

Aplicação de modelo preditivo autoregressivo como ferramenta de gestão do custo da soja

Peterson Diego Kunh (Universidade Tecnológica Federal do Paraná) petersond@utfpr.edu.br
Lucas Duarte Soares (Universidade Estadual do Oeste do Paraná) lucas.2012@alunos.utfpr.edu.br

Resumo: O Brasil é o segundo maior produtor de soja do mundo. Para os agricultores tomarem decisões de forma eficaz relacionados a gestão da cultura é necessário estimar seus custos. Uma forma de estima-los é através de séries temporais por meio de modelos preditivos autoregressivos. Este modelo foi estimado através das funções de autocorrelação e função de autocorrelação parcial. O presente artigo teve como objetivo aplicar um modelo de previsão autoregressivo como ferramenta de apoio à gestão de custos ao produtor de soja. O modelo de previsão obteve resultado satisfatório em relação ao erro mse, com isso o produtor pode realizar decisões relacionados à sua gestão da cultura de forma mais nítida e eficaz.

Palavras chave: Previsão, Gestão Agrícola, Estimação

Application of autoregressive predictive model as a soybean cost management tool

Abstract: Brazil is the second largest soy producer in the world. For farmers to make effective crop management decisions, their costs need to be estimated. One way of estimating them is through time series using autoregressive predictive models. This model was estimated through the autocorrelation functions and partial autocorrelation function. This article aimed to apply an autoregressive forecasting model as a cost management support tool for soybean producers. The prediction model obtained a satisfactory result in relation to the mse error, so that the producer can make clearer and more effective decisions related to his culture management.

Keyword: Forecast, Agricultural Management, Estimation

1. Introdução

De acordo com informações obtidas no endereço eletrônico da EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária, os Estados Unidos são o maior produtor mundial de soja, enquanto que o Brasil ocupa a segunda posição. Na safra brasileira de 2018/2019, segundo o levantamento do CONAB (Companhia Nacional do Abastecimento), a cultura utilizou uma área de 35,822 milhões de hectares, o que gerou uma produção de 114,843 milhões de toneladas e produtividade média de 3.206 kg por hectare.

Os custos são inerentes a todos os tipos de processos produtivos e são fatores determinantes no lucro da empresa. No caso da produção de soja alguns destes custos são: sementes, defensivos, fertilizantes, máquinas e mão-de-obra. Caso o produtor da cultura queira tomar decisões eficazes para aumentar seu lucro, a gestão dos custos é essencial para atingir esse propósito. Portanto o presente artigo tem como objetivo aplicar um modelo preditivo autoregressivo em uma série histórica da estimativa do custo da saca de soja de 60 kgs. As previsões geradas podem ser utilizadas como ferramenta de apoio na gestão de custos do produtor.

Este estudo adota a seguinte estrutura: introdução, que tem como finalidade contextualizar brevemente a produção de soja no Brasil, além da justificativa e objetivo da pesquisa;

referencial teórico, onde é apresentado o método para encontrar o modelo de predição autoregressivo; metodologia utilizada para a realização do artigo e, por fim, são relatadas as considerações finais sobre o trabalho, seguida pelas referências consultadas.

2. Fundamentação teórica

2.1. Gestão de custos

O sucesso das empresas em um mercado tão competitivo é proporcional ao conhecimento relacionado aos seus custos, sejam estes provenientes de suas atividades e/ ou processos. Por consequência, controlar os custos é prioridade para garantir a sobrevivência do negócio.

Martins (2010) conceitua gasto como toda movimentação que ocorre em uma empresa. Desta forma, qualquer entrada ou saída é vista como gasto. Custos são gastos que as empresas desembolsam para produzir e/ou comercializar seus produtos, assim como os gastos gerados para prestação de determinados serviços (TEIXEIRA et al., 2015).

Tendo conhecimento do que são custos e seus impactos nas organizações, sua gestão torna-se de extrema importância para a saúde financeira do negócio. A gestão de custos, de acordo com Bruni e Famá (2004), é definida como um conjunto de ferramentas e técnicas que buscam a análise, o planejamento e a melhoria contínua do desempenho dos processos produtivos de uma empresa.

