito se encontra na literatura, sobre previsão, refere-se ao preço do leite.

## MATERIAL E MÉTODOS

Os dados utilizados foram obtidos no Sistema IBGE de Recuperação Automática (banco de dados SIDRA do IBGE) e referem-se à série histórica de produção em litros de leite cru (resfriado ou não) do estado de Minas Gerais. Foram obtidas 126 observações (de janeiro de 2003 a junho de 2013), em que cada observação representa o valor produzido em litros de leite no dado mês de um determinado ano. Porém, para a modelagem estatística utilizou-se apenas os dados compreendidos no período entre janeiro de 2003 a dezembro de 2012. Desta forma considerou-se que o número de observações da série é  $N = 120$ . Isso ocorreu pois o horizonte de previsão contemplou o primeiro semestre de 2013 e, portanto, os valores reais desses meses foram reservados para comparar com as estimativas obtidas pelos modelos de previsão, bem como para o cálculo dos indicadores de erro.

O procedimento metodológico inicial consistiu em decompor a série em suas componentes de tendência, sazonalidade e parte aleatória (não modelável).

Para comprovar a presença de tendência nos dados procedeu-se o teste não paramétrico de Cox-Stuart (teste do Sinal) ao nível de 5% de significância. Como o número de observações da série ( $N$ ) é par, assumiu-se que  $c = N/2$ . Portanto, as observações foram agrupadas

duas a duas em  $c$  pares, da seguinte maneira:  $(Y_1, Y_{1+c}), (Y_2, Y_{2+c}), \dots, (Y_{N-c}, Y_N)$ . Realizado este procedimento, um sinal positivo foi atribuído sempre que em um par  $Y_i < Y_{i+c}$ , o que resultou em  $T$  sinais positivos. O termo  $n$  recebeu o valor em (1) todas as vezes que não houve empate entre as observações de cada par, ou seja,  $Y_i \neq Y_{i+c}$ . Se  $n > 20$  pode-se utilizar a aproximação da distribuição binomial para a normal ( $z_{\alpha/2} = 1,96$ ) e calcular o valor da estatística  $t = \frac{1}{2}(n + z_{\alpha/2} \cdot \sqrt{n})$ . se  $T \geq n-t$  conclui-se sobre a existência de tendência na série. Maiores detalhes sobre este teste podem ser encontrados em Morretin; Toloí (2004).

A presença de sazonalidade nos dados foi verificada através da análise espectral e pelo teste  $G$  de Fisher. A construção do periodograma da análise espectral ocorreu decompondo a série temporal em uma série de Fourier (domínio de frequências). No eixo das ordenadas ficaram localizadas as  $n_1 = N/2$  frequências, calculadas como  $f_i = \frac{i}{n_1}$  onde  $i$  é uma sequência de valores de 1 a  $n_1$ . No eixo das abscissas encontram-se as respectivas intensidades de cada frequência, sendo

$$I(f_i) = \frac{2}{n_1} \left\{ \left[ \sum_{t=1}^{n_1} \varepsilon_t \cos\left(\frac{2\pi i}{n_1} t\right) \right]^2 \left[ \sum_{t=1}^{n_1} \varepsilon_t \sin\left(\frac{2\pi i}{n_1} t\right) \right]^2 \right\}.$$

Destaca-se  $\gamma_i$  é o componente aleatório da série temporal associado ao tempo  $t$ . Posteriormente, verificou-se qual frequência estava associada ao maior valor de intensidade  $I(f_i)$ . A periodicidade sazonal foi obtida dividindo 1 por este valor de frequência. Para testar sua significância estatística utilizou-se o teste  $G$  de Fisher. Se comprovado que o valor da estatística

$$G = \max I(f_i) / \sum_{i=1}^{n_1} I(f_i) \text{ é superior ao valor de}$$

$$Z = 1 - \left(\frac{\alpha}{n_1}\right)^{1/(n_1 - 1)}, \text{ pode-se concluir que a componente sazonal com periodicidade } s = 1/f_i \text{ é}$$

significante. Maiores detalhes sobre estes procedimentos podem ser encontrados em Morettin; Toloí (2006) e Jenkins; Watts (1968).

