

## Aplicação de modelagem matemática para previsões da cultura de batata no Brasil

## Application of mathematical modelling for potato crop predictions in Brazil

### RESUMO

**Ingridy Maria Xavier Miranda**  
[ingridy\\_smi@hotmail.com](mailto:ingridy_smi@hotmail.com)  
Universidade Tecnológica Federal do Paraná, Medianeira, Paraná, Brasil

**Carla Adriana Pizarro Schmidt**  
[carlaschmidt@utfpr.edu.br](mailto:carlaschmidt@utfpr.edu.br)  
Universidade Tecnológica Federal do Paraná, Medianeira, Paraná, Brasil

**Dayane Regina Trage**  
[dayaneregina@hotmail.com](mailto:dayaneregina@hotmail.com)  
Universidade Tecnológica Federal do Paraná, Medianeira, Paraná, Brasil

**Camila Ciello**  
[camila\\_ciello@outlook.com](mailto:camila_ciello@outlook.com)  
Universidade Tecnológica Federal do Paraná, Medianeira, Paraná, Brasil

**OBJETIVO:** O principal objetivo da pesquisa foi realizar as previsões da produtividade da batata, para os próximos 12 meses, utilizando modelos estatísticos. **MÉTODOS:** a pesquisa foi baseada nos dados disponibilizados pelo IBGE, entre 2007 e 2018. Assim, foram coletadas informações sobre área plantada, área colhida, produção e rendimento médio. Posteriormente, foram realizadas análises com o auxílio dos softwares Microsoft® Excel, Action Stat Pro e NNQ. A estatística descritiva apontou o comportamento dos dados e norteou a aplicação da modelagem matemática. Para calcular as previsões, aplicou-se três modelos, sendo eles: de suavização exponencial simples, suavização exponencial Holt e ARIMA. Por fim, as medidas de acurácia apontaram qual o modelo de melhor ajuste na série de dados. **RESULTADOS:** O melhor método identificado foi o ARIMA. Porém, a não estacionariedade dos dados observados, a ausência da ordem das médias móveis e da autocorrelação dos resíduos, acarretou em valores de previsões constantes. **CONCLUSÕES:** Para este estudo, exceto no rendimento médio, o segundo melhor modelo de previsão é o mais indicado. Neste caso, sugere-se uma análise mais profunda sobre a comercialização e exportação, atrelando a oferta da batata.

**PALAVRAS-CHAVE:** Séries temporais. Estatística. Oferta.

**Recebido:** 19 ago. 2019.

**Aprovado:** 01 out. 2019.

**Direito autorial:** Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



### ABSTRACT

**OBJECTIVE:** The main objective of the research was to make potato productivity predictions for the next 12 months using statistical models. **METHODS:** The research was based on data provided by IBGE between 2007 and 2018. Thus, information on planted area, harvested area, productivity and average yield was collected. Posteriorly analyzes were performed using Microsoft® Excel, Action Stat Pro and NNQ software. Descriptive statistics pointed the data behavior and guided the application of mathematical modeling. To calculate the forecasts, three models were applied: simple exponential smoothing, Holt exponential smoothing and ARIMA. Finally, the accuracy measures indicated which model best fit the data series. **RESULTS:** The best method identified was ARIMA. However, the non-stationarity of the observed data, the absence of the order of moving averages and the autocorrelation of the residuals, resulted in constant prediction values. **CONCLUSIONS:** For this study, except for the average yield, the second best forecasting model is the most indicated. In this case, a deeper analysis on marketing and exportation is suggested, linking the offer of potatoes.

**KEYWORDS:** Time series. Statistic. Offer.

## INTRODUÇÃO

A batata é um alimento importante mundialmente, ocupando a terceira posição na cultura alimentar. Isto ocorre devido a sua composição nutricional e versatilidade na forma de consumo. (SILVA, 2016).

Para manter a produtividade, a batata é plantada em três épocas do ano, na qual se denomina como das águas, da seca e do inverno. O plantio da primeira safra ocorre entre agosto e dezembro, sendo colhida entre dezembro e março. Já a segunda safra inicia a plantação em janeiro até abril, onde a oferta ocorre entre abril e agosto. Já a de inverno, é plantada em maio até setembro, na qual é colhida de setembro a novembro (GODOY, 2001).