A previsão dos custos pode ser uma informação de extrema importância para o gestor que busca um planejamento eficaz, pois a empresa estará melhor preparada para situações futuras, com isso se comportará de maneira proativa enquanto que seus concorrentes ficarão em uma posição de vulnerabilidade em relação à cenários desconhecidos.

A planilha eletrônica que foi utilizada no presente artigo contém as estimativas dos custos (em reais) de produção de soja no Paraná por saca de 60 kgs. As observações estão compreendidas entre o período de 1995 até 2018 e foram coletadas anualmente. O arquivo pode ser encontrado no seguinte endereço eletrônico: <http://www.agricultura.pr.gov.br/modules/conteudo/conteudo.php?conteudo=228>. A Tabela 1 mostra os dados contidos nesta.

SAFRA	MÊS	CUSTO DE PRODUÇÃO POR 60 kg	
		VARIÁVEL(R\$)	TOTAL(R\$)
1994/95	fev/94	3.660,15	6.344,06
1995/96	fev/95	7,65	13,42
1996/97	fev/96	7,65	13,02
1997/98	abr/97	6,21	9,62
1998/99	fev/98	6,16	9,73
1999/00	fev/99	7,65	11,41
2000/01	fev/00	8,28	12,49
2001/02	fev/01	8,51	13,26
2002/03	fev/02	9,95	15,48
2003/04	fev/03	14,04	21,41
2004/05	mai/04	19,59	30,37
2005/06	nov/05	17,23	27,74
2006/07	fev/06	17,34	28,08
2007/08	fev/07	16,00	26,79
2008/09	fev/08	18,93	30,46
2009/10	fev/09	24,49	38,24

2010/11	fev/10	20,68	34,37
2011/12	fev/11	18,91	33,62
2012/13	fev/12	21,69	36,75
2013/14	fev/13	25,44	42,85
2014/15	fev/14	27,32	46,08
2015/16	fev/15	28,50	49,10
2016/17	fev/16	34,41	55,39
2017/18	fev/17	32,90	55,41

Tabela 1 - estimativa dos custos de produção de soja no Paraná por saca de 60 kgs

Fonte: SEAB/DERAL

2.2. Previsão

Previsão consiste em estipular dados futuros, seja de maneira quantitativa ou qualitativa. Estes dados podem ser caracterizados como o custo de fabricação de determinado produto, a demanda por algum serviço, o consumo de matéria-prima para produzir uma peça, a quantidade de horas gastas com paralizações em um processo produtivo, etc. Portanto, conforme Pellegrini e Fogliatto (2001), no gerenciamento da produção as previsões são de extrema importância na operacionalização de vários aspectos, tais como planejamento e desenvolvimento da produção, gestão de estoques e tomadas de decisão.

Os métodos de previsão são divididos entre modelos qualitativos e quantitativos. Os qualitativos, segundo Moreira (2000), são baseados em informações de caráter subjetivo, em outras palavras, no julgamento e experiência pessoal do gestor. Já os quantitativos, conforme Pinto (2006), são modelados matematicamente baseado em dados históricos. Estes dados históricos são componentes de séries temporais que, segundo Morettin e Tolo (1986), “são um conjunto de observações ordenadas no tempo”

Neste artigo utilizou-se o modelo quantitativo Auto-regressivo de Médias Móveis (ARMA) para verificar sua eficiência na previsão dos dados históricos da Tabela 1. A série temporal da estimativa de custo de produção da saca de soja de 60 kgs pode ser vista na Figura 1.

