



## APLICAÇÃO DE MODELO ESTATÍSTICO PARA ESTIMATIVA DA PRODUÇÃO DE SOJA NA REGIÃO DE BALSAS-MA

REGO, Nadyne Gomes<sup>1</sup>  
ARAUJO, Déborah Pimentel de<sup>2</sup>  
SOUSA NETO, João Valério de<sup>3</sup>

**Resumo:** A economia da região sul do Maranhão é caracterizada pela agricultura, tendo como destaque a produção de soja na região de Balsas-MA. Este trabalho propõe a aplicação de um modelo estatístico para estimativa de produções futuras. Tais dados de produções anuais foram requisitados ao IBGE. Após a tabulação dos dados, foi aplicado o Método dos Mínimos Quadrados (MMQ), um modelo de regressão linear simples com o objetivo de encontrar uma função numérica de melhor ajuste à produção. Obteve-se o coeficiente de Pearson de 0,97, comprovando a correlação muito forte entre as variáveis. Com isso, constatou-se que o coeficiente de determinação de 0,94 encontrado, indica que 94% da produção de soja pode ser explicada pelo modelo aplicado. Com base nos resultados numéricos obtidos, através da aplicação de tal modelo, comprova-se que o modelo explica o acentuado crescimento das produções.

**Palavras-chave:** Produção de soja. Modelo estatístico. Estimativa. Regressão Linear. Método dos Mínimos Quadrados (MMQ).

**Abstract:** The economy of Maranhão region south is characterized by agriculture, with the highlight soybean production in the Balsas - MA region. This paper proposes the application of a statistical model to estimate future productions. For such study the data annual productions were required to IBGE. After tabulating the data, the Method of Least Squares (OLS), a simple linear regression model with the goal of finding a numerical function better fit the production was applied. Was obtained the Pearson coefficient of 0.97, proving the strong correlation between the variables. Thus, it was found that the coefficient of determination found 0.94 indicates that 94% of soybean production can be explained by the model applied. Based on the numerical results obtained by applying such a model, it proves that the model explains the sharp increase in annual production.

**Keywords:** Soybean production. Statistical modeling. Estimation. Linear Regression. Method of Least Squares (OLS).

### INTRODUÇÃO

No final dos anos de 1970, as áreas do cerrado maranhense tornaram-se palco da introdução direta de capital, no que diz respeito à agricultura (MOTA, 2017). A chegada dos agricultores do centro-sul promoveu uma transformação sócio espacial na cidade de Balsas-MA, através do surgimento e propagação da agricultura moderna, o que acarretou no aumento da população urbana (FERREIRA, 2011). Mota (2017),

<sup>1</sup>Graduada em Ciência e Tecnologia pela Universidade Federal do Maranhão (2018) e atualmente é graduanda Engenharia Civil na Universidade Federal do Maranhão. E-mail: nadynegomes9@hotmail.com

<sup>2</sup>Graduada em Interdisciplinar em Ciência e Tecnologia pela Universidade Federal do Maranhão (2018) E-mail: deborahh.pda@gmail.com

<sup>3</sup> Mestre em Engenharia de Materiais pelo Instituto Federal do Piauí- IFPI. Atualmente Professor da Universidade Federal do Maranhão – UFMA. E-mail: joao.valerio@ufma.br.

também nos diz que a modernização da agricultura no cerrado, além de modificar as relações no processo produtivo, contribuiu também para o processo de urbanização das cidades.

De acordo com Mota e Pessoa (2012), fatores como os baixos preços da terra, topografia plana da região, mão de obra barata, incentivos fiscais, esgotamento de terra do sul do país, proporcionaram essa migração do centro-sul para região de Balsas. Mota (2017), afirma que os gaúchos chegaram em Balsas a partir de 1974, implantando projetos agropecuários com ênfase na agricultura.

As tecnologias de previsão de safra, com as quais contam-se hoje, geram muitos benefícios. Tais como, um planejamento estratégico de produção com maior eficiência juntamente com uma melhor avaliação de custos, promovendo maior estabilidade e equilíbrio do setor agrícola da região.

De acordo com Rizzi e Rudorff (2005), o levantamento de estatísticas sobre as safras agrícolas é um requisito fundamental para que sejam traçadas medidas que favorecem todo o processo produtivo, evitando especulações e favorecendo preços justos. A utilização de modelos estatísticos para a elaboração de funções permite uma estimativa que pode ser usada para beneficiar várias áreas da economia. Nesta linha, ressaltam-se estudos de modelos de previsão de produtividade. (MIRANDA et al., 2011)

Sell (2005), afirma que a regressão linear, como uma das técnicas de métodos quantitativos, é usada para melhorar a capacidade de analisar o comportamento dos custos e o aumento da objetividade e confiabilidade das informações contábeis, através da análise individual das variáveis que possam estar relacionadas com a composição do custo.