Identificadas as características presentes na série temporal, optou-se por utilizar o

modelo SARIMA multiplicativo, útil para modelagem de dados com autocorrelação periódica e inferior a um ano (sazonalidade), pois além dos operadores autoregressivos e de médias móveis de ordem  $p$  e  $q$  do modelo ARIMA, tal modelo conta com operadores autoregressivos e de médias móveis sazonais de ordem  $P$  e  $Q$ . O termo  $W_t$  representa os valores de  $Y_t$  após a aplicação de  $d$  e/ou  $D$  diferenças de ordem  $l$  e  $s$  (valor da periodicidade sazonal), ou seja,  $W_t = (1 - B)^d \cdot (1 - B^s)^D \cdot Y_t$ , necessárias para remover a tendência e/ou sazonalidade da série, tornando-a estacionária, o que segundo Makridakis et al. (1998), é uma pressuposição para aplicação da metodologia Box & Jenkins. O referido modelo é descrito na equação (1).

$$\Phi_p(B)\Phi_p(B^s)W_t = \Theta_q(B)\Theta_q(B^s)\varepsilon_t \quad (1)$$

$$\Phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (2)$$

$$\Theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (3)$$

$$\Phi_p(B^s) = 1 - \Phi_s B^s - \Phi_{2s} B^{2s} - \dots - \Phi_{Ps} B^{Ps} \quad (4)$$

$$\Theta_q(B^s) = 1 - \Theta_s B^s - \Theta_{2s} B^{2s} - \dots - \Theta_{Qs} B^{Qs} \quad (5)$$

$\varepsilon_t$  é o ruído aleatório no tempo  $t$ ,  $\Phi_p(B)$  são os operadores autoregressivos e de médias móveis definidos em (2) e (3) e  $\Phi_p(B^s)$  e  $\Theta_q(B^s)$  os operadores autoregressivos e de médias móveis sazonais, conforme em (4) e (5).

A aplicação da metodologia Box & Jenkins consistiu em três etapas usuais: identificação dos parâmetros do modelo, estimação e diagnóstico. Segundo Morettin; Toloí (2004), com a série estacionária, a identificação dos parâmetros ocorre como em um processo Autoregressivo de Médias Móveis (ARMA), via funções de Autocorrelação (ACF) e Autocorrelação Parcial (PACF). No entanto, Makridakis et al. (1998) afirmam que obter a ordem exata dos parâmetros através destes gráficos é algo complexo e subjetivo, devido ao grande número de modelos que podem ser ajustados. Neste trabalho esta etapa foi conduzida

simulando diversos modelos e selecionando os seis mais parcimoniosos por meio do Critério de Informação de Akaike (AIC) (Akaike, 1974), calculado como  $C = -2 \cdot (\text{LMV}) + 2 \cdot (N_1)$ , no qual LMV é o logaritmo da máxima verossimilhança e é número de parâmetros do modelo.

A estimação dos parâmetros ocorreu pela distribuição Gaussiana (normal) e método da máxima verossimilhança, satisfazendo as condições de invertibilidade e unicidade dos parâmetros (REZENDE et al., 2005).

O diagnóstico consistiu em verificar se os erros seriam um ruído branco, ou seja,  $\varepsilon \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ . Para avaliar a presença de autocorrelação serial utilizou-se o teste estatístico Ljung-Box (Q) (Ljung; Box, 1978) além da ACF para os resíduos estimados. A estatística Q possui distribuição Qui-Quadrado com  $k - N_1$  graus de liberdade e é calculada como  $Q = N \cdot (N+2) \cdot \sum_{i=1}^k (N-K)^{-1} r_k^2$ , onde  $k$  é o número de defasagens utilizadas (normalmente 15) e  $r_k$  é valor estimado do  $k$ -ésimo coeficiente de autocorrelação amostral. Logo, o valor-p é obtido como Probabilidade de  $Q \geq \chi^2_{G, I}$ . A normalidade foi também verificada por meio do valor-p para o teste Shapiro-Wilk, que segundo Torman et al. (2005) é um dos mais eficientes testes de normalidade.

Após esses procedimentos calculou-se o Erro Quadrático Médio (EQM) e o Erro Percentual Médio Absoluto (MAPE) baseado nos valores estimados e reais de produção de leite dos seis primeiros meses de 2013. O cálculo ocorreu da seguinte maneira:

$$\text{EQM} = \frac{1}{6} \sum_{t=1}^6 e_t^2 \quad \text{e} \quad \text{MAPE} = \frac{1}{6} \sum_{t=1}^6 \left| \frac{e_t}{Y_t} \right| \times 100. \quad \text{Onde}$$

é o resíduo (erro estimado) da previsão no tempo  $t$ .