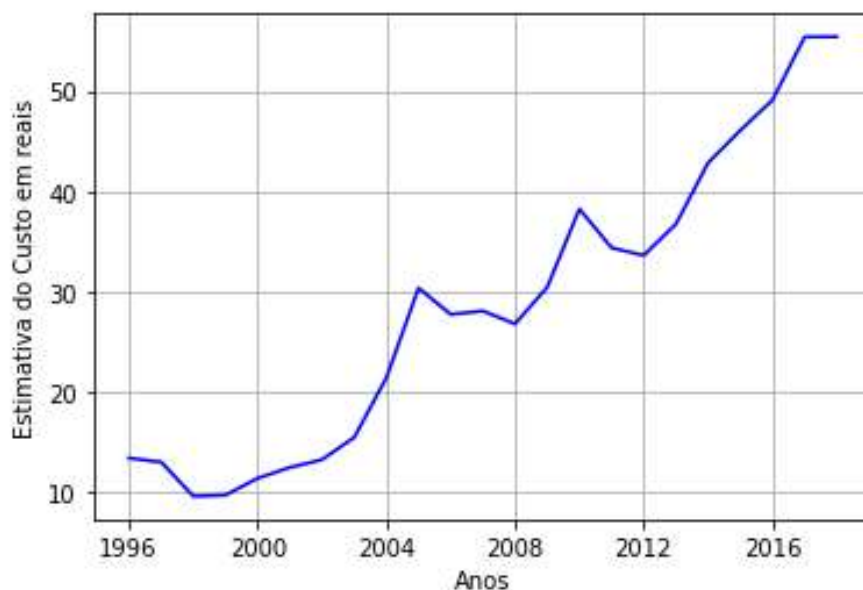


Figura 1 - série temporal da estimativa de custo de produção da saca de soja de 60 kgs

Fonte: autor

2.3. Boxplot

Antes de realizar a previsão de uma série histórica é necessário fazer uma análise estatística dos dados para verificar se não há alguma observação que possa comprometer a eficiência do modelo preditivo. Triola (2008) comenta que os diagramas de caixa, conhecidos com Boxplots, são convenientes para revelar tendências centrais, dispersão, distribuição dos dados e a presença de outliers (valores extremos). A Figura 2 mostra o diagrama de caixa da série histórica da Tabela 1.

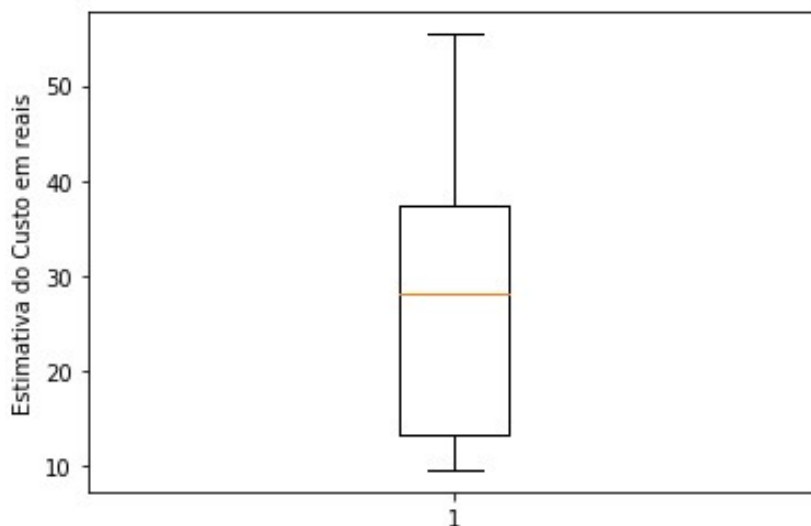


Figura 2 - boxplot da estimativa de custo de produção da saca de soja de 60 kgs
Fonte: Autor

Observa-se que nenhum valor da série temporal ultrapassou os limites outliers estabelecidos; portanto, não houve necessidade de remover qualquer valor dessa série.

2.4. Autocorrelação

As análises de correlação permitem verificar se as variáveis de estudo variam no mesmo sentido (coeficiente de correlação positivo), em sentidos opostos (coeficiente de correlação negativo) ou se não há correlação entre as variáveis (coeficiente de correlação zero) (Baptista, 2009).

Quando trata de autocorrelação não existe duas variáveis, mas, sim, uma. A análise é feita entre a variável em um determinado tempo com ela mesmo em um instante de tempo anterior, essa diferença de tempo é chamada de lag.