Este artigo é resultado de estudos realizados para identificar e aplicar um modelo estatístico que estime a produção de soja na região de Balsas-MA em datas futuras, utili-

zando regressão linear simples por meio do Método dos Mínimos Quadrados (MMQ). De tal forma, verificando se o comportamento da produtividade em função do tempo estimada pelo modelo acompanhará a variação das produtividades observadas, se possibilitará estabelecer, conhecer e administrar os custos de produção de soja na região e se o modelo estatístico poderá ser aplicado na região de Balsas-MA, tendo em vista suas previsões mais rápidas em períodos antecedentes à colheita.

Assim, esta pesquisa justifica-se pela importância de estimar produções futuras de soja na região de Balsas-MA, uma vez que, os resultados desta podem representar um significativo interesse para a população no que se refere ao cultivo da soja na região, além disso, poderá facilitar no planejamento estratégico dos agricultores, para assim, alcançar a máxima eficiência e o mínimo de custos.

## 1. METODOLOGIA

Primeiramente foi feito um levantamento histórico da produção de soja da região de Balsas-MA através dos dados do IBGE. Tais dados dispostos na Tabela 1 foram utilizados para a obtenção do modelo estatístico por meio do MMQ e manipulados por meio do Excel.

Tabela 1 - Produção de soja por ano

Ano	Produção (t)
2001	162714
2002	186286
2003	216053
2004	262980
2005	289655
2006	246468
2007	324300
2008	331515
2009	319248
2010	376524
2011	400452
2012	417325
2013	386196
2014	457760
2015	501668

Fonte: IBGE (2015)

Depois de feito o levantamento, foi

medido o grau de relação entre o ano e a quantidade de produção através do coeficiente de correlação de Pearson conforme equação 1.

$$r = \frac{(n\sum xy - \sum x \sum y) / (\sqrt{(n(\sum x^2) - (\sum x)^2)} \cdot \sqrt{(n(\sum y^2) - (\sum y)^2)})}{(1)}$$

Em que,

r – coeficiente de correlação de Pearson;

n – quantidade de variáveis;

x – variável independente;

y – variável dependente;

Aplicando-se a regressão por mínimos quadrados (equações 2 e 3), é possível determinar os parâmetros estimados de modo que os desvios entre os valores examinados e estimados sejam mínimos, estabelecendo-se assim a relação linear entre o ano e a produtividade.

$$a = \frac{(\sum xy - n\bar{x}\bar{y}) / (\sum x^2 - n(\bar{x})^2)}{(2)}$$

$$b = \bar{y} - a\bar{x} \quad (3)$$

Em que,

a – coeficiente angular;

b – coeficiente linear;

$\bar{x}$  – média aritmética de x;

$\bar{y}$  – média aritmética de y.

Substituindo-se os valores encontrados para os coeficientes na equação da regressão linear (equação 4), obtém-se a função.

$$y = ax + b \quad (4)$$

Em seguida, foi calculado o coeficiente de explicação [ $r^2$ ] a fim de determinar a qualidade do ajuste do modelo, sendo possível quantificar a variabilidade nos dados que é explicada pelo modelo de regressão e a variabilidade que não é explicada.

De posse dos valores estimados pela função encontrada, analisaram-se os erros

para cada ano conforme equação 5.

$$\varepsilon = \text{valor real} - \text{valor estimado} \quad (5)$$

## 2. RESULTADOS E DISCUSSÃO

O levantamento histórico da produção de soja foi feito de 1990 até 2015, inicialmente realizou-se todos os cálculos utilizando 26 variáveis conforme a Tabela 2.

Tabela 2 - Resultados da aplicação do modelo estatístico

Variáveis	Ano (x)	Produção (y)	y (t)	$\varepsilon$
1	1990	1607	-50054,26	51661,26
2	1991	2070	-29337,80	31407,80
3	1992	5406	-8621,34	14027,34
4	1993	18552	12095,12	6456,88
5	1994	32888	32811,58	76,42
6	1995	36794	53528,04	-16734,04
7	1996	40698	74244,50	-33546,50
8	1997	65877	94960,96	-29083,96
9	1998	68003	115677,42	-47674,42
10	1999	128259	136393,89	-8134,89
11	2000	152141	157110,35	-4969,35
12	2001	162714	177826,81	-15112,81
13	2002	186286	198543,27	-12257,27
14	2003	216053	219259,73	-3206,73
15	2004	262980	239976,19	23003,81
16	2005	289655	260692,65	28962,35
17	2006	246468	281409,11	-34941,11
18	2007	324300	302125,58	22174,42
19	2008	331515	322842,04	8672,96
20	2009	319248	343558,50	-24310,50
21	2010	376524	364274,96	12249,04
22	2011	400452	384991,42	15460,58
23	2012	417325	405707,88	11617,12
24	2013	386196	426424,34	-40228,34
25	2014	457760	447140,80	10619,20
26	2015	501668	467857,26	33810,74