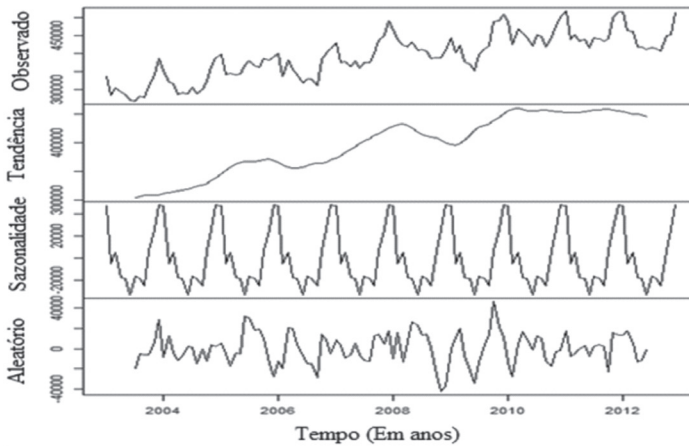
Utilizou-se o software livre R (R Development Core Team, 2012) para gerenciamento e análise estatística dos dados, sendo o stats, forecast e timeSeries os pacotes mais usados.

## RESULTADOS E DISCUSSÃO

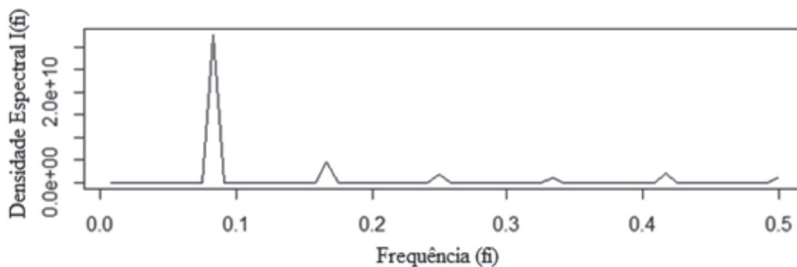
A Figura 1 apresenta o gráfico da série temporal de produção de leite cru (em litros) do estado de Minas Gerais (MG) e sua decomposição temporal. Para aplicação do teste de Cox-Stuart tem-se que o número de observações da série é  $N = 120$ , portanto, o número de pares é  $c = 60$ . O número de sinais positivos atribuídos a cada vez que em um par  $Y_i < Y_{i+c}$  é  $T = 60$ . Como em nenhum dos 60 pares  $Y_i = Y_{i+c}$ , tem-se que  $n = 60$ . Com estas informações a estatística do teste é  $t = 37,5$ . Logo, como  $T \geq n - t$ , conclui-se que existe tendência nos dados. A Figura 2 ilustra o Periodograma da série temporal. O número de frequências é  $n_1 = 60$  e a maior

intensidade ou densidade espectral esta associada à frequência de 0,08333, o que indica a presença de uma componente sazonal com periodicidade de 12 meses. Pelo teste G de Fisher foram obtidos os valores de  $G = 0,7577$  e  $Z = 0,1132$ , o que permite concluir sobre a presença de sazonalidade anual na série.

Alves et al. (2013) obtiveram resultados semelhantes ao analisar o preço do leite para os estados da Bahia e Minas Gerais. Os autores verificaram que a série apresentava tendência e sazonalidade bem definidas. Souza et al. (2010) explicam que no verão, devido à intensidade de chuvas existe, abundância da pastagem. Isso permite que o rebanho consuma e produza maiores quantidades de leite. O inverso ocorre no inverno, devido à escassez