Uma maneira rápida, fácil e gráfica para ver se a série histórica possui autocorrelação é plotar a observação futura, $y(t+1)$, em função da observação presente, $y(t)$. A Figura 3 ilustra esta situação

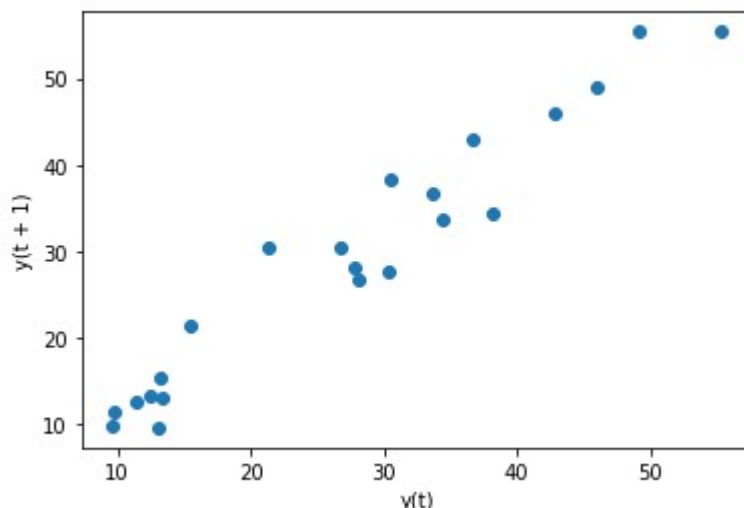


Figura 3- Autocorrelação da estimativa de custo de produção da saca de soja de 60 kgs
Fonte: Autor

Analisando a Figura 3 é possível verificar que os pontos estão alinhados na diagonal de forma crescente. Essa situação demonstra um forte indício de autocorrelação positiva da série temporal. Portanto, um modelo Autoregressivo aparenta ser eficiente nas previsões

2.5. Modelos Autoregressivo

Segundo Morettin e Tolo (2006), um modelo preditivo autoregressivo de ordem p, AR(p), possui a seguinte estrutura demonstrada pela Equação 1.

$$Z_t = \Phi_1 \cdot Z_{t-1} + \Phi_2 \cdot Z_{t-2} + \dots + \Phi_p \cdot Z_{t-p} + a_t \quad (1)$$

onde Z é a observação da variável em determinado instante de tempo t, phi é os parâmetros do modelo e a é um ruído branco, que, segundo Bueno (2008), é caracterizado por possuir média zero, variância constante e autocorrelação igual à zero

2.6 Modelos Médias Móveis

Segundo Morettin e Tolo (2006), um modelo preditivo médias móveis de ordem q, MA(q), possui a seguinte estrutura demonstrada pela Equação 2, para um processo de média nula.

$$Z_t = \mu + a_t - \theta_1 \cdot a_{t-1} + \theta_2 \cdot a_{t-2} - \dots - \theta_q \cdot a_{t-q} \quad (2)$$

onde Z é a observação da variável em determinado instante de tempo t, mi é a constante nível da série temporal, teta é os parâmetros do modelo e a é um ruído branco.

2.7 Modelos Autoregressivo e de Médias Móveis

Segundo Morettin e Tolo (2006), um modelo preditivo autoregressivo e de médias móveis representado por ARMA(p,q), é a combinação das Equações 1 e 2, resultando na Equação 3.

$$Z_t = \Phi_1 \cdot Z_{t-1} + \Phi_2 \cdot Z_{t-2} + \dots + \Phi_p \cdot Z_{t-p} - \theta_1 \cdot a_{t-1} - \theta_2 \cdot a_{t-2} - \dots - \theta_q \cdot a_{t-q} + a_t \quad (3)$$

2.8 Identificação do Modelo

Para a estimação da ordem p e q do modelo ARMA, Bueno (2006) mostra que as características da FAC e da FACP em cada caso, é que indicarão qual o possível processo gerador da série

Modelo	Comportamento FAC	Comportamento FACP
AR(p)	Infinita, exponencial e/ou senóide amortecida	Finita, corte brusco zerando após lag p
MA(q)	Finita, corte brusco zerando após lag q	Infinita, exponencial e/ou senóide amortecida
ARMA(p,q)	Infinita, exponencial e/ou senóide amortecida após lag (q-p)	Infinita, exponencial e/ou senóide amortecida após lag (p-q)

Quadro 1 - Estimação de Parâmetros (p,q) através das análises de FAC e FACP

Fonte: Adaptado de Bueno (2006) e Morettin e Toloi (2006)

A Figura 4 e 5 apresentam a FAC e FACP da série histórica da estimativa de custos de produção de soja no Paraná por saca de 60 kgs. Com o auxílio do Quadro 1 verifica-se que o modelo mais adequado é o autoregressivo de ordem 1, AR(1).