Fonte: Autores

O crescimento da produção de soja nos primeiros quatro anos teve alta variação, logo utilizou-se os últimos 15 anos para minimização de erros, o que em nada reduz a eficiência do método em vista que erros, neste processo, são aleatórios e devem possuir valor esperado tendendo a zero [ $E(\varepsilon) = 0$ ] e variância [ $\sigma^2$ ] mínima para normalmente serem distribuídos. Com tal quantidade de variáveis e com os valores iniciais tão distantes dos valores finais, o erro só começou a diminuir de modo significativo a partir do décimo primeiro ano, onde as variáveis começaram a ser mais próximas.

Devido ao modelo inicial apresentar

um erro muito alto nos primeiros anos, utilizou-se uma amostra de 15 dados para obter uma melhor homogeneidade com expectativa de alcançar um menor desvio padrão, ou seja, uma menor dispersão entre as variáveis.

Os resultados da manipulação dos dados para a obtenção do coeficiente de correlação de Pearson, do coeficiente de explicação e dos coeficientes linear e angular através do MMQ para as 15 variáveis são exibidos na Tabela 3.

**Tabela 3 - Cálculos para a obtenção da função**

Variáveis	Ano (x)	Produção (y)	xy	x <sup>2</sup>	y <sup>2</sup>
1	2001	162714	325590714	4004001	26475845796
2	2002	186286	372944572	4008004	34702473796
3	2003	216053	432754159	4012009	46678898809
4	2004	262980	527011920	4016016	69138480400
5	2005	289655	580758275	4020025	83900019025
6	2006	246468	494414808	4024036	60746475024
7	2007	324300	650870100	4028049	105170490000
8	2008	331515	665682120	4032064	109902195225
9	2009	319248	641369232	4036081	101919285504
10	2010	376524	756813240	4040100	141770322576
11	2011	400452	805308972	4044121	160361804304
12	2012	417325	839657900	4048144	174160155625
13	2013	386196	777412548	4052169	149147350416
14	2014	457760	921928640	4056196	209544217600
15	2015	501668	1010861020	4060225	251670782224
Σ	30120	4879144	9803378220	60481240	1725308796324

Fonte: Autores

Onde, substituindo os valores nas equações 1, 2 e 3, temos:

$$r = \frac{(15 \cdot 9803378220 - 30120 \cdot 4879144)}{(\sqrt{(15 \cdot (60481240) - (30120)^2)}) \cdot \sqrt{(15 \cdot (1725308796324) - (4879144)^2)}} \quad (6)$$

$$a = \frac{(9803378220 - 15 \cdot 2008 \cdot 325276,27)}{(60481240 - 15 \cdot (2008)^2)} \quad (7)$$

$$b = 325276,26 - 2008a \quad (8)$$

Obteve-se o coeficiente de Pearson de 0,97, comprovando a correlação muito forte entre as variáveis e um coeficiente de explicação de 0,94, indicando que 94% da produção de soja pode ser explicada pelo modelo aplicado. No entanto, 6% da variação da produção de soja é explicada por fatores diferentes da variação do tempo, tais

como: oscilações pluviométricas, déficit hídrico, variações de temperatura e manuseio do controle de invasores.

Aplicando os valores encontrados para os coeficientes angular e linear, 21632,39 e -43112554,25, respectivamente, foi possível determinar a relação entre as variáveis através da equação 9, que melhor se ajustou aos dados, minimizando a soma dos quadrados dos desvios entre os valores tabelados e os valores ajustados pela regressão linear. Sendo também possível determinar o erro.

$$y = 21632,39x - 43112554,25 \quad (9)$$

Na figura 1 está o gráfico de dispersão da produção de soja por ano onde é possível ver a linha de tendência linear que descreve a equação 9 descrita acima, confirmando o melhor ajuste.

**Figura 1 - Gráfico da função regressora estimada**



Fonte: Autores

Aplicando-se os valores de x na função, obtêm-se os valores estimados podendo ser comparados com os valores reais encontrando o erro para cada estimativa, conforme a tabela 4 abaixo.

Tabela 4 - Resultados obtidos pela aplicação do modelo estatístico

y (t)	ε
173849,57	-11135,57
195481,95	-9195,95
217114,34	-1061,34
238746,72	24233,28
260379,11	29275,89
282011,50	-35543,50
303643,88	20656,12
325276,27	6238,73
346908,65	-27660,65
368541,04	7982,96
390173,42	10278,58
411805,81	5519,19
433438,20	-47242,20
455070,58	2689,42
476702,97	24965,03

Fonte: Autores

Logo, utilizando o modelo encontrado pode-se estimar a produção de soja para anos futuros, abaixo na tabela 5 temos a estimativa de 2018 a 2032.