**Figura 1** – Série temporal e sua decomposição em tendência, sazonalidade e parte aleatória



**Figura 2** – Análise espectral da série temporal de estudo

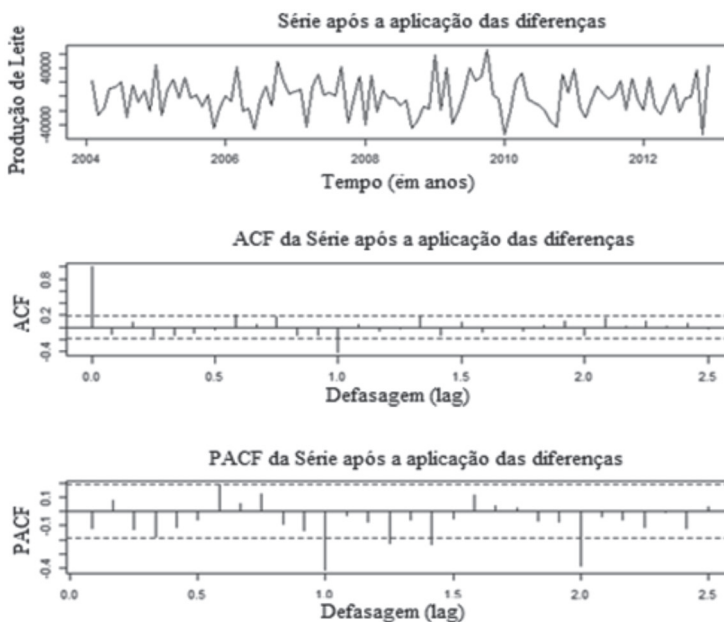
(seca), o que implica na queda do volume de leite produzido.

Comprovada a presença de tendência e sazonalidade nos dados, aplicou-se uma diferença de ordem 1 e uma diferença de ordem  $s = 12$  para remover ambos os componentes e estabilizar sua média e variância. Na Figura 3 verifica-se a série após a aplicação das diferenças e os respectivos correlogramas da ACF e PACF.

Nota-se que a série esteve estacionária após a aplicação das diferenças, visto que os dados oscilaram em torno de um valor médio com variância aproximadamente constante. Esta afirmativa é corroborada por meio da ACF, que se tornou estatisticamente igual à zero rapidamente, o que segundo Pindyck; Rubinfeld (2004) é um indicio de série estacionária.

Verifica-se que após eliminar a componente sazonal, em algumas defasagens (lags) múltiplas de 12 na ACF e PACF os coeficientes de autocorrelação ainda foram significativos, o que conforme Morettin; Toloi (2006) indica a presença de sazonalidade estocástica, que deve ser eliminada através do ajuste de um modelo que considere este tipo de sazonalidade, além da determinística, retirada por diferença.

Neste sentido, aplicou-se a metodologia proposta por Box & Jenkins (1976) ajustando modelos SARIMA  $(p,d,q) \times (P, D, Q)_s$ . Foram testados diversos modelos com os valores de  $p$  e  $q$  variando de 0 a 3, e  $P$  e  $Q$  de 0 a 2. Como foi aplicada uma diferença de ordem 1 e uma diferença de ordem 12,  $d$  e  $D$  assumem o valor de 1. Os resultados podem ser visualizados na Tabela 1.



**Figura 3** – Série, ACF<sup>1</sup> e PACF<sup>2</sup> após a aplicação das diferenças

- 1 Autocorrelation Function (ACF), que significa em português, Função de Autocorrelação.
- 2 Partial Autocorrelation Function (PACF), que significa em português, Função de Autocorrelação Parcial.

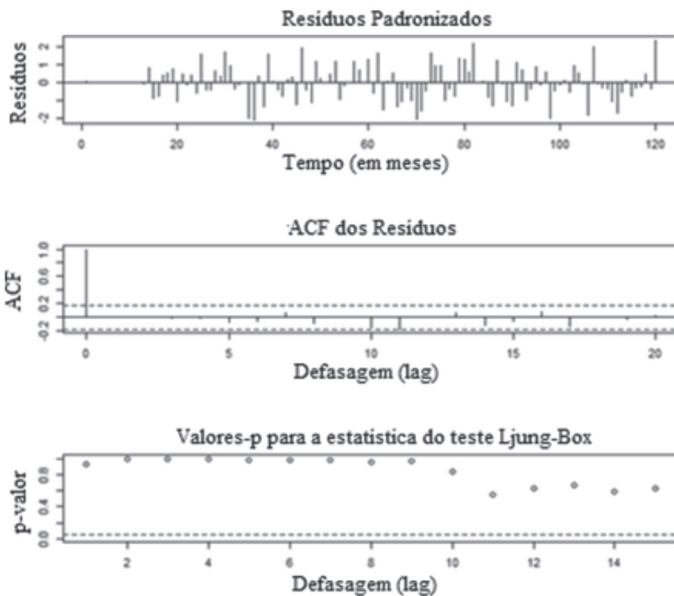
Estes foram os seis modelos mais parcimoniosos selecionados segundo o AIC. Nota-se que os seis modelos estimam resíduos que são ruído branco, visto que o valor-p para o teste Ljung-Box e Shapiro-Wilk foram superiores a 0,05. O modelo SARIMA  $(3,2,1) \times (0,1,2)_{12}$  obteve melhores resultados, apresentando menor EQM = 68.850.450 litros de leite e MAPE = 1,5929%. Utilizar um segundo indicador de erro é importante

quando a escala da variável em estudo é elevada. A Figura 4 corrobora características de ruído branco deste modelo, visto que os resíduos mostram-se não correlacionados e homogêneos.

