

MÉTODOS ESTATÍSTICOS DE REGRESSÃO PARA ESTIMAR O CUSTO DA COLHEITA MECANIZADA DE CANA-DE-AÇÚCAR

EDUARDO S. MAEKAWA¹, MARCOS MILAN²

¹ Matemático, Mestrando em Engenharia de Sistemas Agrícolas, Depto. Engenharia de Biossistemas, ESALQ/USP, Piracicaba-SP, Fone: (0XX19) 3306.8688, eduardo.maekawa@usp.br.

² Engenheiro Agrônomo, Prof. Doutor, Depto. Engenharia de Biossistemas, ESALQ/USP, Piracicaba-SP.

Apresentado no
XLIV Congresso Brasileiro de Engenharia Agrícola - CONBEA 2015
13 a 17 de setembro de 2015- São Pedro – SP, Brasil

RESUMO: A colheita mecanizada de cana-de-açúcar é uma das mais significativas e onerosas operações do seu processo de produção e por isso é importante o entendimento das relações que envolvem o seu custo. Atualmente, as metodologias para estimar o custo da colheita partem do conceito de custo fixo e variável. No entanto, considerando a complexidade desse processo, faz-se necessário avaliar métodos capazes de relacionar os parâmetros operacionais com o custo final. Neste contexto, a modelagem estatística por meio da regressão permite tratar tais relações e prever tendências. O objetivo deste trabalho foi desenvolver um modelo empírico para o cálculo do custo da colheita mecanizada de cana-de-açúcar. Desenvolveu-se um modelo linear generalizado (MLG) com distribuição gama, utilizando indicadores operacionais e dados de custo de 20 usinas do setor sucroalcooleiro. Por meio do MLG, obteve-se uma aderência satisfatória quando comparado aos modelos nulo (média) e linear (supondo normalidade). Os indicadores que explicaram o custo foram: produtividade ($t\text{ maq}^{-1}$), hora-horímetro (h) e número de operadores por colhedora (nop).

PALAVRAS-CHAVE: colhedora de cana, custo operacional, modelos lineares generalizados.

STATISTICAL METHODS OF REGRESSION TO ESTIMATE THE COST OF MECHANICAL HARVESTING OF SUGARCANE

ABSTRACT: The mechanical harvesting of sugarcane is one of the most significant and costly operations of the production process, thus it is important to understand the relationships involving its cost. Currently, methods to estimate these costs rise from the concept of fixed and variable cost. However, considering the complexity of the harvesting process, it is necessary to evaluate techniques to relate the operating parameters with the final cost. In this context, statistical modeling by regression allows to treat such relationship and predict trends. The objective of this study was to develop an empirical model to calculate the cost of mechanical harvesting of sugarcane. A generalized linear model (GLM) with gamma distribution was developed using operational indicators and cost data from 20 plants in the sugarcane industry of crop 14/15. Through the GLM, satisfactory adhesion was obtained when compared to the null model (average) and linear (assuming normality). The indicators that explained the cost were: productivity ($t\text{ mach}^{-1}$), hourmeter (h) and number of operators per harvester (nop).

KEYWORDS: generalized linear models, operational cost, sugarcane harvester.

INTRODUÇÃO: A competitividade das empresas envolve a busca constante por eficiência operacional, principalmente em um cenário onde demanda e preço estão sujeitos a diversos fatores externos. Por isso, para que se torne viável economicamente, a colheita mecanizada de cana deve ser analisada visando a redução de custos (MINETTE, 2008), uma vez que a colheita representa cerca de 35% do custo total de produção. Quando se trata dos modelos para estimativa do custo de colheita mecanizada, diferentes abordagens são citadas. No entanto, todas tem em comum a mesma premissa para formação das linhas de custo que são determinadas em duas vertentes: os custos fixos e os custos variáveis (RODRIGUES & SAAB, 2007; GARCIA & SILVA, 2010). Dada à complexidade do processo de colheita, faz-se necessário avaliar métodos que relacionam parâmetros operacionais com o custo final. Uma ferramenta útil que permite analisar tais relações e prever tendências é a modelagem

estatística por meio da regressão (FRYTAK *et al.*, 2009; VELOPULOS *et al.*, 2013). O objetivo desta pesquisa é desenvolver modelo empírico para o cálculo do custo da colheita mecanizada de cana-de-açúcar, verificar quais variáveis que melhor o explicam e aferir estatisticamente a qualidade dos resultados obtidos.