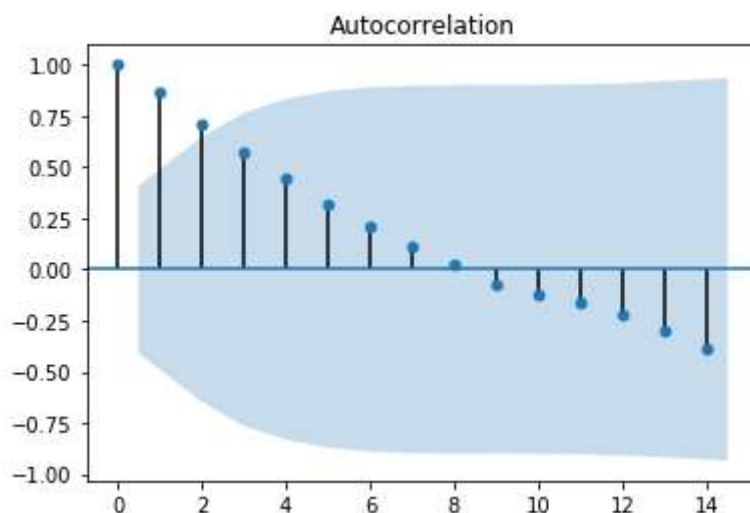


Figura 4 - FAC da estimativa de custo de produção da saca de soja de 60 kgs

Fonte: Autor

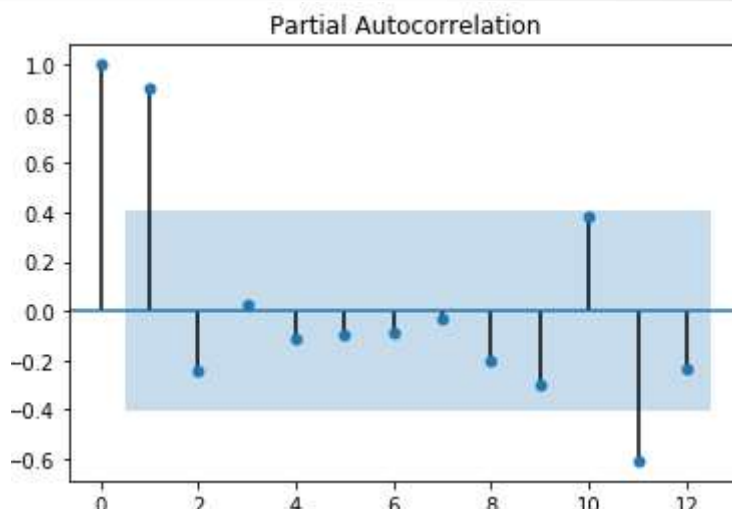


Figura 5 - FACP da estimativa de custo de produção da saca de soja de 60 kgs
Fonte: Autor

2.9 Estimação dos parâmetros e previsão

Para estimar os parâmetros do Modelo Autoregressivo de primeira ordem foi necessário dividir a amostra da Tabela 1 em duas. A primeira amostra, chamada de amostra treino, foi construída a partir da remoção da última observação (55,41) da série histórica e mantendo os valores restantes. O valor excluído da amostra treino foi guardado para compor a amostra teste.

O novo conjunto de dados formado pela amostra treino foi inserido em um código desenvolvido pelo autor em python para encontrar os parâmetros que são mostrados na Equação 4.

$$Z_t = 1,05061159Z_{t-1} + 0,7709119 \quad (4)$$

Os coeficientes da Equação 4 foram definidos pelo erro MSE (mean squared error) apresentado pela Equação 5.

$$\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (5)$$

Esta medida de acurácia foi utilizada como função objetivo a ser minimizada pelas variáveis de decisão (parâmetros do modelo AR(1)), ou seja, os parâmetros apresentados pela Equação 4 são os que minimizam o erro MSE entre a amostra treino com os seus valores previstos pela Equação 4.