Na metodologia Box & Jenkins (1976) é adequado que os coeficientes estimados sejam estatisticamente significativos segundo o teste t de Student. Porém, neste trabalho essa premissa não foi uma condição fundamental

**Tabela 1** – Comparação entre modelos SARIMA

Modelo	SARIMA (s=12)	AIC	Ljung- Box (valor-p)	Shapiro- Wilk (valor-p)	EQM (litros)	MAPE (%)
1	(0,1,1) x (2,1,1)	2.417,22	0,2376	0,0699	202.246.675	2,5883
2	(3,1,2) x (0,1,1)	2.416,90	0,4310	0,3530	111.147.778	1,6567
3	(3,1,2) x (0,1,2)	2.417,24	0,6232	0,2519	68.850.450	1,5929
4	(1,1,0) x (2,1,1)	2.417,63	0,2580	0,0655	233.057.649	2,8062
5	(1,1,1) x (0,1,2)	2.416,68	0,4362	0,2184	256.949.811	3,1469
6	(3,1,3) x (0,1,1)	2.420,56	0,5746	0,2633	265.209.711	3,5228



**Figura 4** – Pressuposições para o Modelo SARIMA  $(3,1,2) \times (0,1,2)_{12}$

para aprovação e seleção de modelos, e sim o princípio de parcimônia e resíduos com características de ruído branco. A Tabela 2 apresenta os coeficientes do modelo e suas

respectivas estatísticas. Verifica-se que os parâmetros  $\phi_3$  e  $\Theta_2$  não foram significativos (valor-p > 0,05), porém, destaca-se que o modelo foi testado sem o parâmetro  $\phi_3$ , sem

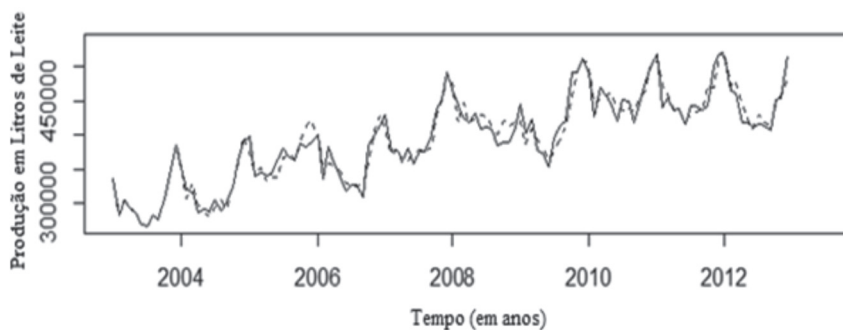
**Tabela 2** – Estatísticas para o Modelo SARIMA (3,1,2) x (0,1,2)<sub>12</sub>

Parâmetros	Coefficientes	Erro Padrão	t de Student	valor-p
$\phi_1$	1,2979	0,1005	12,9148	0,0000*
$\phi_2$	-0,6323	0,1476	-4,2830	0,0000*
$\phi_3$	-0,1707	0,1012	-1,6878	0,0914
$\theta_1$	-1,6375	0,0663	-24,7125	0,0000*
$\theta_2$	0,9996	0,0777	12,8572	0,0000*
$\Theta_1$	-1,1770	0,2411	-4,8822	0,0000*
$\Theta_2$	0,1776	0,1417	1,2535	0,2100

\*significativo a 0,05

**Tabela 3** – Estimativa de Produção de Leite cru para os seis primeiros meses de 2013

Meses	Limite Inferior (5%)	Previsto	Limite Superior (95%)	Valor Real	Resíduos	Dentro do Intervalo de Confiança?
Janeiro	472.264,6	503.858,2	535.451,9	505.246	1.388	Sim
Fevereiro	412.564,8	450.596,1	488.627,3	439.667	-10.929	Sim
Março	423.002,0	465.454,8	507.907,6	460.825	-4.630	Sim
Abril	405.902,8	451.635,9	497.369,0	443.607	-8.029	Sim
Mai	406.871,0	455.954,5	505.038,1	460.707	4.753	Sim
Junho	391.988,2	445.155,1	498.321,9	458.692	13.537	Sim



**Figura 5** – Valores Reais (Linha Contínua) e Valores Previstos (Linha Tracejada)



o parâmetro  $\Theta_2$  e sem os dois parâmetros simultaneamente e os resultados obtidos foram inferiores para as três situações.