Tabela 5 – Estimativa futura da produção de soja na região de Balsas-MA

Ano	Produção (t)
2018	541609
2019	563241
2020	584874
2021	606506
2022	628138
2023	649771
2024	671403
2025	693036
2026	714668
2027	736300
2028	757933
2029	779565
2030	801197
2031	822830
2032	844462

Fonte: Autores

## CONSIDERAÇÕES FINAIS

Neste estudo, verificou-se a relação entre o ano e a produção de soja na região de Balsas-MA. A equação encontrada, utilizando regressão linear simples, por meio do MMQ, permite prever futuras produções, se utilizado as mesmas ou melhores tecnologias de produção e cultivo. Os resultados confirmam a validade da equação, tendo em vista que os valores estimados se aproximam dos reais. Dessa forma, nota-se que os valores possuem uma boa aproximação, contribuindo para que a equação seja capaz

de estimar a produção de soja. Com base na equação encontrada, estima-se uma produção de aproximadamente 541609 toneladas de soja para este ano de 2018, quase 40 mil toneladas a mais que em 2015. Sendo assim, esse modelo estatístico é um recurso que pode possibilitar um melhor planejamento agrícola a fim de reduzir custos e facilitar o desenvolvimento das atividades dos agricultores.

## REFERÊNCIAS

DEVORE, Jay L. **Probabilidade e estatística: para engenharia e ciências**. São Paulo: Cengage Learning, 2006.

FERREIRA, Maria da Glória Rocha. **Mudanças no Urbano de Balsas (Ma) Decorrentes da Agricultura Moderna**. Costa Rica: Revista Geográfica de América Central Número Especial EGAL, 2011.

HELENE, Otaviano. **Métodos dos Mínimos Quadrados**. Editora Livraria da Física, 2006.

MIQUELLUTI, Daniel Lima. **Métodos alternativos de previsão de safras agrícolas**. 86 f. Dissertação (Mestrado). Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, 2015.

MIRANDA, José M.; REINATO, Rosicler A. O.; SILVA, Adriano B. da. **Modelo matemático para previsão da produtividade do cafeeiro**. Campina Grande: Revista Brasileira de Engenharia Agrícola e Ambiental, 2011.

MOTA, Francisco Lima. **O Rural eo Urbano no Cerrado Sul-Maranhense: Balsas enquanto cenário de reprodução das transformações socioespaciais no pós 1980**. Disponível em: <http://www.periodicoeletronicos.ufma.br/index.php/interespaco/article/view/793>. Acesso em: 16 de março de 2017.

MOTA, Francisco Lima; PESSÔA, Vera Lúcia Salazar. **O Agronegócio como (re) produtor e um novo território**: Balsas no Contexto do Agronegócio da soja. Disponível em: <http://www.uff.br/vsinga/trabalhos/Trabalhos%20Completos/FRANCISCO%20LIMA%20M%20OTA.pdf>. Acesso em: 16 de março de 2017.

PALUDZYSZYN FILHO, E. **A cultura da soja no sul do Maranhão**. Balsas: EMBRAPA-CNP- SO, 1995. 34 p. (EMBRAPA-CNP- SO. Documentos, nº 84).

RIZZI, R. **Geotecnologias em um sistema de estimativa da produção de soja**: estudo de caso no Rio Grande do Sul / R. Rizzi. São José dos Campos: INPE, 2004.

RIZZI, Rodrigo; RUDORFF, Bernardo Friedrich Theodor. **Estimativa da produtividade de soja por meio de um modelo agrometeorológico-espectral**: estudo de caso no Rio Grande do Sul. Goiânia: Anais XII Simpósio Brasileiro de Sensoriamento Remoto, 2005.

SELL, Isair. **Utilização da regressão linear como ferramenta de decisão na gestão de custos**. 2005. IX Congresso Internacional de Custos - Florianópolis, SC, Brasil - 28 a 30 de novembro de 2005. Disponível em: [http://www.ime.usp.br/~salles/fatec/estatistica/trabalho/Utilização da regressão linear como ferramenta de decisão na gestão de custos.pdf](http://www.ime.usp.br/~salles/fatec/estatistica/trabalho/Utilização%20da%20regressão%20linear%20como%20ferramenta%20de%20decisão%20na%20gestão%20de%20custos.pdf). Acesso em: 16 de março de 2017.

Sistema IBGE de Recuperação Automática - SIDRAIBGE. **Produção Agrícola Municipal**. Disponível em: <https://sidra.ibge.gov.br/tabela/1612#resultado>. Acesso em: 16 de março de 2017.