Analisando o desempenho positivo no saldo da balança comercial é de extrema importância o desenvolvimento e manutenção do agronegócio (FILHO E ALVES, 2005). Sendo assim, visando a tomada de decisões estratégicas, faz-se essencial a estimação dos índices de produtividade para os próximos períodos.

## MATERIAL E MÉTODOS

O estudo será realizado com base nos dados coletados pelo Instituto Brasileiro de Geografia e Estatística (IBGE, 2019), disponibilizados no Sistema IBGE de Recuperação Automática (SIDRA). Logo, foram acessados dados referentes a área plantada, área colhida, quantidade produzida e rendimento médio nacional da 1ª safra, de 2007 a 2018. Apesar do período de plantio e colheita da primeira safra ser específico, o IBGE apresenta, na tabela 6588, dados mensais.

Os dados brutos foram analisados com o auxílio dos softwares Microsoft® Excel (MICROSOFT, 2016), Action Stat Pro® (PORTAL ACTION, 2018) e NNQ - Estatística (NNQ, 2019), onde respectivamente foram responsáveis pela estatística descritiva e desenvolvimento dos modelos de séries temporais.

O primeiro método de previsão aplicado foi o de suavização exponencial simples. Este método é eficaz para previsões de curto prazo, onde os dados observados recentes têm pesos maiores ao realizar o cálculo. Segundo Dias (1993), o método de suavização exponencial simples não considera tendência, sazonalidade ou quaisquer reações de valores aleatórios.

Outro método aplicado, foi o de Suavização Exponencial de Holt. Também conhecido como Suavização Exponencial Dupla, considera a tendência dos dados observação ao realizar o cálculo de previsão.

O último modelo utilizado foi o ARIMA, conhecido como a metodologia de Box e Jenkins. Segundo Morettin e Tolo (2004), “consiste em ajustar modelos auto-regressivos integrados a médias móveis – ARIMA(p, d, q) – a um conjunto de dados”, seguindo um procedimento definido.

Então, através das medidas de acurácia foi possível determinar qual o melhor modelo a ser ajustado. Logo, foram calculados o Mean Absolute Percentage Error (MAPE), o Mean Absolute Deviation (MAD) e o Mean Squared Deviation (MSD).

## RESULTADOS E DISCUSSÃO

A primeira análise realizada foi a estatística descritiva, auxiliando a identificar o comportamento dos dados. Logo, uma das medidas observadas é o desvio padrão. Para definir a dimensão do desvio padrão, o coeficiente de variação é um dos melhores parâmetros. Portanto, o coeficiente de variação é utilizado para mensurar o grau de dispersão dos dados em torno da média, tendo como níveis: baixa ( $CV \leq 15\%$ ), média ( $15\% < CV < 30\%$ ) e alta ( $CV \geq 30\%$ ) (MEDRI, 2011). Então, as quatro variáveis estudadas apresentaram coeficiente de variação entre 5 e 10%, ou seja, considerado baixo.

Outro aspecto a ser observado em uma série de dados é a presença de tendência e sazonalidade. A tendência é o comportamento de longo prazo, que pode ser causada por infinitos fatores, afetando a variável de interesse (NNQ, 2019).

A área plantada e colhida, apresentam uma tendência decrescente. Ou seja, ao longo dos anos, houve uma redução das duas variáveis. Entretanto, o rendimento e a produção se comportaram de forma contrária, tendo um aumento significativo.

De acordo com o Boteon e Deleo (2004), isto ocorre devido as “novas regiões produtoras, estrutura fundiária, nível tecnológico e aspectos gerenciais das propriedades”. Além do mais, houve uma baixa nos preços da batata devido a interferência climática, acarretando produtos de baixa qualidade. Godoy (2001) destaca que houve avanços tecnológicos significativos, resultando no aumento constante da produtividade ao longo dos últimos anos.