A Figura 6 mostra a comparação dos valores previstos e esperados da amostra treino bem como o seu erro MSE associado.

```
-----  
Previsto: 14.45405412801163 Esperado: 13.02  
-----  
Previsto: 10.881974713306825 Esperado: 9.62  
-----  
Previsto: 10.997541988488452 Esperado: 9.73  
-----  
Previsto: 12.76256946398965 Esperado: 11.41  
-----  
Previsto: 13.897229983954706 Esperado: 12.49  
-----  
Previsto: 14.706200910226089 Esperado: 13.26  
-----  
Previsto: 17.038558645709813 Esperado: 15.48  
-----  
Previsto: 23.268685389592022 Esperado: 21.41  
-----  
Previsto: 32.68216525893175 Esperado: 30.37  
-----  
Previsto: 29.919056770498322 Esperado: 27.74  
-----  
Previsto: 30.276264711968803 Esperado: 28.08  
-----  
Previsto: 28.920975757566097 Esperado: 26.79  
-----  
Previsto: 32.77672030226217 Esperado: 30.46  
-----  
Previsto: 40.95047849238082 Esperado: 38.24  
-----  
Previsto: 36.884611629172696 Esperado: 34.37  
-----  
Previsto: 36.09665293475251 Esperado: 33.62  
-----  
Previsto: 39.38506721946606 Esperado: 36.75  
-----  
Previsto: 45.79379793408351 Esperado: 42.85  
-----  
Previsto: 49.18727337805307 Esperado: 46.08  
-----  
Previsto: 52.36012038758499 Esperado: 49.1  
-----  
Previsto: 58.96846730478888 Esperado: 55.39  
-----  
O erro MSE foi de: 5.255680740135167
```

Figura 6 - Valores previstos e esperados da amostra treino
Fonte: Autor

O gráfico contendo as informações da Figura 6 é apresentado na Figura 7, onde verifica-se que o modelo Auto regressivo de primeira ordem conseguiu aderir bem as variações da amostra treino. Entretanto todas as previsões ficaram acima das observações.

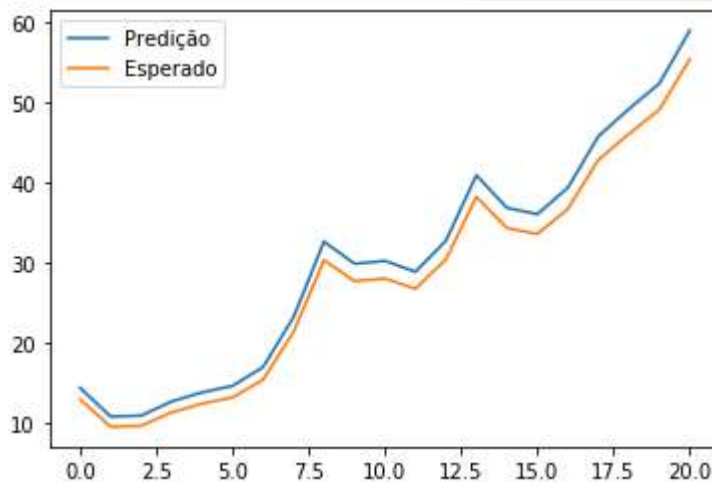


Figura 7- Gráfico dos valores previstos e esperados da amostra treino
Fonte: Autor

O modelo autoregressivo de primeira ordem encontrado pela amostra treino foi utilizado na amostra teste para verificar sua eficiência na extrapolação dos dados.

A Figura 8 mostra a comparação do valor previsto e esperado da amostra teste bem como o seu erro MSE associado.

```
-----
Previsto: 52.36012038758499 Esperado: 55.41
-----
O erro MSE foi de: 9.301765650224699
```

Figura 8 - Valores previstos e esperados da amostra treino
Fonte: Autor

Analisando a Figura 8 verifica-se que o valor previsto (52,36) pela Equação 4 na amostra teste foi menor que o esperado (55,41) resultando em um erro mse de 9,30.

3. Metodologia

A natureza da pesquisa proposta é classificada como sendo aplicada pois, segundo Silva e Menezes (2005), pesquisas dessa natureza possuem interesse na utilização dos resultados obtidos para que possam auxiliar na solução de problemas específicos reais. A problemática da pesquisa pode ser caracterizada como prever com precisão a estimativa de custo de produção da saca de soja de 60 kgs.