MATERIAL E MÉTODOS: Para estimar o custo da colheita mecanizada será desenvolvido um modelo linear generalizado (MLG) (NELDER & WEDDERBURN,1972) utilizando informações extraídas de 20 usinas. Os dados de cada uma das usinas, referentes à safra 2014/2015, serão tratados, agrupados mensalmente e analisados por meio de software estatístico R. Para desenvolver o modelo serão utilizadas duas bases, uma com indicadores operacionais (variáveis explanatórias) e outra com informações de custo (variável resposta). Os custos de amortização foram desconsiderados da análise. A seleção das variáveis será através do método da verossimilhança (PAULA, 2011). O diagnóstico do modelo será realizado por meio de análises de resíduos e gráficos em três abordagens: a) análise gráfica do *half normal plot* com envelope simulado (ATKINSON, 1985), b) critério de Akaike, AIC (MCCULLAGH & NELDER, 1989) e c) métrica proposta por Austin (2003), MPEQ – média predita do erro quadrado. Para verificar a aderência do modelo MLG, ele será comparado com o nulo (média) e com o linear (distribuição normal).

RESULTADOS E DISCUSSÃO: Inicialmente, análises prévias foram realizadas para entender o comportamento do custo da colheita e verificar qual a melhor técnica estatística para ser aplicada. Na Figura 1 apresenta-se a distribuição da variável resposta(custo).

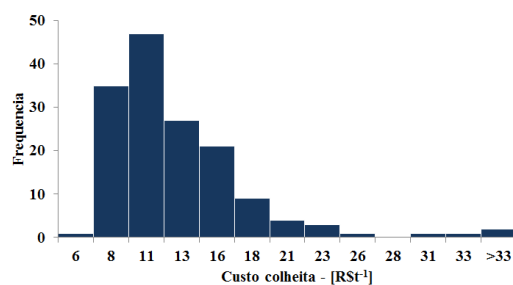


FIGURA 1. Histograma do custo da colheita mecanizada de cana-de-açúcar.

Percebe-se uma distribuição assimétrica do custo, o que pressupõe que as técnicas referentes ao modelo linear não terão boa aderência. O MLG com distribuição gama pode ser uma boa opção devido a sua característica de flexibilidade possibilitada pela variação dos parâmetros da sua função de densidade de probabilidade (HANDORF, 2013). A distribuição dos custos, por ordem de importância é apresentada na Figura 2.

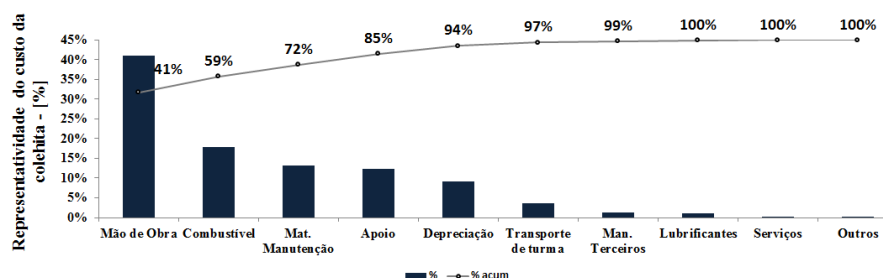


FIGURA 2. Ordem de importância das categorias do custo.

Na categoria de mão de obra, estão contidos os custos de ordenados dos operadores de colhedora. Em combustível, de abastecimento de óleo diesel. Em mat. manutenção, estão os custos de peças. Na categoria de apoio, custos de suporte como por exemplo: comboios e caminhão prancha. Em depreciação, a desvalorização da colhedora. Em transporte de turma, estão os custos de transporte dos operadores da usina até o campo. Em man. terceiro, os de manutenções externas. Na categoria de lubrificantes, os custos são referentes a lubrificantes. Em serviços, referem-se a serviços de terceiros. E em outros, os custos de categorias complementares. As categorias de mão de obra, combustível, mat.

manutenção, apoio e depreciação representam 94% do custo total. Sendo assim, os seguintes indicadores serão testados: número de operadores por colhedora (nop), produtividade (tonelada colhida por máquina), consumo (litros de diesel por hora), utilização (horas da colhedora realizando corte ou manobra) e horímetro (horas acumulada da colhedora). Uma matriz de correlação dos indicadores operacionais com o custo foi calculada a fim de analisar quais devem entrar no modelo. Quanto maior correlação com o custo, maior a chance de ser significativo (Tabela 1).