A sazonalidade também é um fator importante a ser considerado ao realizar previsões com séries temporais. A sazonalidade são variações repetitivas anuais e com durabilidade máxima de um ano. (MORETTIN E TOLOI, 2004). Assim, o software NNQ apresentou que para a área colhida, plantada e produção, o melhor modelo para o índice de sazonalidade foi o multiplicativo. Já para o rendimento, foi a sazonalidade aditiva.

Para a área colhida e plantada, há uma queda no primeiro semestre com o maior pico nos meses de janeiro e fevereiro e aumento no segundo semestre, com maior pico em dezembro.

Para a produção, de setembro a janeiro os índices foram menores, com enfoque em fevereiro. Já o rendimento, de agosto a dezembro apresentaram saldos negativos, principalmente em dezembro. Porém, no primeiro semestre houve um aumento, sendo janeiro o pico.

O período de plantio e colheita da primeira safra propicia essa sazonalidade. Portanto, a maioria produz nesta época, pois não utilizam frequentemente a irrigação, devido à alta pluviosidade nas regiões (CEASA, 2017).

## MÉTODOS DE PREVISÕES

Primeiramente, foi verificado qual é o melhor método pelo NNQ. Assim, para cada variável analisada obteve – se o melhor modelo de suavização exponencial. Para as três primeiras variáveis, o software apontou que o MNM seria o mais

apropriado, sendo com erros multiplicativos, sem tendência e sazonalidade multiplicativa. Já para o rendimento médio, o MNA (erros multiplicativos, nenhuma tendência e sazonalidade aditiva) foi o melhor ajuste.

Os modelos obtidos pelo método de suavização exponencial de Holt foram através do Action. Além do nível, foi levado em conta no cálculo a tendência. Porém, as medidas de acurácia da produção e rendimento médio foram maiores que o método de suavização simples. Por isso, nestes casos, o primeiro modelo foi o melhor ajuste.

Para todas as variáveis estudadas, o modelo autorregressivo integrado que o Action ajustou foi o ARIMA(0,1,0). Logo, foi possível observar que a série de dados não apresentou estacionaridade, devido a tendência e sazonalidade. Assim, conclui-se que não há ordem das médias móveis e autocorrelação dos resíduos ( $p=q=0$ ) e há apenas uma diferenciação ( $d=1$ ). Esta diferenciação é o suficiente para tornar o modelo estacionário, alcançando um bom ajuste. Logo, comparado aos demais modelos, este foi o que obteve o melhor ajuste, conforme a tabela 1.

Tabela 1. Medidas de acurácia do ARIMA (0,1,0) no Action.

Medidas de acurácia	Área plantada (Hectares)	Área colhida (Hectares)	Produção (Hectares)	Rendimento médio (Quilograma por Hectare)
MAPE	0,8921	0,8949	1,1896	0,57
MAD	596,64	598,16	20071	145,01
MSD	1838616,51	1835902,44	2044877464	116358,50

Fonte: Autoria Própria.

Ao simular as previsões para os próximos 12 meses, verificou-se que o modelo ARIMA apresentou resultados constantes para a área plantada, área colhida e produção. Isso ocorre devido a não estacionaridade e repetição constante de dados observados mensalmente. Por isso, será demonstrado na tabela 2, os valores previstos correspondentes ao segundo melhor modelo para cada variável, exceto para o rendimento médio.

Tabela 2. Previsão dos melhores modelos para cada variável.

	Área plantada (Hectares) - Suavização Exponencial de Holt	Área colhida (Hectares) - Suavização Exponencial de Holt	Produção (Toneladas) - Suavização Exponencial (MNM)	Rendimento médio (Quilograma por Hectare) –ARIMA
1	58758,891	58784,996	1645660,60	28040,0909
2	58676,757	58708,21178	1639559,57	28085,1818
3	58594,623	58631,42694	1662932,67	28130,27273
4	58512,48859	58554,6421	1664123,20	28175,36364
5	58430,35419	58477,85726	1660992,65	28220,45455
6	58348,21978	58401,07242	1657510,98	28265,54545
7	58266,08538	58324,28758	1657462,98	28310,63636
8	58183,95097	58247,50274	1657815,08	28355,72727
9	58101,81656	58170,7179	1660197,59	28400,81818
10	58019,68216	58093,93306	1649239,38	28445,90909
11	57937,54775	58017,14822	1648885,08	28491
12	57855,41335	57940,36338	1648367,83	28536,09091

Fonte: Autoria Própria.

