



Estimativa do preço de mandioca na região de Assis-SP através de dados de insumos e fatores climáticos com aplicação de modelos lineares generalizados

Lucas Valle Mielke¹

ICMC - USP

Paulino Ribeiro Villas Boas²

Embrapa Instrumentação

ICMC - USP

1 Introdução

A mandioca é um dos principais alimentos de aproximadamente 700 milhões de pessoas, principalmente em países em desenvolvimento. No Brasil, ela é cultivada em todas as regiões e possui importante papel de geração renda da agricultura familiar. Como em todo cultivo agrícola, o custo de produção da mandioca é dependente de insumos, como área plantada, fertilizantes, defensivos, diesel, entre outros, e é afetada por fatores climáticos, tais como precipitação, insolação e temperatura. A oscilação no preço dos insumos e variações climáticas podem, portanto, impactar a produção, a oferta e, consequentemente, o preço da mandioca, causando problemas econômicos e sociais importantes, como redução do poder de compra dos consumidores ou redução da renda no campo. Desta forma, possibilitar melhor previsão da variação de preço da mandioca pode trazer benefícios sociais e econômicos ao permitir melhor planejamento dos envolvidos.

A tentativa de predição do preço dessa cultura poderia ser feita através de diversos modelos matemáticos de regressão disponíveis. Entretanto, o preço de produtos agrícolas nem sempre possuem distribuição normal [5] e, portanto, não pode ser bem explicado por modelos lineares [3]. Esse problema fez com que pesquisadores usassem algum tipo de transformação para alcançar a normalidade nesse tipo de situação [4]. Este procedimento, porém, muda a forma de interpretar os resultados, já que as análises estatísticas valem apenas para os dados transformados. Assim, os modelos lineares generalizados (MLGs) surgem como alternativa para elaborar modelos matemáticos de regressão quando a suposição de normalidade não é plausível sem a necessidade de

¹lucas.mielke@usp.br

²paulino.villas-boas@embrapa.br

transformação de variáveis, desde que a variável resposta pertença à família exponencial. A modelagem da média também não é feita de forma direta, sendo necessário escolher uma função de ligação que é capaz de relacionar as co-variáveis com a média da variável resposta [2,4].

O objetivo desse estudo foi, portanto, modelar o preço da mandioca em função de variáveis de insumos e de fatores climáticos com uso de MLGs.

2 Materiais e Métodos

Base de dados

Para elaboração do modelo proposto, foram coletados preços de mandioca, para uso como variável preditiva, e dados de insumos agrícolas, de área plantada e de fatores climáticos, como variáveis explicativas. A variável preditiva foi obtida a partir de pesquisa de preço à vista da raiz de mandioca para a região de Assis (SP) na base de dados do Centro de Estudos Avançados em Economia Aplicada (CEPEA).

As observações das variáveis relativas aos insumos agrícolas foram obtidas na base de dados de custos da Companhia Nacional de Abastecimento (CONAB) [1], onde se verificou que os maiores custos, excluídos mão-de-obra, correspondem aos agrotóxicos, fertilizantes e operações com máquinas. Os custos médios de agrotóxicos e fertilizantes foram coletados das estatísticas de importação do Ministério do Desenvolvimento, Indústria, Comércio Exterior e Serviço (MDIC). Para representar os custos de operações com máquinas agrícolas, foi utilizado a pesquisa de custos de diesel da Agência Nacional de Petróleo (ANP) da região de Assis (SP).

Coletou-se a área destinada ao cultivo de mandioca na região de Assis utilizando a Pesquisa Agrícola Municipal (PAM) realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Dentre os fatores climáticos, foram obtidos os dados de temperatura e de precipitação de estações meteorológicas localizadas nas cidades de Rancharia (SP), Ourinhos (SP), Lins (SP) e Londrina (PR), coletados pelo Instituto Nacional de Meteorologia (INMET). Também foi coletado a cotação do dólar comercial do Banco Central (BACEN) para conversão dos dados de Fertilizantes e Agrotóxicos de dólar para reais. Todos os dados foram coletados para o período entre dezembro de 2004 e janeiro de 2021.

