Universidade de São Paulo
Escola Superior de Agricultura "Luiz de Queiroz"

Estimativa do custo da colheita mecanizada de cana-de-açúcar utilizando modelos de regressão

Eduardo Shigueiti Maekawa

Dissertação apresentada para obtenção do título de Mestre em Ciências. Área de concentração: Engenharia de Sistemas Agrícolas

Piracicaba
2016

Eduardo Shigueiti Maekawa
Licenciado em Matemática

Estimativa do custo da colheita mecanizada de cana-de-açúcar utilizando modelos de regressão

Orientador:
Prof. Dr. **MARCOS MILAN**

Dissertação apresentada para obtenção do título de Mestre em Ciências. Área de concentração: Engenharia de Sistemas Agrícolas

Piracicaba
2016

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Maekawa, Eduardo Shigueiti
Estimativa do custo da colheita mecanizada de cana-de-açúcar utilizando modelos de regressão / Eduardo Shigueiti Maekawa. - - Piracicaba, 2016.
69 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Colhedora de cana 2. Custo operacional 3. Modelos lineares generalizados 4. Modelos lineares generalizados mistos I. Título

CDD 633.61
M184e

DEDICATÓRIA

Dedico a minha esposa e minha filha.

AGRADECIMENTOS

Agradeço a minha esposa pelo companheirismo, paciência, força e orientação. Minha filha pelos sorrisos que davam forças durante o processo trabalho-estudo-família. Ao Maclovin pelos latidos e ações de carinho.

Agradeço aos meus pais pela educação e criação dada. As minhas três irmãs que mesmo longe sempre me deram suporte e motivação. Aos meus sobrinhos e sobrinhas pelas boas energias que toda criança proporciona.

Aos meus cunhados, cunhadas, sogro e sogra pelo suporte, amizade e convívio que mostraram que persistir e trabalhar duro compensam a difícil jornada entre o início e o fim.

Ao meu orientador Marcos Milan pela parceria e por ter acreditado no meu projeto. Pelas orientações, revisões, transmissão de conhecimento e boas conversas.

Agradeço a ESALQ por ter possibilitado a oportunidade de aprendizado e a realização do mestrado. Ao Departamento de Engenharia de Biossistemas e todos os professores que agregaram valor ao conhecimento.

Aos amigos pelos estudos em grupos, discussões, ajudas e colaboração. As secretárias e toda a infraestrutura da ESALQ pela facilitação dos serviços.

Agradeço ao meu gerente e a empresa que acreditam no desenvolvimento das pessoas entre a academia e mundo corporativo e pelas liberações para assistir as aulas.

EPÍGRAFE

"Um mar calmo nunca formou bons marinheiros."

SUMÁRIO

RESUMO

Estimativa do custo da colheita mecanizada de cana-de-açúcar utilizando modelos de regressão

A colheita mecanizada é uma das mais significativas e onerosas operações do processo de produção de cana-de-açúcar, tornando-se 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) e um modelo linear generalizado misto (MLGM) ambos com distribuição gama, utilizando indicadores operacionais e dados de custo de 20 usinas do setor sucroalcooleiro. Por meio do MLGM, obteve-se uma aderência satisfatória quando comparado aos modelos MLG, nulo (média) e linear (supondo normalidade). Os indicadores que explicaram o custo foram: produtividade (t maq^{-1}), consumo (l t^{-1}), horímetro (h) e número de operadores por colhedora (nop).

Palavras-chave: Colhedora de cana; Custo operacional; Modelos lineares generalizados; Modelos lineares generalizados mistos

ABSTRACT

Estimated cost of mechanized harvesting of sugarcane using regression models

The mechanized 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) and a generalized linear mixed model (GLMM) both with gamma distribution was developed using operational indicators and cost data from 20 plants in the sugarcane industry. Through the GLMM, satisfactory adhesion was obtained when compared to the GLM, null model (average) and linear (assuming normality). The indicators that explained the cost were: productivity (t $mach^{-1}$), consumption (l t^{-1}), hourmeter (h) and number of operators per harvester (nop).

Keywords: Generalized linear models; Generalized linear mixed models; Operational cost; Sugarcane harvester

LISTA DE FIGURAS

LISTA DE TABELAS

1 INTRODUÇÃO

A cana-de-açúcar transformou-se em uma das principais culturas da economia brasileira, tornando o país o maior produtor de açúcar do mundo. Além disso, a produção de etanol, menor apenas do que a dos Estados Unidos, vem se destacando no mercado externo como alternativa ao uso dos combustíveis fósseis.

Além do açúcar e o do etanol, outros subprodutos da cana estão ganhando a atenção do mercado. Entre eles, a energia cogerada a partir da queima do bagaço, os bioplásticos como garrafas de bebida e pacotes de alimentos, os biocombustíveis como o etanol de segunda geração e os resíduos da produção como a torta de filtro e a vinhaça, que são utilizados como fertilizantes.

Diante dessas possibilidades de produtos e visando equilibrar o desenvolvimento socioeconômico com a sustentabilidade, a cadeia produtiva da cana deve ser conduzida de forma eficiente, principalmente no processo da colheita uma vez que 35% do custo total da produção encontram-se nesta etapa.

Do ponto de vista agrícola, a transição da colheita manual para colheita mecanizada está sendo o maior desafio. As áreas que demandam essa transição sofrem mudanças significativas na gestão agronômica desde a variedade da cana até a aplicação de herbicidas e fertilizantes.

Por outro lado, a mecanização da colheita ajudou a aquecer o mercado, fazendo com que as usinas elevassem os investimentos na aquisição de colhedoras. A inovação tecnológica trouxe também notáveis avanços na diminuição da erosão, aumento do teor de matéria orgânica dos solos e redução do consumo de água.

O grande desafio para os próximos anos será o aumento significativo da demanda tanto do açúcar quanto do etanol, enfrentando temas como escassez de água, impactos ambientais e produtividade agrícola. Desse modo, o setor sucroenergético deverá ser expandido intensificando a produção, inovando práticas agronômicas e melhorando a gestão de processos.

Além disso, em um cenário onde demanda e preço estão sujeitos a diversos fatores externos, a competitividade das empresas envolve naturalmente a busca constante por redução de despesas. Por isso, é imprescindível a análise e entendimento dos custos e suas relações com o processo produtivo.

Diferentes abordagens são citadas no que se refere à estimativa do custo da colheita mecanizada. No entanto, para a formação desse custo, são determinadas

duas vertentes utilizando uma mesma premissa através do cálculo dos custos fixos e dos custos variáveis. Em geral, os custos fixos são calculados levando em consideração a mão de obra, a depreciação, os impostos e as taxas de alojamento, enquanto que os custos variáveis consideram o combustível, o lubrificante, o material e a manutenção.

Apesar das abordagens determinarem boas estimativas de custo, o cálculo torna-se muito pulverizado e encontram-se dificuldades para realizar testes de sensibilidade com mais de duas variáveis a fim de verificar o comportamento do custo. Dessa forma, com o intuito de contornar esses problemas e dada à complexidade do processo da colheita mecanizada, 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 as relações entre os parâmetros e prever tendências. Assim, o objetivo desta pesquisa visa estimar o custo da colheita mecanizada de cana-de-açúcar a partir de dados referentes a indicadores operacionais, extraídos dos computadores de bordo instalados em colhedoras.

2 REVISÃO BIBLIOGRÁFICA

A cana-de-açúcar é uma das principais culturas do mundo, cultivada em mais de 100 países e representa uma importante fonte de mão de obra no meio rural. Apesar desta pulverização, os cinco maiores países produtores já somam 75% da produção mundial. O Brasil é o maior produtor representando 38%, seguido de Índia com 17%, China com 12%, Tailândia com 5% e Paquistão com 3%. (FAO, 2016)

No Brasil, a produção de cana cresceu de forma acelerada após o estabelecimento do Proálcool, em novembro de 1975, dobrando sua produção na safra de 1986/1987. Na safra 1993/1994 também teve um crescimento considerável motivado pelo aumento das exportações de açúcar. Em 2003, com o lançamento dos veículos *flex*, a produção de cana-de-açúcar teve um crescimento acelerado para atender ao aumento da demanda de álcool hidratado (NOVACANA, 2016).

Nos últimos anos, o aumento da demanda interna por álcool hidratado e ampliação da exportação do açúcar, impulsionou a expansão de aproximadamente 100 novas usinas, principalmente no estado de São Paulo. De 2001 a 2014 a produção de açúcar teve um aumento de 87%, saindo de 18,9 milhões de toneladas para 35,6. Já para o etanol o aumento foi de 52% partindo de 11,5 milhões de m^3 para 28,9 (CONSELHO DE PRODUTORES DE CANA-DE-AÇÚCAR, AÇÚCAR E ETANOL DO ESTADO DE SÃO PAULO - CONSECANA, 2016).

Diante desse cenário, fica clara a necessidade das organizações estarem preparadas às mudanças, sabendo onde investir e reduzir custos para melhorar os resultados. De acordo com Chagas et al. (2016) é evidente como a melhora nas práticas de gestão agrícola na mecanização da colheita influenciam na diminuição dos custos além do benefício ao meio ambiente. Para Lanna e Reis (2012) o estudo dos custos através da viabilidade econômico-financeira torna-se relevante para acompanhar rentabilidade e estimar impactos na gestão do processo.