TABELA 1. Matriz de correlação das variáveis com o custo.

	Custo	Nop	Produtividade	Utilização	Horímetro	Consumo
Custo	1					
Nop	0,4586	1				
Produtividade	-0,5127	-0,1318	1			
Utilização	-0,3256	-0,0665	0,6517	1		
Horímetro	-0,0693	-0,1333	-0,0199	-0,0627	1	
Consumo	-0,1083	-0,0071	0,0757	-0,0846	0,1337	1

Nop – número de operadores por colhedora.

Número de operadores, produtividade e utilização são as que mais têm correlação com o custo. No entanto, produtividade e utilização tem correlação forte, o que implica escolher somente uma delas. Como a produtividade tem correlação maior com o custo, ela será priorizada. Horímetro e consumo por ter correlação baixa, serão testadas de forma categorizada. Na Figura 3 podem ser observados os gráficos de dispersão que auxiliaram para determinar os intervalos das variáveis.

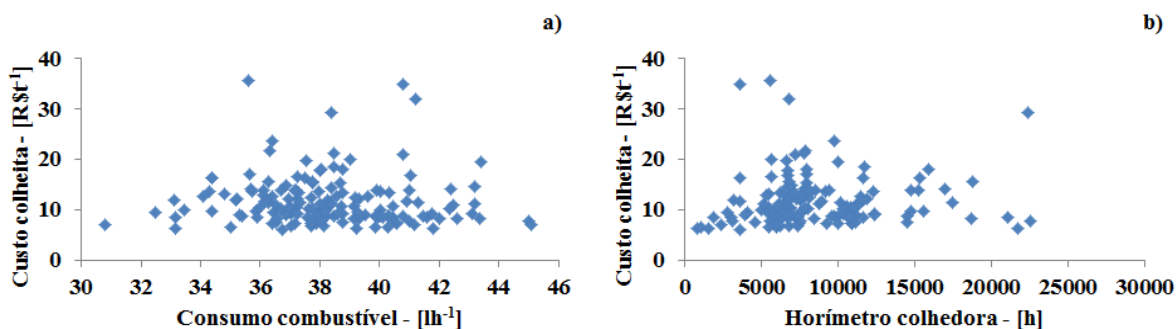


FIGURA 3. Dispersão das variáveis a) consumo e b) horímetro.

A partir das figuras 3a) e 3b) a categorização foi realizada de forma que a média do custo de cada intervalo ficasse ordenada. Tabela 2.

TABELA 2. Categorização das variáveis horímetro e consumo.

Horímetro	Consumo
até 3000h	até 35 lh ⁻¹
de 3000h-6000h	de 35-40 lh ⁻¹
de 6000h-12000h	acima de 40 lh ⁻¹
acima 12000h	

Com a escolha e categorização das variáveis, verificou-se significância das seguintes variáveis: produtividade, número de operadores e horímetro. Na comparação do modelo MLG com o nulo e com o linear feita com base na análise de que quanto mais pontos dentro dos limites do envelope melhor o ajuste, o MLG obteve somente um ponto fora (Figura 4). E a Tabela 3 mostra os diagnósticos de critério e métrica dos modelos. Quanto menor o valor, melhor o ajuste do modelo.

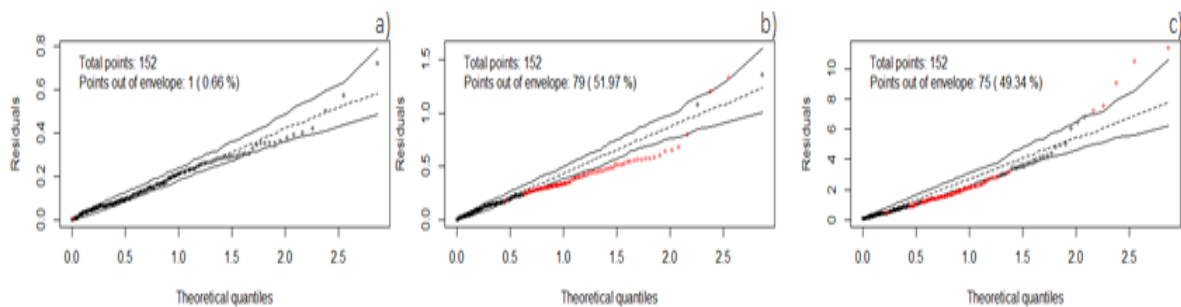


FIGURA 4. Diagnósticos dos modelos a) MLG, b) linear e c) nulo.