Ajuste e análise dos dados

Os dados de Fertilizantes e Agrotóxicos foram convertidos de Dólar para Reais utilizando o dólar médio do mês da importação. Os dados temperatura e chuva da região de Assis foram obtidos através da interpolação, pelo inverso da distância, dos dados das estações meteorológicas coletadas. Foram calculadas variáveis extras a partir das variáveis dependentes de Temperatura, Diesel, Agrotóxicos, Fertilizantes e chuva para representar os valores médios (ou totais no caso da chuva) dos últimos 8 meses antes do preço de venda, que é o tempo médio de cultivo da mandioca. As análises foram feitas no software estatístico R. Todos os dados tratados e combinados para análise estatística encontram-se descritos na Tabela 1.

Distribuição Não-normal e Modelos Lineares Generalizados (MLGs)

Neste trabalho, usamos o teste de Shapiro-Wilk para avaliar se a distribuição dos preços da mandioca é normal. Neste teste, a hipótese nula é que a população possui distribuição normal,

Id	Descrição	Unidade	Sazonalidade	Fonte	Ajustes
area	Área agrícola	Hectares	Anual	IBGE	
diesel	Custo Diesel na revenda	R\$/L	Semanal	ANP	
temp_assis	Temperatura	°C	Diária	INMET	Interpolação pelo inverso da distância
chuva_assis	Pluviometria	mm	Diária	INMET	Interpolação pelo inverso da distância
adubos	Custo médio dos adubos	R\$/kg	Mensal	MDIC	Conversão para reais usando dólar comercial médio do mês
agrotóxicos	Custo médio dos agrotóxicos	R\$/kg	Mensal	MDIC	Conversão para reais usando dólar comercial médio do mês
diesel_8meses	Custo médio de diesel para o período de 8 meses	R\$/L	Semanal	ANP	Média móvel dos últimos 8 meses
temp_assis_8meses	Custo médio de temp_assis para o período de 8 meses	°C	Diária	INMET	Média móvel dos últimos 8 meses
chuva_assis_8meses	chuva_assis acumulada nos últimos 8 meses	mm	Diária	INMET	Total acumulado dos últimos 8 meses
adubos_8meses	Custo médio de adubos nos últimos 8 meses	R\$/kg	Mensal	MDIC	Média móvel dos últimos 8 meses
agrotóxicos_8meses	Custo médio de agrotóxicos nos últimos 8 meses	R\$/kg	Mensal	MDIC	Média móvel dos últimos 8 meses

Tabela 1: Descrição das variáveis utilizadas.

e a alternativa, o contrário. Portanto, um valor- $p < 0,05$ indica que os dados não seguem uma distribuição normal.

A situação do estudo pode ser expressa na forma de equação onde temos uma variável Y de interesse que pode ser explicada por vetores X , chamados de variáveis explicativas ou independentes da forma que ocorre com um modelo linear clássico. Entretanto, os MLGs são usados quando as observações não possuem distribuição normal, mas alguma outra da família exponencial, como a normal-inversa, a gama, a binomial, a Poisson, entre outras. Além disso, os MLGs requerem uma função de ligação, como identidade, inversa, logarítmica, entre outras, que relaciona a média da variável resposta a uma combinação linear das variáveis explicativas [2].

Para a seleção de MLG, em geral é utilizado o critério de informação de Akaike (AIC) que se baseia na função de verossimilhança e é amplamente utilizado para selecionar o melhor entre os modelos paramétricos alternativos. Essa seleção é feita considerando o AIC mínimo como melhor ajuste e tendo como objetivo a minimização desse valor ao testar modelos e variáveis [2].

3 Resultados e discussão

No teste de normalidade de Shapiro-Wilk obteve-se um valor- $p = 1,23 \times 10^{-14}$ que é menor que 0,05. Portanto, concluímos que a variável preditiva não possui distribuição normal a um nível de significância de 5%.

Pelo critério AIC verificou-se que dentre as diversas combinações de distribuições, funções de ligação e seleção de variáveis testadas dos MLGs, a distribuição ‘gama’ com função de ligação ‘inversa’ e variáveis ‘temp-assis-8meses’, ‘chuva-assis-8meses’, ‘diesel-8meses’, ‘adubos-8meses’ e ‘agrotóxicos-8meses’ melhor descreveu a variável dependente ‘preço’ ao ter o menor AIC entre todos os testes ($AIC = 3840,8$) e 27% dos pontos fora no gráfico de envelope da distribuição dos dados versus a teórica (Figura 1).