As próximas seções irão abordar três técnicas utilizadas para a estimativa de custo. A primeira com abordagens nas vertentes de custos fixos e variáveis. A segunda com a metodologia de regressão do modelo linear generalizado (MLG) e a terceira com a metodologia de regressão do modelo linear generalizado misto (MLGM).

2.1 Custos da colheita mecanizada

A análise e entendimento da colheita mecanizada são vitais para redução de custos, objetivando máximo aproveitamento das funções nas operações de forma contínua (MINETTE et al., 2008). De acordo com Higgins e Davies (2004), para redução de custos, o setor canavieiro está explorando oportunidades dentro da colheita, implementando boas práticas e removendo processos ineficientes.

Em qualquer processo de produção, devem-se evitar desperdícios e excessos de recursos, visando à lucratividade e garantindo a sobrevivência no mercado (ZANLUCA, 2009). Para Cunha e Rodrigues (2012) é preciso ter conhecimento de dados para planejamento e controle de custos, pois a ausência de informações pode resultar no fracasso da organização. A contabilidade de custos, além de fornecer elementos para avaliar estoques e acompanhar resultados, trouxe grande importância à contabilidade gerencial, auxilio no controle e tomada de decisões (NEVES; VICECONTI, 2010).

Neste contexto de acompanhamento e análise de custo, muitos autores abordaram o seu impacto na colheita mecanizada. Vieira (2003) avaliou de forma comparativa as relações do custo do corte de cana-de-açúcar manual e mecanizado com a produtividade de trabalho e a geração de empregos. As metodologias utilizadas para estimativa do custo foram as propostas por Mialhe (1974), Noronha (1987), Witney (1988) e Hoffmann et al. (1992), concluindo evidente vantagem para o sistema mecanizado com custo 54% menor.

Rodrigues e Saab (2007) avaliaram a viabilidade técnica e econômica da utilização das colhedoras automotrizes de cana-de-açúcar comparando o custo da colheita manual queimada com o custo da colheita mecanizada. Os autores utilizaram a metodologia de desempenho econômico proposta por Ripoli e Mialhe (1982), concluindo que a colheita mecanizada para área e condições avaliadas, é técnica e economicamente promissora.

Garcia e Silva (2010) compararam o desempenho operacional nos sistemas de colheita manual e mecanizado de cana-de-açúcar e sua relação com os custos, identificando-se maior capacidade no corte mecânico em relação ao manual e com um lucro 31% maior. Para isso, a estimativa dos custos utilizada foi através do método de Noronha (1987).

Dos Santos et al. (2015) avaliaram o impacto econômico causado pela perda de cana de acordo com a variação da velocidade da colhedora. Concluíram que a velocidade ideal fica entre 2,01 e 2,99 quilômetros por hora para compensar o ganho em custo e a perda da matéria-prima. Para análise do impacto econômico estimaram o custo da colheita mecanizada da cana através da metodologia proposta pela American Society of Agricultural and Biological Engineers – ASABE (2011).

Oliveira et al. (2007) avaliaram os custos operacionais da colheita mecanizada do café e verificaram que os fatores que mais influenciaram na composição dos custos foram gastos com depreciação, mão de obra e combustível. Além disso, concluíram que quanto maior a eficiência da colheita, menores são os custos operacionais. Para o cálculo dos custos, a metodologia empregada foi proposta por Tourino (2000) e Silva (2004).

Cunha et al. (2011) compararam o custo operacional e as perdas de produção entre a colheita mecanizada e semi-mecanizada da cultura de batata. Para estimar os custos operacionais, utilizaram a abordagem proposta por Centeno e Kaercher (2010), concluindo que o custo da colheita mecanizada em relação à semi-mecanizada teve redução de 49% e 4%, para custo operacional e perdas de produção respectivamente.

Burla et al. (2012) avaliaram técnica e economicamente a colhedora de árvores *harvester* em diferentes condições de terreno e produtividade florestal. A análise técnica consistiu em estudo de tempos e movimentos do ciclo operacional da máquina, enquanto que a análise econômica consistiu na determinação dos custos operacionais por meio da metodologia proposta por Machado e Malinovski (1988).

Apesar das diferentes abordagens, uma mesma premissa é adotada para formação do custo. Essa premissa em comum pode ser resumida na Figura 1.

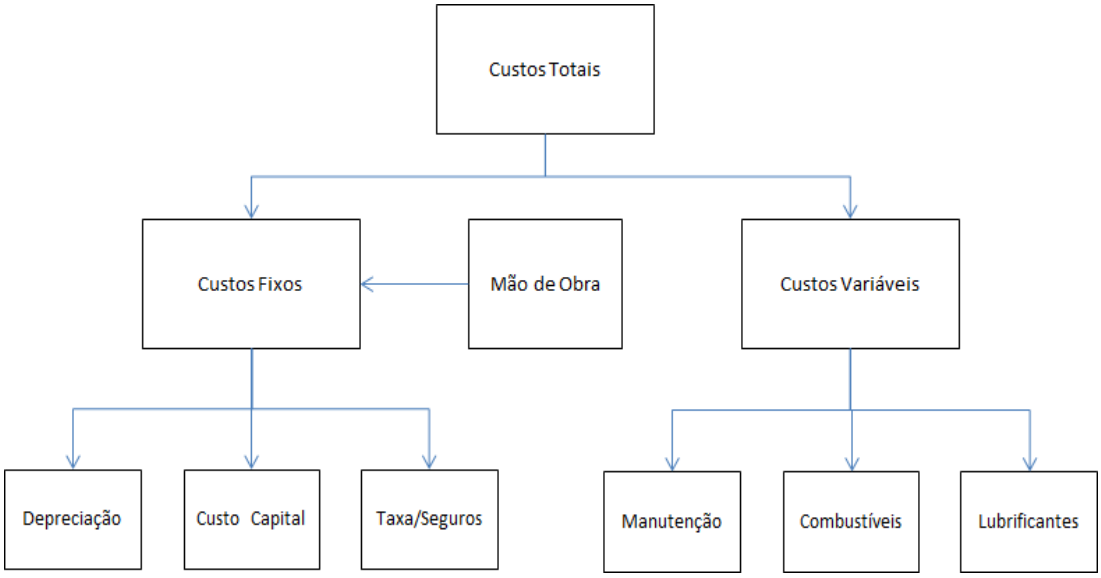

Figura 1 - Organograma para estimativa de custos da colheita mecanizada

2.2 Modelos lineares generalizados

Durante muito tempo, os modelos lineares foram utilizados para descrever a maioria dos fenômenos aleatórios, mesmo quando não apresentava uma resposta para a qual a suposição da normalidade fosse assumida. O que geralmente ocorria era a transformação da variável a fim de obter a normalidade procurada, visando trazer a constância da variância e a linearidade para o preditor. A mais conhecida era a de Box e Cox (1964)

$$z = \begin{cases} y^{\lambda}-1 \text{ , se } \lambda \neq 0 \\ \log y \text{ , se } \lambda = 0 \end{cases} \qquad (1)$$

No entanto, enquanto a transformação tem a vantagem de boas propriedades e inferências de modelos lineares, há uma dificuldade em interpretar os coeficientes do modelo que estão em escala diferente da métrica inicial. Uma saída é a transformação inversa da resposta, mas de acordo com Baldi (2013, apud MANNING, 1998), qualquer transformação da variável resposta implica em uma investigação profunda no erro total do modelo, pois poderá conduzir a estimadores viesados.

Nelder e Wedderburn (1972) propuseram os modelos lineares generalizados (MLG), que consiste em aumentar as opções para a distribuição da variável

resposta, pertencendo à família exponencial de distribuições, dando maior flexibilidade na relação média da variável resposta e o preditor linear η.

Cordeiro e Demétrio (2014) definem os MLG em 3 partes:

1) Variável resposta ou componente aleatório e sua distribuição (normal, gama, binomial, Poisson). (μ_i)
2) Variáveis explanatórias ou componente sistemático, que entram de forma linear no modelo. (η_i)
3) Ligação entre componente aleatório (1) e sistemático (2), através da função de ligação g(.)

$$g(\mu_i) = \eta_i \tag{2}$$

em que,

$\eta_i = X\beta$ – valor esperado da variável resposta;

$X=(X_1,\ldots, X_p)$ – vetor das variáveis explanatórias;

$\beta=(\beta_1,\ldots, \beta_n)^T$ – vetor dos parâmetros estimadores.

Tabela 1 - Função de ligação das distribuições de probabilidade mais utilizadas

	Função padrão	Outras funções
Normal	μ	-
Lognormal	μ	-
Gamma	log(μ)	1/μ
Exponencial	log(μ)	1/μ
Beta	log[μ/(1-μ)]	-
Binomial	log[μ/(1-μ)]	log[-log(1-μ/n)]
Poisson	log(μ)	-

Os modelos de regressão também tem como vantagem a identificação de fatores que mais influenciam a variável dependente. Além disso, a variável dependente é modelada na sua forma bruta sem precisar realizar qualquer transformação linear (PRENZLER et al., 2011).

Halpern et al. (2013) também mostraram que com a estimativa de custos através da regressão é possível realizar testes de sensibilidade. E quando estão devidamente especificados, obtém-se estimativas consistentes na definição dos parâmetros de regressão.