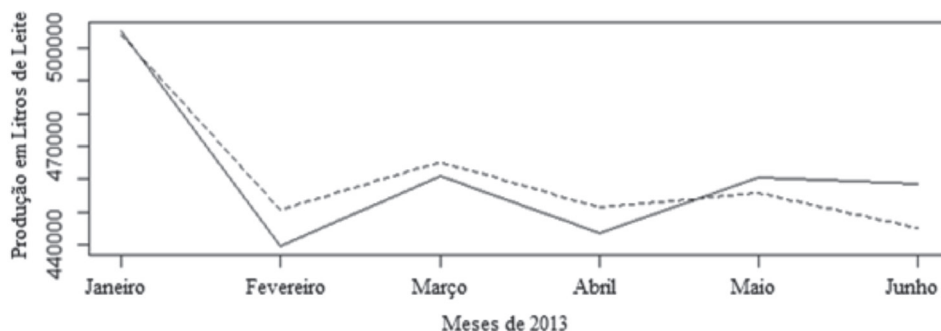
A ordem dos parâmetros  $p$ ,  $P$ ,  $q$ ,  $Q$  do modelo SARIMA está relacionada à memória do processo, ou seja, de quantos valores passados são necessários para poder explicar o futuro. O modelo com  $p = 3$  coeficientes autoregressivos ( $\phi_1, \phi_2, \phi_3$ ) e  $q = 2$  coeficientes de médias móveis ( $\theta_1, \theta_2$ ) obtém estimativas para o tempo  $t$  considerando a contribuição de 3 valores passados regredidos e a soma ponderada de 2 erros aleatórios passados, mais o erro aleatório associado ao tempo  $t$ . O mesmo ocorre para a ordem dos parâmetros sazonais, no entanto, para valores defasados (passados) múltiplos de  $s = 12$ . Neste caso tem-se  $Q = 2$  coeficientes de médias móveis sazonais ( $\theta_1, \theta_2$ ), logo, considera-se ainda a contribuição da soma ponderada dos erros aleatórios defasados em 12 e 24 períodos.

A Figura 5 ilustra o ajuste do modelo SARIMA  $(3,2,1) \times (0,1,2)_{12}$  aos dados históricos da série em estudo. Verifica-se que o modelo conseguiu captar a dinâmica temporal, pois os valores estimados ficaram bem ajustados aos reais. Na Tabela 3 são apresentados os valores previstos para os seis primeiros meses de 2013, os resíduos e

intervalos de confiança de 95%. Na Figura 6, para melhor visualização, é apresentada uma comparação visual da previsão com os valores reais dos seis primeiros meses de 2013.

## CONCLUSÕES

Verificou-se que a série temporal de produção de leite cru no estado de Minas Gerais (MG) apresenta tendência de crescimento e uma componente sazonal com periodicidade de 12 meses. A metodologia Box & Jenkins foi empregada no intuito de obter estimativas de produção de leite para os seis primeiros meses de 2013. O modelo SARIMA  $(3,2,1) \times (0,1,2)_{12}$  obteve os melhores resultados, apresentado o menor erro de acordo com os indicadores selecionados, EQM = 68.850.450 e MAPE = 1,5929%, além de resíduos com características de ruído branco, o que corrobora a qualidade do ajuste. Observa-se que os seis valores reais de produção de leite para o ano de 2013 no estado de Minas Gerais (MG) situam-se dentro dos intervalos de confiança de 95% fornecidos pelo modelo ajustado. Isso é um ótimo resultado, principalmente devido às incertezas e imprevistos associados ao mercado lácteo e também para questões práticas, que envolvem tomadas de decisões e a realização de planejamentos estratégicos futuros.