TABELA 3. Critério e métrica dos modelos MLG, linear e nulo.

Medidas	Modelo ajustado	Modelo Linear	Modelo Nulo
AIC	685.44	746.81	865.35
MPEQ	8.76	7.26	25.23

AIC – Critério de Akaike; MPEQ – Média predita do erro quadrado.

O modelo MLG tem diagnóstico gráfico e critério melhor que os demais. A métrica MPEQ do modelo linear tem o valor mais baixo, assim como foi percebido por Austin (2003). Isso ocorre, pois sob a função de erro quadrático, a média é um ótimo estimador da variável aleatória.

CONCLUSÕES: O Modelo MLG desenvolvido ajustou-se bem e quando comparado aos modelos linear e nulo através das análises de diagnósticos, obteve melhor resultado. Os indicadores que explicaram o custo foram: produtividade da colhedora, horímetro da colhedora e número de operadores por colhedora.

REFERÊNCIAS

- ATKINSON, A.C. (1985). Transformations and Regression. Oxford University Press, Oxford.
- AUSTIN, P.C; GHALI, W.A; TU, J.V. A comparison of several regression models for analysing cost of CABG surgery. Statist. Med. 2003; 22 p2799–2815.
- FRYTAK, J.R; HENK, H.J; CASTRO, C.M; HALPERN, R; NELSON, M. Estimation of economic costs associated with transfusion dependence in adults with MDS. CURRENT MEDICAL RESEARCH AND OPINION. VOL. 25, NO. 8, p1941–1951, 2009.
- GARCIA, R.F; SILVA, L.S. AVALIAÇÃO DO CORTE MANUAL E MECANIZADO DE CANA-DE-AÇÚCAR EM CAMPOS DOS GOYTACAZES, RJ. Engenharia na agricultura, viçosa - MG, V.18 N.3, p.234-240, MAIO / JUNHO 2010.
- HANDORF, E.A; BEKELMAN, J.E; HEITJAN, D.F; MITRA, N. EVALUATING COSTS WITH UNMEASURED CONFOUNDING: A SENSITIVITY ANALYSIS FOR THE TREATMENT EFFECT. The Annals of Applied Statistics, 2013, Vol. 7, No. 4, P2062–2080.
- MCCULLAGH, P.; NELDER, J. A. (1989). Generalized Linear Models. Chapman and Hall, London, second edition. 511p.
- MINETTE, L.J.; SILVA, E.N.; FREITAS, K.E.; SOUZA, A.P.; SILVA, E.P. Análise técnica e econômica da colheita florestal mecanizada em Niquelândia, Goiás. Revista Brasileira de Engenharia Agrícola e Ambiental, Campina Grande, v.12, n.6, p.659-665, 2008.
- NELDER, J. A; WEDDERBURN, R. W. M. (1972). Generalized linear models. Journal of the Royal Statistical Society, A, 135, 370–384.
- Paula, G. A. (2011). Modelos de Regressão com Apoio Computacional. IME/USP, São Paulo.
- RODRIGUES, E.B; SAAB, O.J.G.A. Avaliação técnico-econômica da colheita manual e mecanizada da cana-de-açúcar (*saccharum spp*) na região de Bandeirantes – Pr. Semina: Ciências Agrárias, Londrina, v. 28, n. 4, p. 581-588, out./dez. 2007.
- VELOPULOS, C.G; ENWEREM, N.Y; OBIRIEZE, A; HUI, X; HASHMI, Z.G; SCOTT, V.K; CORNWELL, E.E; SCHNEIDER, E.B; HAIDER, A.H. National cost of trauma care by payer status. Journal of surgical research 184 (2013) P.444-449.