As análises de estimativa de custo através da regressão tornaram-se mais evidentes a partir do início dos anos 2000. Austin et al. (2003), utilizaram 7 modelos de regressão diferentes para estimar o custo de uma cirurgia através de variáveis como idade, sexo, diabete, acesso ao hospital via emergência. Os modelos testados foram: regressão linear, regressão linear com log-transformada, modelos lineares generalizados com distribuições poisson, binomial negativa e gama, regressão mediana e modelos proporcionais *hazards*, concluindo que qualquer um dos modelos pode ser utilizado para identificar os fatores associados com o aumento do custo. No entanto, para assertividade de valores altos, a melhor opção são os modelos lineares generalizados, enquanto que a regressão mediana traz melhores resultados para valores baixos.

Unützer et al. (2009) desenvolveram um MLG com distribuição gama, para estimar o custo hospitalar em pacientes depressivos. Através das inferências do modelo, observaram que o custo de quem tem depressão é duas vezes maior. Frytak et al. (2009) analisaram a relação de transfusão de sangue em pacientes com síndrome da mielodisplasia (SMD – doença que afeta o funcionamento dos ossos) e o seu impacto financeiro. Através de informações dos pacientes como dados demográficos, situação de saúde, utilização de procedimentos médicos, um MLG com distribuição gama foi desenvolvido para estimativa dos custos.

Rockhill et al. (2012) acompanharam custos médicos durante 3 anos de pacientes que passaram por um trauma cerebral (TC) e/ou doenças psiquiátricas (DP). O objetivo era verificar o comportamento dos custos com a presença de TC e DP. Por meio de um MLG com distribuição gama, concluíram que os pacientes com TC tinham custos 5,75 vezes maior em relação aos que não tinham e com a presença da DP os custos médicos dobravam.

Murakami et al. (2013) estudaram a relação de custos médicos com doenças cardiovasculares a fim de reduzir gastos a partir da criação de políticas públicas de prevenção de saúde. Os autores desenvolveram um MLG com distribuição gama para estimar o custo através de fatores de risco como hipertensão, colesterol, glicose no sangue e tabagismo.

Para estimar os custos até cinco anos após o tratamento de pacientes com hepatite C que atingiram a resposta virológica precoce (RVP) - queda de 100 vezes na carga viral após 12 semanas consecutivas de tratamento - Manos et al. (2013) utilizaram informações como idade, sexo e raça para desenvolver um modelo de

regressão gama. Em seguida, compararam os custos obtidos concluindo que os custos médicos foram significativamente mais baixos em pacientes que atingiram RVP.

Velopulos et al. (2013) verificaram o comportamento das despesas médicas sob uma perspectiva de tipo de pagamento do paciente (plano de saúde, pagamento particular) desenvolvendo MLG. A análise mostrou que os pacientes particulares tem custo muito maior que os usuários de plano de saúde.

Goren et al. (2014) avaliaram o impacto de incontinência urinária através de utilização de recursos médicos por meio de um MLG para estimar os seus custos associados. Já Orueta et al. (2014) desenvolveram um MLG com distribuição gama para estimar o custo médio de pacientes de acordo com a quantidade de patologias que cada um possuía.

Upatising et al. (2015) através do desenvolvimento de um MLG com distribuição gama, estimaram os custos médicos em idosos a fim de verificar se existia diferença significativa entre dois grupos: um grupo que era tratado com monitoramento médico e outro sem o monitoramento. Concluíram que a intervenção do monitoramento não tinha diferença significativa de custo.

Wakeam et al. (2015) desenvolveram um MLG com distribuição gama para estimar o custo de pacientes que removeram o pulmão a fim de comparar se determinados grupos de pessoas separados por idade teriam diferenças significativas. Já Soerensen et al. (2015) desenvolveram um MLG para estimar o custo da implementação de tratamento de câncer em um grupo de pessoas na Dinamarca, concluindo que não houve diferenças significativas em termos de custo com a presença ou ausência do tratamento.

Pela distribuição assimétrica encontrada em dados de custo, para Kwong et al. (2010) e Handorf et al. (2013) o uso da regressão gama com função de ligação logarítmica torna-se uma boa opção devido a flexibilidade possibilitada pela variação dos parâmetros de sua função de densidade de probabilidade. De acordo com Basu et al. (2004) o uso dessa regressão para estimar custos médicos traz melhores resultados do que outras técnicas estatísticas como a análise de sobrevivência. O uso do MLG para estimativa de custo é comumente utilizada na área médica. Em outras áreas como a agrícola é praticamente nulo.

2.3 Modelos lineares generalizados mistos

Os estudos de McCullagh e Nelder (1989) trouxeram atenção aos MLG para toda comunidade estatística com o início das pesquisas adicionando efeitos aleatórios, originando o desenvolvimento dos modelos lineares generalizados mistos (MLGM).

Os MLGM são definidos na mesma estrutura do MLG, sendo compostos por i) variável resposta, ii) variáveis explanatórias e iii) função de ligação. A diferença é que nas variáveis explanatórias entram efeitos aleatórios (GBUR et al., 2012).

Efeitos fixos são aqueles que permitem inferir características para toda a população. Por outro lado, efeitos aleatórios apesar de não possibilitar inferir características diretamente, auxiliam na explicação da dependência entre as respostas do mesmo indivíduo, e esta variabilidade reflete a heterogeneidade devido a fatores não mensurados ou não mensuráveis.

Assim, considera-se um modelo como MLGM quando possui ao menos uma variável de efeito fixo e uma variável de efeito aleatório, definindo-se da seguinte forma (JIANG, 2006):

$$g(\mu_i) = x_i'\beta + z_i' U \tag{3}$$

em que,

$g(\mu_i)$ – valor esperado da variável resposta;

x_i' – vetor das variáveis explanatórias fixas;

z_i' – vetor das variáveis explanatórias aleatórias;

β – vetor dos coeficientes das explanatórias fixas;

U – vetor dos coeficientes das explanatórias aleatórias.

Os principais elementos dos MLGM de acordo com Silva (2014) são: i) as respostas são independentes quando condicionados aos efeitos aleatórios, ii) cada resposta segue uma distribuição condicional pertencente a família exponencial e iii) os efeitos aleatórios seguem uma distribuição normal.

Utilizar o MLG quando a premissa de independência da variável resposta não é respeitada, leva a inferências estatísticas com distorção (viés). Já com o uso do MLGM, além de flexibilizar a suposição de independência possibilita a inferência de estatísticas sem viés. (FORTIN, 2013).

A premissa básica do MLGM é a suposição de heterogeneidade dos indivíduos na população, além de possibilitar maior precisão nos intervalos de confiança e maior utilidade para realizar inferências com características específicas ao invés de inferências na média geral da população (MITTAL et al., 2015).

Para Baldi et al. (2013) a presença do efeito aleatório induz uma correlação com a variável resposta. Os efeitos aleatórios são essenciais para ajustar a dependência da resposta, a dispersão da variância e assimetria da distribuição. De acordo com Fausto et al. (2008) o uso de efeitos aleatórios nos modelos de regressão é especialmente adequado para dados em que a variabilidade entre indivíduos é maior que a variabilidade dentro do indivíduo. Além disso, para Nunes et al. (2004) os MLGM têm se mostrado muito eficiente para ajustes de modelos com superdispersão (quando a variância observada é muito maior do que a esperada).

De acordo com Bautista (2014) a modelagem da parte aleatória se realiza com a inclusão de uma matriz de variâncias e covariâncias sendo que os principais motivos são: i) unidades experimentais colocadas em um mesmo grupo e ii) medidas repetidas sobre uma mesma unidade experimental. Em ambos os casos, ocorre uma correlação entre as observações e para Da Costa (2003) a análise dessa correlação faz com que o modelo tenha melhor ajuste captando melhor sua variabilidade.

Para Kurusu (2013) deve-se ter cuidado, pois resultados inválidos podem ser obtidos quando algum efeito aleatório é ignorado. E ainda, diferentes escolhas de funções de ligação resultam em diferentes aderências. De acordo com Prates et al. (2013) uma escolha errada pode arruinar a relação da variável resposta com os efeitos fixos e aleatórios.

Os modelos de regressão MLGM com distribuição gama, assim como os MLG, também têm sido estudados majoritariamente em análises de custos médicos proporcionando bons ajustes aos dados (ONG et al., 2013).

Yau et al. (2002) desenvolveram um modelo MLGM com distribuição gama para estimar o custo de acidentes de trabalho no hospital público Western da Austrália. Grieve et al. (2005) desenvolveram um MLGM com distribuição gama para estimar o custo de tempo de estadia bem como o custo total de pacientes que entraram no hospital com princípio de infarto. Além do MLGM também desenvolveram o MLG e o modelo linear, concluindo que o MLGM foi o que mais se adequou para a estimativa do custo.

Liu et al. (2010) estimaram o custo farmacêutico em pacientes da região central dos Estados Unidos durante o período de um ano. Nesse tipo de custo há ocorrência de muitos zeros em casos onde os pacientes não utilizaram medicamentos. Com isso, desenvolveram um modelo em duas partes. A primeira estimou a probabilidade de o custo ocorrer através de um MLG com distribuição binomial e a segunda com um MLGM gama para estimar os valores das despesas, concluindo que a inclusão de efeitos aleatórios auxilia na caracterização da variabilidade dentro de agrupamentos, trazendo boas inferências para definição de políticas médicas.