Pelos resultados, verifica-se que o MLG descreveu razoavelmente a variável preditiva em função das variáveis explicativas, o que indica que o modelo desenvolvido é promissor. Entretanto, notamos que as variáveis explicativas não captaram toda a variabilidade da variável preditiva, necessitando, portanto, incluir outras variáveis explicativas no modelo. Faltou ainda realizar uma análise de resíduos para encontrar possíveis observações aberrantes e obter um modelo com menor AIC e com menos pontos fora do gráfico de envelope. Pretendemos também realizar uma validação cruzada para avaliar a capacidade preditiva do modelo gerado e comparar os resultados com os de outros modelos de aprendizado de máquina, tais como, máquina de suporte de vetores e redes neurais.

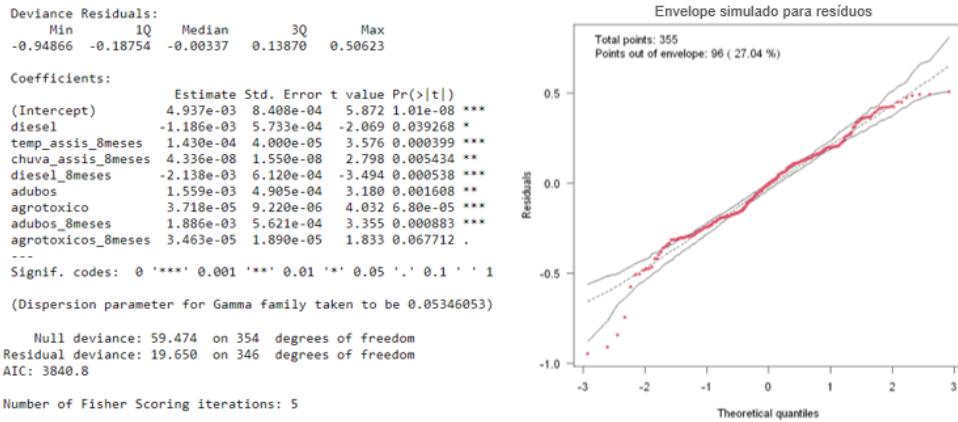


Figura 1: Resultados do MLG com menor AIC e gráfico de envelope

4 Conclusões

Neste trabalho, concluímos que os preços de mandioca foram melhores modelados pelos MLGs, porque as observações não possuíram distribuição normal a um nível de significância de 5% pelo teste de Shapiro-Wilk. Dentre as distribuições exponenciais avaliadas, a ‘gama’ com função de ligação ‘inversa’ foi a que melhor modelou os dados deste trabalho, considerando a variáveis médias no período de cultivo da mandioca, exceto a de precipitação que foi a acumulada. Entretanto, é necessário explorar outras variáveis e realizar análise de resíduos a fim de obter MLG com melhor capacidade de descrever a variável preditiva ‘preço’.

Referências

- [1] Companhia Nacional de Abastecimento. *Série de Custos de Mandioca*. Disponível em: <https://www.conab.gov.br/info-agro/custos-de-producao/planilhas-de-custo-de-producao/item/1979-serie-historica-custos-mandioca-2003-2017>. 2021.
- [2] A. F. Hess et al. Aplicação dos modelos lineares generalizados para estimativa do crescimento em altura *Brazilian Journal of Forestry Research*. Disponível em <https://www.researchgate.net/publication/290472250>. 2015.
- [3] E. M. Lima *Modelos Lineares Generalizados*. Universidade Federal de Minas Gerais. Disponível em: <http://www.est.ufmg.br/~enricoc/pdf/categoricos/mlg.pdf>.
- [4] G. A. Paula. *Livro do Instituto de Matemática e Estatística Universidade de São Paulo*. Disponível em: <https://www.ime.usp.br/~giapaula/texto2013.pdf>. 2013.
- [5] A. C. Silva. Previsão do preço da commodity café arábica *Revista Espacios Vol 39* disponível em: <https://www.revistaespacios.com/a18v39n04/a18v39n04p18.pdf>. 2017.