**Figura 6** – Valores Reais (Linha Contínua) e Valores Previstos (Linha Tracejada)

## REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716-723, 1974.
- ÁLVARES, J. G. Pagamento do leite por sólidos. In: SANTOS, F. A. P.; MOURA, J. C.; FARIA, V. P. **Visão técnica e econômica da produção leiteira**. 5 ed. Piracicaba: FEALQ, 2005. p. 129-140.
- ALVES, F. F.; ERVILHA, G. T.; TOYOSHIMA, S. H. Sazonalidade e previsão do preço do leite recebido pelos produtores da Bahia e de Minas Gerais. In: ENCONTRO DE ECONOMIA BAIANA, 9., 2013, Salvador. **Anais eletrônicos...** Salvador: UFBA, 2013. Disponível em: <[www.eeb.ufba.br/includes/download-anais.php?id=201](http://www.eeb.ufba.br/includes/download-anais.php?id=201)>. Acesso em: 15 out. 2013.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Sistema IBGE de Recuperação Automática (SIDRA): banco de dados agregados**. Disponível em: <<http://www.ibge.gov.br>>. Acesso em: 18 nov. 2013.
- BOX, G. E. P.; JENKINS, G. M. **Time series analysis: forecasting and control**. San Francisco: Holden-Day, 1976. 375 p.
- CARNEIRO, A. V.; YAMAGUCHI L. C. T.; CARVALHO G. R. Sistemas referências de produção de leite: região norte de minas gerais. **Panorama do Leite**, v. 2, n. 23, outubro 2008. Disponível em: <<http://www.ibge.cileite.com.br/panorama/especial23.html>>. Acesso em: 18 nov. 2013.
- FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS (FAO). **Statistical database**. Roma: FAO, 2009.
- HEMME, T.; OTTE, J. **Status and prospects for smallholder milk production – a global perspective**. Rome: Food and Agriculture Organization of the United Nations, 2010. 181 p.
- JENKINS, G. M.; WATTS, D. G. **Spectral analysis and its applications**. San Francisco: Holden-Day, 1968. 525p.
- LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika**, v. 65, n. 2, p. 297-303, 1978.
- MAKRIDAKIS, S.; WHEELWRIGHT, S.; HYNDMAN, R. J. **Forecasting: Methods and Applications**. 3 ed. New York: John Wiley & Sons, 1998. 642 p.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2 ed. **Associação Brasileira de Estatística**. São Paulo: Edgard Blücher, 2006. 538p.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. São Paulo: Edgard Blücher, 2004. 535p.
- PINDYCK, R. S.; RUBINFELD, D. L. **Econometria: modelos e previsões**. Ed. 4. São Paulo: Atlas, 2004. 726 p.
- R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2012. Disponível em: <<http://www.Rproject.org>>. Acesso em: 15 out. 2013
- REZENDE, J. L. P. de et al. **Análise dos preços de carvão vegetal em quatro regiões no**

estado de Minas Gerais. **CERNE**, v. 11, n. 3, p. 237-252, 2005.

SILVA, M. R. et al. Doenças transmitidas pelo leite e sua importância em saúde pública. **Revista do Instituto de Laticínios Cândido Tostes**, v. 62, n. 358, p. 3-18, 2007.

SOUZA, R. et al. Produção e qualidade do leite de vacas da raça Holandesa em função da estação do ano e ordem de parto. **Revista Brasileira de Saúde e Produção Animal**, v. 11, n. 2, p. 484-495, 2010.

TEIXEIRA, N. M. et al. Influência de fatores de meio ambiente na variação mensal da composição e contagem de células somáticas do leite em rebanhos no estado de Minas Gerais. **Arquivo Brasileiro de Medicina**

**Veterinária e Zootecnia**, v. 55, n. 4, p. 491-499, 2003.

TORMAN, V. B. L.; BIRCK, A. R.; RIBOLDI, J. Comparação dos Testes de Aderência à Normalidade Kolmogorov-Smirnov, Anderson-Darling, Cramer-von Mises e Shapiro-Wilk por simulação. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 50.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA (SEAGRO), 11., 2005, Londrina. **Anais eletrônicos...** Londrina: UEL/ Departamento de Estatística e Matemática Aplicada: IBS, 2005. Disponível em: <[http://www.professores.uff.br/cecilia/disciplinas/Testes\\_aderencia.pdf](http://www.professores.uff.br/cecilia/disciplinas/Testes_aderencia.pdf)>. Acesso em: 02 jan. 2014.