Baldi et al. (2013) desenvolveram um MLGM gama para estimar o custo médico em dois cenários: o primeiro determinou custos de pacientes com úlcera e o segundo o custo de diferentes estratégias de tratamento em pacientes com infarto do miocárdio. O MLGM foi comparado com o modelo linear misto, resultando em aderências muito melhores com o viés nunca excedendo 1%, enquanto que no linear misto chegava a passar de 10%.

Engblom et al. (2012) estimaram o custo logístico considerando-o como uma proporção do faturamento de 241 empresas da Europa. Para isso, desenvolveram o MLGM com distribuição beta-binomial através de componentes individuais de custos logísticos como transporte, armazenamento, inventário, administrativo e empacotamento. Apesar de na própria metodologia descrever a melhor aderência da distribuição gama para estimar custos, o uso da beta-binomial foi devido ao fato do custo ser uma razão e não valor absoluto.

De acordo com Swanson et al. (2013) os MLGM quando comparados ao MLG propiciam melhor ajuste aos dados, é melhor para eliminar a correlação dos resíduos e com intervalos mais confiáveis. Para Kwak et al. (2012) a precisão do MLGM melhora consideravelmente pois considera a dependência das observações enquanto que o MLG assume a independência entre elas. E para Krueger e Mongomery (2014) os MLGM são boas abordagens para grandes bases de dados, provendo melhores predições do que os MLG.

3 MATERIAL E MÉTODOS

Para estimar o custo da colheita mecanizada de cana-de-açúcar, foram desenvolvidos dois modelos de regressão. Um modelo linear generalizado (MLG) e um modelo linear generalizado misto (MLGM). Os modelos de regressão possibilitam relacionar variáveis explanatórias (independentes) com uma variável resposta (dependente) através de uma equação matemática.

No presente estudo, são consideradas como variáveis independentes os indicadores operacionais da colheita e como variável dependente o custo em reais por tonelada. Tanto no modelo MLG quanto no modelo MLGM, utilizou-se a distribuição gama, pois quando se trata de dados de custo, suas características são sempre positivas e na maioria das vezes, sua distribuição de probabilidade é altamente assimétrica. Além disso, essa distribuição é contínua e flexível para acomodar diferentes formas de distribuição de acordo com seus parâmetros de média e variância.

Os dados utilizados para modelagem do custo da colheita foram de 20 usinas localizadas na região centro-sul do Brasil no período de cinco safras (11/12 a 15/16, entre os meses abril a novembro). Os dados foram empregados para o desenvolvimento das bases a partir dos indicadores operacionais e de custo, para a análise de *outliers,* a seleção das variáveis, as análises de diagnósticos e a validação do modelo.

3.1 Indicadores operacionais

Os dados referentes aos indicadores operacionais foram extraídos a partir de computadores de bordo instalados nas colhedoras das usinas. As informações foram transmitidas via satélite nas áreas cobertas ou via GPRS em áreas com falta de sinal, utilizando-se cartões de memória para o descarregamento e envio dos dados. Esses dados foram recebidos em sistemas internos que trataram as informações através de regras e agrupamentos, disponibilizando a informação um dia após o recebimento, organizando em indicadores médios por usina. Os dados referem-se ao tempo de corte, tempo de manobra, velocidade da colhedora, consumo de combustível, quantidade média de cana colhida por colhedora, horímetro médio, tempo de manutenção e espaçamento da cana.

3.2 Indicadores de custo

Cada uma das usinas possui um código identificador de custo onde são alocadas todas as despesas referentes à colheita mecanizada. Estruturalmente, os custos da colheita seguem a dinâmica apresentada na Figura 2.

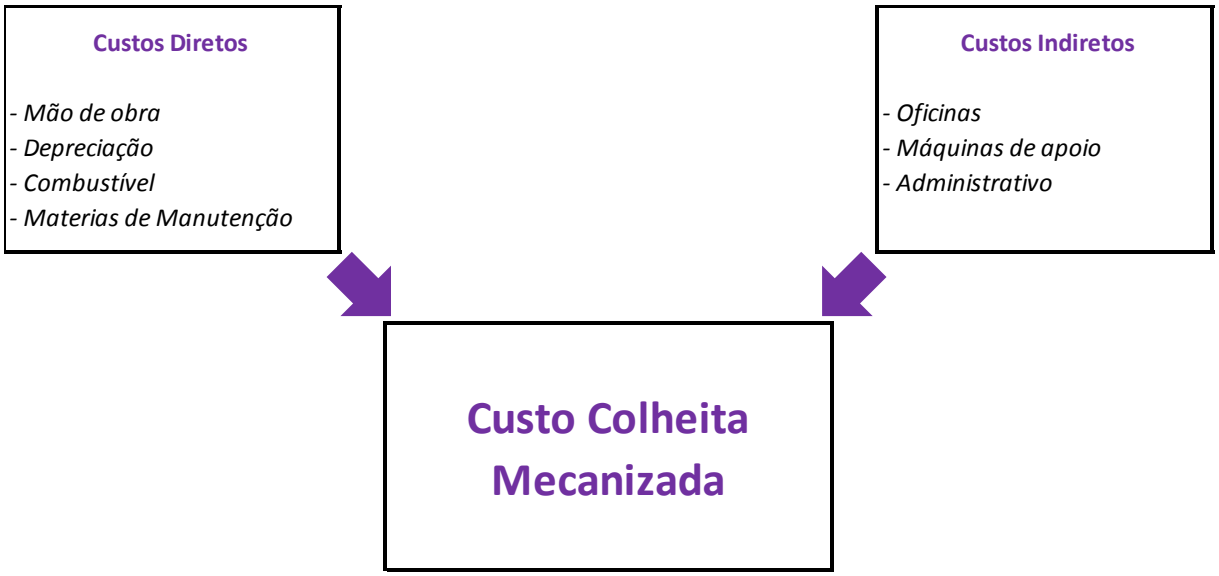

Figura 2 - Dinâmica da estrutura de custos da colheita mecanizada de cana

Os custos diretos são atribuídos em primeira instância. Todas as despesas de operadores de colhedora, depreciação da máquina, utilização de combustível e materiais de manutenção, são alocados imediatamente para o código identificador.

Os custos indiretos são alocados em forma de rateio. Todo o custo da estrutura que dá apoio à colheita, mas não exerce a operação em si, são distribuídos proporcionalmente de acordo com quantidade de trabalho exercida. Por exemplo, a oficina pode ter trabalhado para a colheita, para o plantio e para o transporte. No entanto, somente o custo proporcional às horas trabalhadas para a colheita é rateado e alocado de forma indireta para formação do custo final da colheita.

As informações de custo são extraídas via sistema SAP, sendo disponibilizadas somente após fechamento contábil, uma vez a cada mês. Nesse mesmo sistema, além das informações de custo, é possível extrair também informação referente à quantidade de operadores de colhedora, líderes e fiscais de frente, o tipo de escala de trabalho e nível de senioridade do cargo.

3.3 Desenvolvimento da base com variáveis explanatórias

Duas bases foram utilizadas para desenvolver o modelo. Uma base chamada a priori e outra chamada a posteriori. A base a priori continha indicadores operacionais que foram extraídos antes de fechar o custo mensal. Já a posteriori continha informações de custo da colheita, sendo obtida após o fechamento do custo mensal.

A base para desenvolvimento do modelo foi obtida através da junção da base a priori com a base a posteriori utilizando como chave a usina, o mês de referência e ano safra com os indicadores operacionais e o custo. Os custos de amortização foram desconsiderados na análise.

3.4 *Outliers*

Finalizada a base de desenvolvimento, antes de utilizá-la para criação dos modelos, realizou-se uma análise de *outliers*, que são observações na base que se diferenciam muito da dispersão das outras observações. Assim, para cada indicador operacional, bem como para o custo da colheita, considerou-se um *outlier* (sendo excluído da base), o valor referente à observação que estivesse fora do intervalo em:

$$[Q1-2,5(Q3-Q1),\ Q3+2,5(Q3-Q1)] \qquad (4)$$

em que, Q1, Q2 e Q3 são quartil 1, quartil 2 e quartil 3 respectivamente.

Além disso, as usinas que optaram pela terceirização da colheita foram desconsideradas da análise pela falta de informações dos indicadores operacionais que são utilizados para estimar o custo. Nas safras 11/12 e 12/13, 16 usinas realizaram a operação com recurso próprio e 4 optaram pela terceirização. Já nas safras 13/14 e 14/15 nenhuma das usinas optou pela terceirização. E na safra 15/16, 19 usinas utilizaram recurso próprio e 1 usina foi desativada.

3.5 Seleção de variáveis

Após a extração dos *outliers*, para desenvolver os modelos realizou-se a seleção de variáveis. Neste passo, selecionaram-se os indicadores operacionais que auxiliaram na explicação do comportamento do custo.

Essa escolha foi realizada em duas etapas. Na primeira avaliou-se a distribuição do custo em suas categorias de mão de obra, depreciação, combustível,

materiais e manutenção para verificar as mais relevantes, possibilitando assim, a inferência de variáveis relacionadas ao custo.

Na segunda etapa, construiu-se uma matriz de correlação do custo com os indicadores da etapa anterior a fim de avaliar o grau de dependência entre eles. Os indicadores com alto grau de correlação com o custo aumentaram suas chances de entrar no modelo como uma variável explanatória.