## CONCLUSÃO

O uso dos modelos de previsões são de extrema importância para a agricultura, visto que movimentam boa parte da economia brasileira. Mesmo com vários fatores incontrolláveis que afetam o cultivo, a “antecipação do futuro” auxilia nas tomadas de decisões e precauções, adequando os locais e épocas de plantio e mantendo a produtividade que atendam o mercado.

O estudo apresentou que o melhor modelo para realizar as previsões para os próximos 12 meses é o ARIMA. Porém, devido a ausência de estacionaridade, as previsões permaneceram constantes. Por isso, o segundo melhor modelo é o mais aplicável.

Para pesquisas futuras, sugere-se analisar minuciosamente a comercialização e exportação da batata, relacionando oferta e demanda. Também, estudar a série de dados temporais da segunda e terceira safra, verificando se possuem o mesmo comportamento que a primeira safra e possíveis relações

## REFERÊNCIAS

- BOTEON, M.; DELEO, J.P. **Comercialização de batata e consumo no Brasil**. 2004. Disponível em: <https://www.cepea.esalq.usp.br/br/documentos/texto/comercializacao-de-batata-consumo-no-brasil-novembro-2004.aspx>. Acesso em: 9 de julho de 2019.
- CEASA PARANÁ. **Banco de dados. Preços médios dos hortifrutícolas**. 2019. Disponível em: <https://www.hfbrasil.org.br/br/banco-de-dados-precos-medios-dos-hortifruticolas.aspx>. Acesso em: 16 de agosto de 2019.
- DIAS, M. A. (1993). **Administração de materiais: uma abordagem logística**. 4. ed. São Paulo: Atlas.
- FILHO, W. P. C; ALVES, H. S. **Mercado de batata no Brasil: análise de produção, importação e preços**. São Paulo, RS. 2005. Disponível em: <http://www.iea.sp.gov.br/ftp/iea/ie/2005/seto2-0505.pdf>. Acesso em: 13 de junho de 2019.
- GODOY, R. C. B. **Batata**. 2001. Disponível em: <http://www.agricultura.pr.gov.br/arquivos/File/deral/cultura7.pdf>. Acesso em: 9 de julho de 2019.
- IBGE – Instituto Brasileiro de Geografia e Estatística. **Banco de tabelas estatísticas do IBGE**. 2019. Disponível em: <https://sidra.ibge.gov.br>. Acesso em: 04 de fevereiro de 2019.
- MEDRI, W. **Análise exploratória de dados**. 2011. Disponível em: [http://www.uel.br/pos/estatisticaquantitativa/textos\\_didaticos/especializacao\\_e\\_statistica.pdf](http://www.uel.br/pos/estatisticaquantitativa/textos_didaticos/especializacao_e_statistica.pdf). Acesso em: 9 de junho de 2019.
- MICROSOFT. **Planilha Microsoft Excel**. Versão 2016. Disponível em: <https://products.office.com/pt-br/excel>. Acesso em: 06 de março de 2019.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. São Paulo: Edgard Blucher, 2004
- NNQ – Núcleo de Normalização e Qualimetria. **Software NNQ - Estatística**. Florianópolis: Universidade Federal de Santa Catarina, 2019. Disponível em: <http://qualimetria.ufsc.br/publicacoes/software/previsao/>. Acesso em: 14 de julho de 2019.
- PORTAL ACTION. **Action Stat**. 2018. Disponível em: <http://www.portalaction.com.br>. Acesso em: 14 de maio de 2019.
- SILVA, P. N.L; JUNIOR, R. F.; SANTOS, E. F. **Conhecimento do consumidor e forma de apresentação da batata no mercado no Estado de São Paulo**. 2016. Disponível em: <http://oaji.net/articles/2017/5879-1523910487.pdf>. Acesso em: 9 de junho de 2019.