Quanto à abordagem, a pesquisa é quantitativa. Sampieri, Collado e Lucio (2006) definem uma pesquisa quantitativa como aquela que utiliza de aferição numérica para validar o método proposto por meio de recursos e técnicas estatísticas. A linguagem de programação python foi utilizada para desempenhar essa função no presente projeto.

Nas palavras dos mesmos autores, uma pesquisa é considerada descritiva em relação aos objetivos quando busca especificar características importantes de um determinado fenômeno. O fenômeno de estudo se refere à estimativa de custo de produção da saca de soja de 60 kgs.

Os procedimentos técnicos da pesquisa são caracterizados como modelagem, pois, segundo

Morabito e Pureza (2010), visam representar a realidade de maneira sistemática. O método Autoregressivo foi utilizado para realizar essa representação.

4. Conclusão

Estimar valores futuros do custo de produção de soja é fundamental para o planejamento, gestão e tomada de decisão do produtor rural. Portanto, o presente artigo expôs o modelo de previsão autoregressivo como ferramenta para auxiliar a gestão de custos de produção agrícola do grão.

O modelo preditivo autoregressivo de primeira ordem, AR(1), demonstrou-se satisfatório para representar de maneira determinística o comportamento estocástico dos custos de produção de soja, quando utilizado o erro mse como métrica, tanto para a amostra treino (erro 5,25) quanto para a amostra teste (erro 9,30). Logo, o produtor da cultura pode realizar decisões de mercado de forma mais consciente e eficaz.

Para futuros trabalhos pretende-se aplicar na série histórica o modelo estatístico de previsão ARIMA com o objetivo de minimizar o erro mse. A minimização deste erro é resultante de previsões mais próximas da realidade, com isso melhores decisões poderão ser tomadas.

Refências

BAPTISTA, J.A.A. **Nível de conhecimento sobre o protocolo de Kyoto pelas empresas componentes do índice de sustentabilidade empresarial da Bovespa**. São Caetano do Sul, 2009

BRUNI, Adriano Leal; FAMÁ, Rubens. **Gestão de custos e formação de preços**. 3. ed. São Paulo: Atlas, 2004.

BUENO, R. L. S. **Econometria de Séries Temporais**, Editora Cengage Learning, 2008

MARTINS. Eliseu. **Contabilidade de custos**. 10. ed. São Paulo: Editora Atlas, 2010.

MORABITO, R.; PUREZA, V.; Modelagem e Simulação. In: MIGUEL, P.A.C. **Metodologia de Pesquisa em Engenharia de Produção e Gestão de Operações**. Rio de Janeiro: Elsevier, 2010

MOREIRA, Daniel A. **Administração da Produção e Operações**. São Paulo. Pioneira.2000

MORETTIN, P. A.; TOLOI, C. M. **Séries Temporais**. São Paulo: Atual, 1986

MORETTIN, P; TOLOI, C. **Análise das Séries Temporais**. 2. ed. São Paulo. Editora Universidade de São Paulo, 2006.

PELLEGRINI, F. R. P.; FOGLIATTO, F. S. **Passos para implantação de sistemas de previsão de demanda: técnicas e estudo de caso**. Produção, v. 11, p. 43-64, 2001

PINTO, K. A. da C. O desafio do planejamento de compra em um hospital oncológico com foco em sua previsão de demanda: estudo de caso. **SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO**, 21, 2014. Bauru, SP, Brasil, 10 a 12 de Novembro de 2014. Anais... XXI SIMPEP, 2014.

SAMPIERI, R. H.; COLLADO, C. F.; LUCIO, P. B. **Metodologia de Pesquisa**. 3.ed. São Paulo: McGraw Hill Book, 2006.

SILVA, Edna Lúcia da; MENEZES, Estera Muszkat. **Metodologia da Pesquisa e Elaboração de Dissertação**. 4.ed. rev. atual. Florianópolis: UFSC, 2005.

TEIXEIRA, A. C. C. P.; ASSIS NETO, A. G.; FERREIRA, F. J. A utilização de conceitos de custos e sua influência na decisão do preço de venda nas micro empresas e empresas de pequeno porte. **Revista de Micro e Pequenas Empresas e Empreendedorismo da Fatec Osasco**, Osasco, v. 1, n. 2, p. 206-222, 2015

TRIOLA, M. F. **Introdução à estatística**. 10. ed. Rio de Janeiro: LTC, 2008