Assim, as variáveis de alta correlação com o custo foram testadas na forma contínua, enquanto que as de baixa correlação foram testadas tanto na forma contínua quanto na forma categorizada como uma forma alternativa de entrar no modelo. Definiu-se a categorização das variáveis em intervalos através de uma análise exploratória de gráficos de dispersão, de forma que a média de cada intervalo estivesse ordenada.

Com as variáveis tratadas, o teste estatístico utilizado para avaliar se as mesmas deveriam entrar no modelo levou em consideração o nível de significância a 5%. A seleção das variáveis e estimação dos parâmetros para o MLG foi obtida de acordo com Paula (2011):

$$U(\beta)=\frac{\partial l(\beta)}{\partial \beta} \tag{5}$$

$$l(\beta)=\frac{1}{\phi}\sum_{i=1}^{n}\left[y_i\theta_i-b(\theta_i)\right]+\sum_{i=1}^{n}c(y_i,\phi) \tag{6}$$

sendo, ϕ parâmetro de dispersão, θ_i parâmetro denominado canônico, b(.) e c(.) funções conhecidas.

Para os MLGM, a seleção das variáveis foi obtida por (JIANG, 2006):

$$\int \exp\{-q(x)\}\,dx \tag{7}$$

onde q(.) é uma função que atinge seu mínimo valor em $\mathbf{x}=\dot{\mathbf{x}}$, com q'($\dot{\mathbf{x}}$)=0 e q''($\dot{\mathbf{x}}$)=0. Então, tem-se uma aproximação,

$$\int \exp\{-q(x)\}\,dx \approx \sqrt{\frac{2\pi}{q''(\dot{x})}}\,\exp\{-q(\dot{x})\} \tag{8}$$

3.6 Desenvolvimento dos modelos

Os modelos foram desenvolvidos utilizando o software R versão 3.1.2, utilizando os pacotes *glm* (para MLG), *glmer* (para MLGM) e *hnp* (para diagnóstico do modelo). A diferença entre os dois modelos foi a inclusão de variáveis aleatórias para o modelo MLGM. A análise das variáveis aleatórias visa identificar como suas características podem influenciar a variabilidade da variável resposta e causar correlação entre as observações.

Através da utilização do software, foi possível estimar os coeficientes das variáveis selecionadas permitindo modelar a equação para estimativa do custo a partir dos indicadores operacionais.

3.7 Análise de Diagnósticos

Atualmente não há técnica ou método padrão para análise do ajuste do modelo. Em geral, os diagnósticos podem ser com análise de resíduos ou gráficos (CORDEIRO; DEMÉTRIO, 2014), com análise de critérios (MCCULLAGH; NELDER, 1989) e com análise de métricas (AUSTIN, 2003).

O diagnóstico do modelo foi realizado por meio de três abordagens: a) análise gráfica *half normal* com envelope simulado (ATKINSON, 1987), b) critério de Akaike, AIC (MCCULLAGH; NELDER, 1989) e c) métricas propostas por Austin et al. (2003), média predita do erro quadrado (MPEQ) e média absoluta do erro predito (MAEP).

O gráfico *half normal com envelope simulado* é obtido seguindo-se as etapas:

1) Ajustar um determinado modelo a um conjunto de dados obtendo os valores absolutos dos resíduos;
2) Simular 19 amostras da variável resposta, usando os parâmetros obtidos na etapa anterior com as mesmas variáveis que ajustaram o modelo;
3) Para cada uma das 19 amostras, ajustar o mesmo modelo e calcular os valores absolutos ordenados dos resíduos;
4) Em cada amostra, calcular o mínimo e máximo dos resíduos;
5) Construir um gráfico onde no eixo das ordenadas estão os valores dos resíduos do modelo ajustado juntamente com os valores mínimo e

máximo obtidos. No eixo das abcissas estão os quantis teóricos dado pelo valor esperado da distribuição *half normal:*

$$\Phi^{-1} = \frac{i + n - \frac{1}{8}}{2n + \frac{1}{2}} \tag{9}$$

em que i=1…n, com n tamanho da amostra.

Sob o modelo correto, os resíduos observados da amostra original estão dentro dos limites do envelope entre os valores mínimos e máximos dos resíduos das 19 amostras. Atkinson (1987) sugere gerar 19 amostras, pois desse modo, a probabilidade do maior resíduo de um envelope particular exceder o limite superior fica sendo = 1/20 = 0.05. Para o estudo em questão foram utilizadas 99 amostras.

As métricas foram definidas por Austin et. al (2003):

$$MPEQ = \frac{1}{n}\sum_{k} (\dot{Y}_k - Y_k)^2 \tag{10}$$

MPEQ – média predita do erro quadrado;

$\dot{Y}_k$ – média predita;

Y_k – média observada.

$$MAEP = \frac{1}{n}\sum_{k} |\dot{Y}_k - Y_k| \tag{11}$$

MAEP – média absoluta do erro predito;

$\dot{Y}_k$ – média predita;

Y_k – média observada.

O critério utilizado (MCCULLAGH e NELDER, 1989):

$$AIC = -2 \log L(\theta|y) + 2d \tag{12}$$

em que,

AIC – Akaike´s information criterion;

$L(\theta|y)$ – máximo valor do logaritmo da verossimilhança;

d – número de parâmetros do modelo.

Quanto menor o valor das métricas MPEQ e MAEP e quanto menor o valor do critério AIC, melhor o ajuste do modelo.

Para verificar a aderência dos modelos MLG e MLGM em suas três abordagens, realizou-se uma análise comparativa com o modelo nulo (média) e com o modelo linear (distribuição normal).

3.8 Validação do modelo

A validação do modelo foi realizada por meio da metodologia da validação cruzada. A validação cruzada consiste em dividir uma base de dados de tamanho n em 2 partes disjuntas. Uma parte para desenvolvimento e calibração do modelo de tamanho n-m e outra parte de tamanho m para validar o modelo. Na parte de validação do modelo, os valores observados e os valores preditos são comparados para avaliar a discrepância entre eles. (ZUCCHINI, 2000).

De acordo com Aguiar (2013), as partições mais comuns são de 70% para desenvolvimento e 30% para validação ou 90% para desenvolvimento e 10% para validação. No presente estudo será utilizado o segundo caso. A escolha das bases de desenvolvimento e de validação foi definida de acordo com as seguintes etapas: i) para cada observação da base desenvolvida em 3.3, gerou-se um número aleatório entre 0 e 1 distribuídos uniformemente utilizando a função aleatório do excel e ii) a base foi ordenada de acordo com o valor aleatório gerado e as primeiras observações que acumularem 90% da base total foram utilizadas para desenvolver o modelo e as 10% restantes para validação.

O modelo foi ajustado na base 90%, realizando em seguida, a análise de diagnósticos. Com isso, determinaram-se os coeficientes da equação que foram utilizados na base 10% para comparar o valor predito com o realizado através das métricas MAEP e MPEQ.

4 RESULTADOS E DISCUSSÃO

A base de dados para desenvolvimento dos modelos (junção das bases a priori com a posteriori) resultou em 731 observações. Cada observação coletada refere-se a uma das 20 usinas, um dos meses de safra (abril a novembro) e uma das safras (11/12 a 15/16). Os indicadores coletados foram quantidade de colhedoras, quantidade de operadores, horímetro médio das colhedoras, quantidade de litros de combustível das colhedoras, quantidade de horas realizando corte ou manobra, volume de cana colhida e custos da colheita em reais por tonelada.

4.1 Análise de *Outliers*

A análise de *outliers* reduziu 104 observações conforme mostrado na Figura 3.

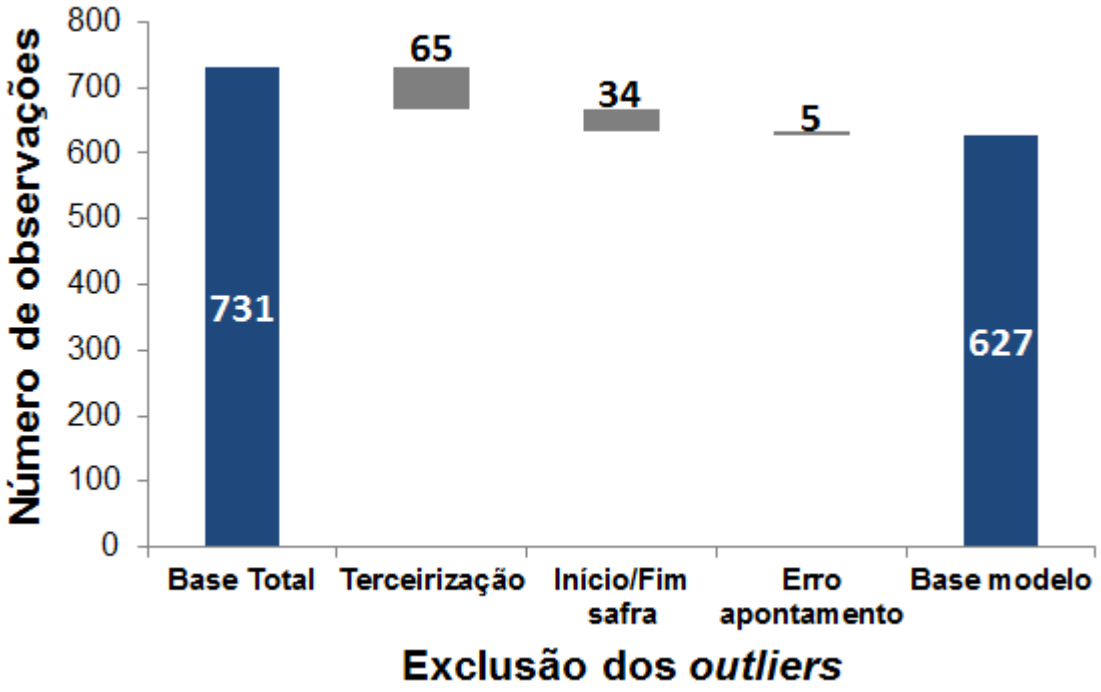

Figura 3 – Extração de dados inconsistentes e *outliers* na base de dados

Dos 104 casos excluídos, 65 foram devido à terceirização da colheita, 34 casos foram devidos ou ao início da safra (abril) ou fim da safra (novembro). Em unidades em que a safra não começa no início do mês ou não termina no fim do mês, é realizado um ajuste manual transferindo parte dos custos para entressafra, diluindo o custo fixo e causando inconsistências nos valores do custo

por tonelada. Os 5 casos restantes foram excluídos por erros de apontamento (2 inconsistências de quantidade de operadores, 2 inconsistências de horímetro da colhedora e 1 inconsistência por horas de utilização da colhedora). Em todos esses casos, os valores estavam fora do intervalo definido em (4).

Dos 627 casos remanescentes realizou-se a seleção da base para desenvolvimento do modelo com 90% das observações, resultando em 564 no total. As análises referentes ao ajuste dos modelos e diagnósticos foram relacionadas na base 90%. A base com os 10% restantes foi utilizada na validação do modelo.

4.2 Seleção de variáveis

A Figura 4 apresenta a distribuição da variável resposta (custo).

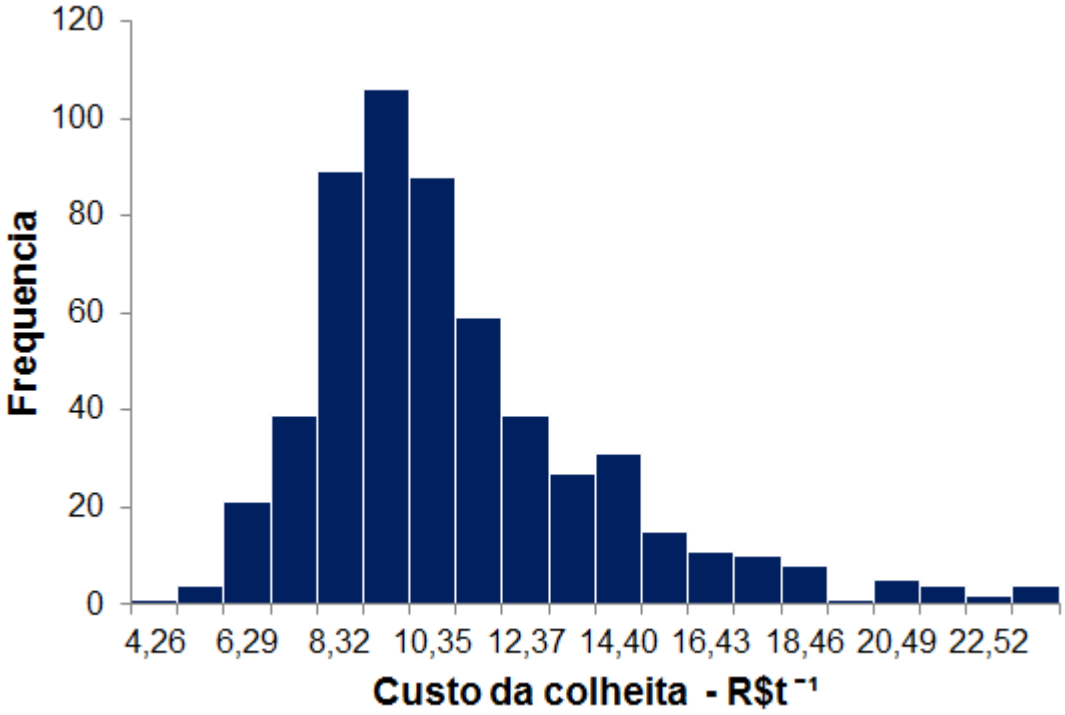

Figura 4 – Histograma do custo da colheita mecanizada de cana-de-açúcar

O custo da colheita apresentou uma distribuição assimétrica, o que pressupõe que as técnicas referentes ao modelo linear não terão boa aderência. O MLG e o MLGM com função de ligação gama são boas opções devido a sua distribuição ter característica de flexibilidade possibilitada pela variação dos parâmetros da sua função de densidade de probabilidade.

Para auxiliar na escolha de variáveis, a Figura 5 mostra proporcionalmente a representatividade de cada categoria de custo.

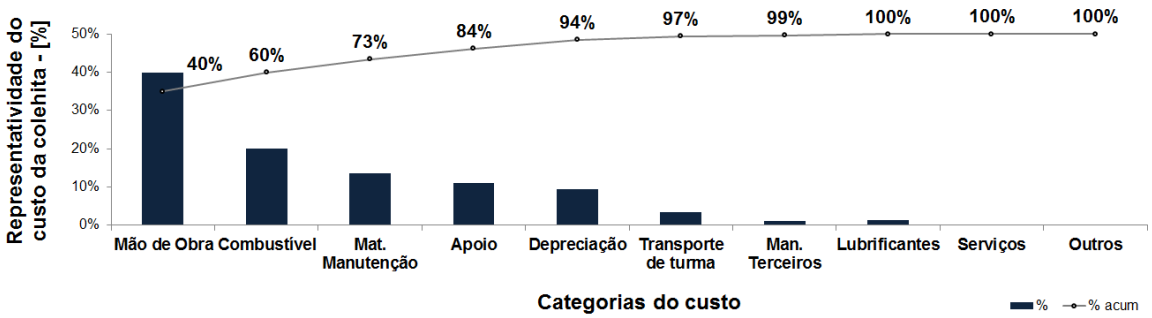

Figura 5 – Distribuição das categorias do custo em ordem de representatividade

Em cada categoria, estão contidos os custos de: a) mão de obra: ordenados de operadores de colhedora, líderes e auxiliares agrícolas; b) combustível: abastecimento de óleo diesel na colhedora; c) mat. manutenção: materiais e manutenção da colhedora; d) apoio: suporte da colheita, como por exemplo, comboios, caminhão prancha entre outros; e) depreciação: desvalorização da colhedora; f) transporte de turma: transporte dos operadores da usina até o campo; g) man. terceiro: manutenções externas; h) lubrificantes: lubrificantes utilizados na colhedora; i) serviços: serviços de terceiros; j) outros: categorias complementares.

De acordo com a Figura 5, as categorias de mão de obra, combustível, materiais e manutenção, apoio e depreciação já representam grande parte do custo contemplando 94% do total.

Assim, os seguintes indicadores foram testados para desenvolver o modelo: número de operadores por colhedora (nop); produtividade (tonelada colhida por máquina por mês); consumo (litros de diesel por tonelada); utilização (horas da colhedora realizando corte ou manobra); horímetro (horas acumuladas da colhedora).

Uma matriz de correlação dos indicadores operacionais com o custo foi calculada a fim de analisar quais possuem maior probabilidade de entrar no modelo. Quanto maior correlação com o custo, maior a chance de ser significativo (Tabela 2).

Tabela 2 – Matriz de correlação das variáveis com o custo

	Índice de correlação dos indicadores operacionais com o custo					
	Custo	Nop	Produtividade	Utilização	Horímetro	Consumo
Custo	1					
Nop	0,1209	1				
Produtividade	-0,6710	-0,0561	1			
Utilização	-0,5591	0,1103	0,8397	1		
Horímetro	0,2848	-0,2039	0,0208	-0,0431	1	
Consumo	0,4195	0,2997	-0,5427	-0,0770	-0,1593	1

Nop - número de operadores por colhedora

A produtividade, utilização e consumo são as variáveis que mais têm correlação com o custo. Além disso, a relação entre utilização e produtividade tem correlação forte (|0,7| a |0,9|) e a relação entre consumo e produtividade tem correlação moderada (|0,5| a |0,7|). Quando a correlação é considerada moderada, forte ou muito forte (|0,9| a |1,0|), em geral somente uma das variáveis torna-se significativa para o modelo.

As Figuras 6 e 7 permitem inferir uma tendência dos indicadores com o custo através de uma correlação negativa. Quanto maior a produtividade, menor o custo e quanto maior a utilização, menor o custo. Já na Figura 8, a correlação é positiva, quanto maior o consumo, maior o custo. Essas tendências estão de acordo com os valores dos índices de correlação da Tabela 2 (-0,6710, -0,5591 e 0,4195 para produtividade, utilização e consumo respectivamente). Quanto maior o valor do índice, mais visível a tendência.

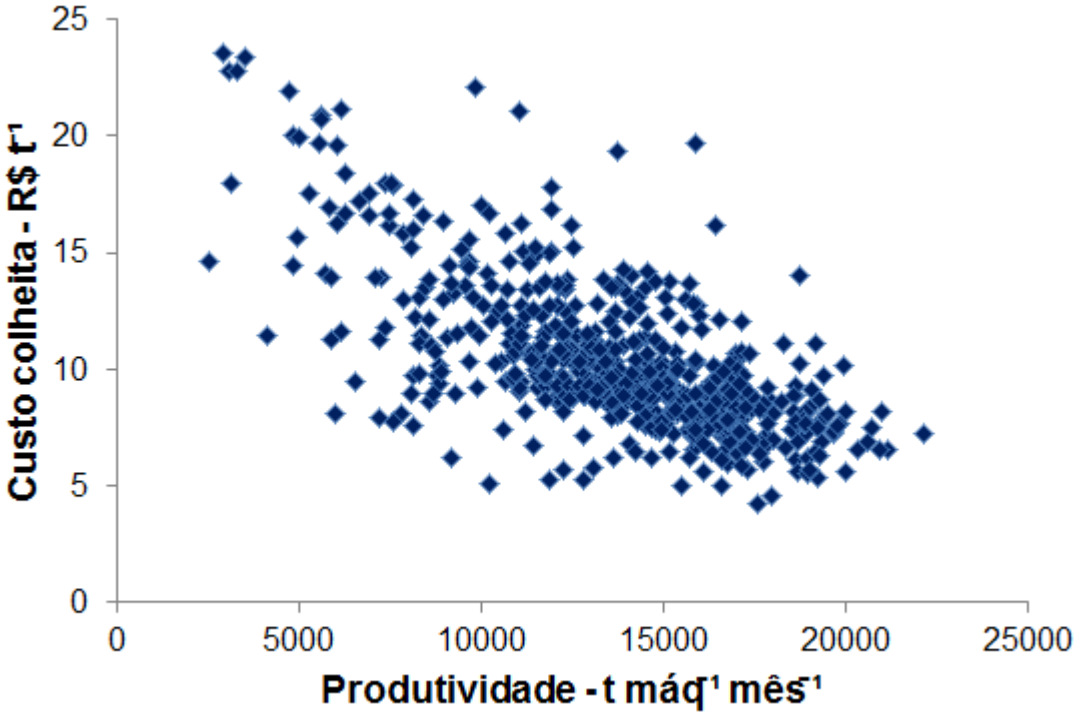

Figura 6 – Produtividade da colhedora versus custo da colheita

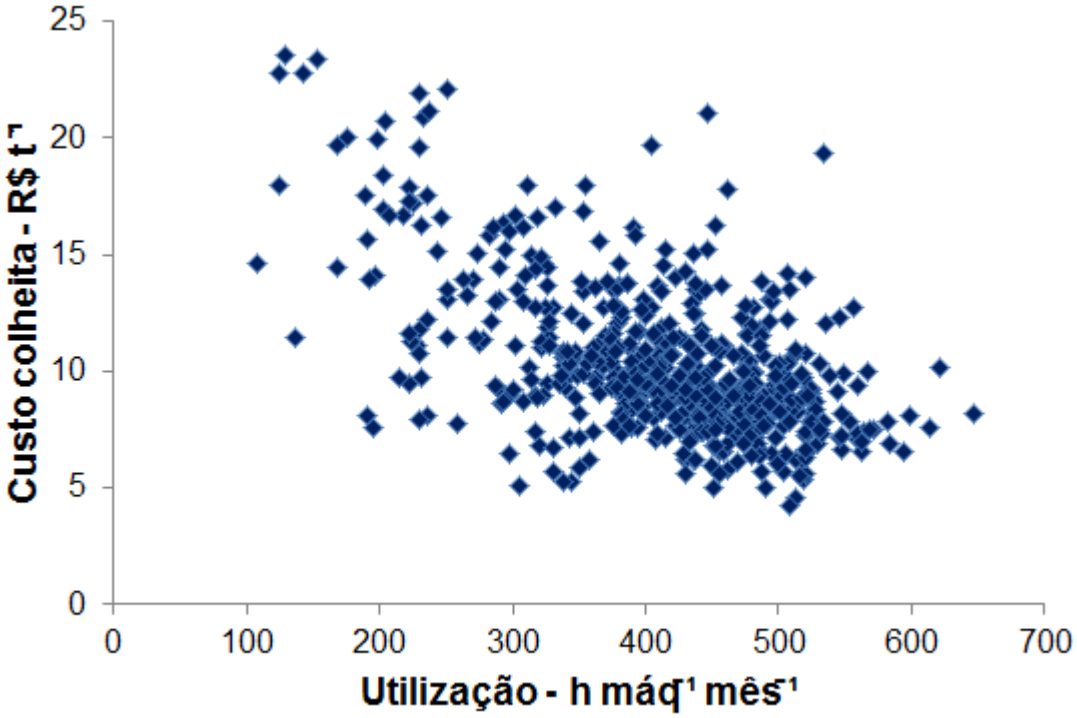

Figura 7 – Utilização da colhedora versus custo da colheita

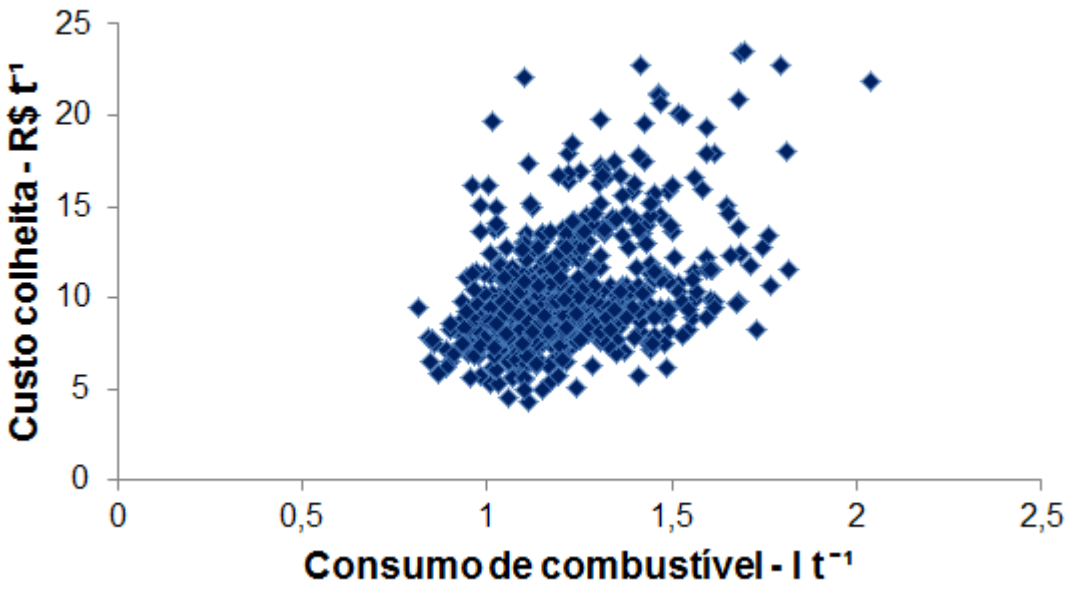

Figura 8 – Consumo de combustível versus custo da colheita

As tendências visíveis nas Figuras 6 a 8 já não são percebidas nas Figuras 9 e 10 e isso se reflete também nos valores dos índices de correlação da Tabela 2 (0,1209 e 0,2848 para número de operadores por colhedora e horímetro médio respectivamente). Por possuírem correlação baixa com o custo (0 a |0,3|), horímetro e nop foram testados tanto na forma contínua como na forma categorizada. Essa categorização é uma forma alternativa para fazer com que as variáveis entrem no modelo melhorando sua predição.

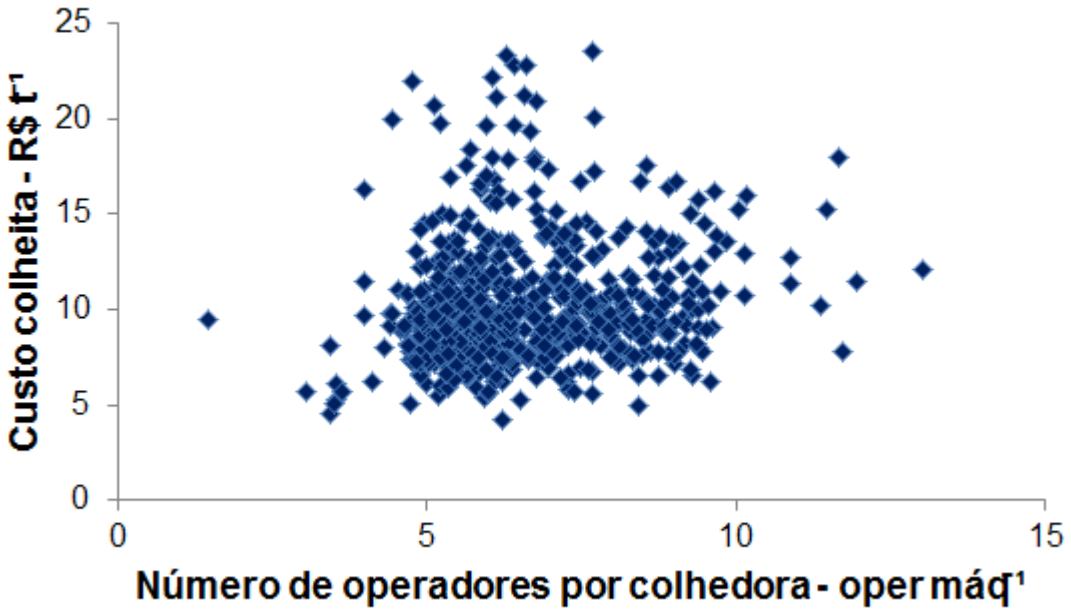

Figura 9 – Relação de operadores por colhedora versus custo da colheita

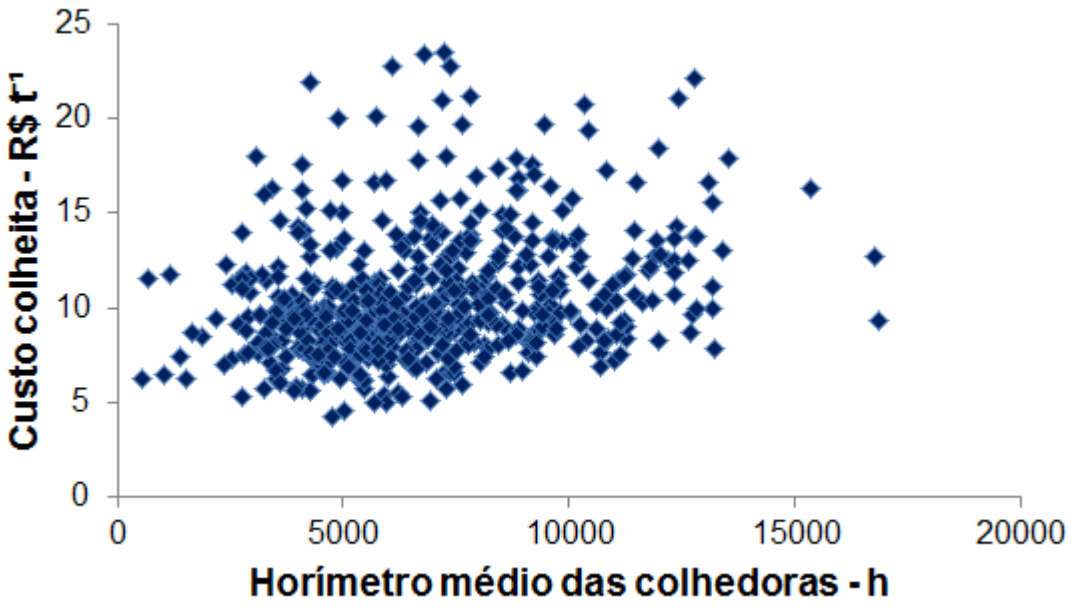

Figura 10 – Horímetro médio da colhedora versus custo da colheita

A Tabela 3 mostra as categorizações para as variáveis horímetro e nop. As médias dos custos da colheita para a variável horímetro ficaram em 9,22; 9,44; 10,95 e 13,06 reais por tonelada para até 3000h, de 3000h-6000h, de 6000h-12000h e

acima 12000h respectivamente. Para a variável número de operadores por colhedora, as médias do custo foram de 9,38; 10,40 e 12,90 para até 5, de 5 a 10 e acima de 10 respectivamente.

Tabela 3 - Categorização das variáveis horímetro e nop

Categoria das variáveis			
Horím. (h)		nop	
De	A	De	A
0	3000	0	5
3000	6000	5	10
6000	12000	10	15
12000	∞	15	∞

Horím.: horímetro

nop: Número de operadores

4.3 Desenvolvimento dos modelos

O modelo foi testado com as variáveis produtividade, consumo e utilização de forma contínua e horímetro e nop de forma contínua e categórica conforme Tabela 3. O pacote estatístico do software R utilizado para a construção do modelo MLG foi o *glm* e para o modelo MLGM o *glmer*. Em ambos os casos, a distribuição utilizada foi a gama e a função de ligação foi a logaritmica[1].

As variáveis testadas de forma contínua sofreram uma transformação logarítmica a fim de atenuar suas variabilidades.

4.4 Modelo MLG

Na Tabela 4 pode-se observar os coeficientes estimados do modelo MLG, o seu desvio padrão, o valor t (proporção coeficiente/desvio padrão) e a significância da variável. A legenda abaixo da Tabela mostra o grau de significância de cada variável. Quanto menor o valor, maior a chance dela explicar o modelo.

O número de operadores trouxe melhor resultado como variável contínua apesar da sua correlação com o custo ter valor considerado baixo. Já o horímetro teve melhor desempenho de forma categorizada.

[1] Log(μ) – Tabela 1

Tabela 4 – Valores estimados dos coeficientes das variáveis para o modelo MLG

Variável	Valores estimados				
	Coeficiente	Desvio	t value	Pr(>|t|)	Significância
Intercepto	7,2732	0,2832	25,6840	< 2e-16	***
Produtividade	- 0,5456	0,0293	-18,6500	< 2e-16	***
Consumo	0,1693	0,0668	2,5330	0,0116	**
Número de operadores	0,2128	0,0375	5,6700	0,0000	***
horímetro [até 3000h]	- 0,3282	0,0540	- 6,0750	0,0000	***
horímetro (3000h, 6000h]	- 0,3361	0,0407	- 8,2600	0,0000	***
horímetro (6000h, 12000h]	- 0,1431	0,0403	- 3,5520	0,0004	***

Significância	*** 0,001	**0,01	*0,05	˙0,1

Assim, a partir dos dados da Tabela 4, pode-se estimar o custo da colheita, utilizando a seguinte equação:

$$Custo = \exp(7{,}2732 - 0{,}5456p + 0{,}1693c + 0{,}2128nop + ho_i) \quad (13)$$

em que,

p – produtividade da colheita, $tmaq^{-1}$mês;

c – consumo de combustível, lt^{-1};

nop – número de operadores por colhedora;

ho_i – horímetro médio da colhedora, h.

com i=1,2,3 representando até 3000h, de 3000h-6000h e 6000h-12000h com valores de coeficiente 0,3282; 0,3361 e 0,1431 respectivamente .

A categoria do horímetro acima de 12000h foi tomada como referência pelo software e é representada pelo valor do intercepto. Quando a observação possui mais de 12000h no horímetro da colhedora, o valor ho_i é igual a 0. O uso da exponencial deve-se ao fato de ser a inversa da função de ligação logarítmica utilizada no desenvolvimento do modelo.

4.5 Modelo MLGM

Para o modelo MLGM foram utilizadas as mesmas variáveis do MLG adicionando as variáveis aleatórias de mês de safra e usina. As características intrínsecas das unidades que ficam alocadas em diferentes regiões do país, bem como a sazonalidade inerente aos meses do ano influenciam na variabilidade e

correlação das observações. Por esse motivo, ambas as variáveis foram consideradas como fator aleatório no modelo.

Tabela 5 – Valores estimados dos coeficientes das variáveis para o modelo MLGM

Efeitos Fixos	Valores estimados				
	Coeficiente	Desvio	t value	Pr(>|t|)	Significância
Intercepto	7,5134	0,3323	22,6100	< 2e-16	***
Produtividade	- 0,5711	0,0345	-16,5330	< 2e-16	***
Consumo	0,1878	0,0724	2,5950	0,0094	*
Número de operadores	0,2088	0,0379	5,5070	0,0000	***
horímetro [até 3000h]	- 0,3594	0,0544	- 6,6020	0,0000	***
horímetro (3000h, 6000h]	- 0,3440	0,0407	- 8,4490	< 2e-16	***
horímetro (6000h, 12000h]	- 0,1381	0,0380	- 3,6320	0,0003	***

Significância	*** 0,001	**0,01	*0,05	˙0,1

Efeitos Aleatórios	Variância	Desvio
Unidade	0,0028	0,0528
Mês	0,0009	0,0307

As Tabelas 6 e 7 mostram os valores dos coeficientes relativos de cada mês e de cada usina.

Tabela 6 – Coeficiente de variabilidade de cada mês de safra

Mês	Coeficiente de variabilidade
Abril	- 0,0532
Maio	- 0,0005
Junho	- 0,0002
Julho	0,0206
Agosto	- 0,0154
Setembro	0,0159
Outubro	0,0416
Novembro	0,0024

Tabela 7 – Coeficiente de variabilidade de cada usina

Usina	Coeficiente de variabilidade
1	0,0136
2	0,0410
3	0,0764
4	- 0,0786
5	- 0,0926
6	0,0421
7	- 0,0583
8	- 0,0622
9	0,0539
10	0,0340
11	0,0271
12	0,0599
13	0,0643
14	0,0850
15	- 0,0223
16	- 0,0651
17	- 0,0550
18	0,0083
19	- 0,0208
20	- 0,0176

O custo da colheita para o modelo MLGM é estimado de forma semelhante ao MLG adicionando os efeitos aleatórios:

$$\text{Custo} = \exp\left(\begin{matrix}7{,}5134-0{,}5711p+0{,}1878c+0{,}2088nop+\\ ho_i+mês_k+usina_j\end{matrix}\right) \quad (14)$$

em que,

p – produtividade da colheita, $tmaq^{-1}mês$;

c – consumo de combustível, lt^{-1};

nop – número de operadores por colhedora;

ho_i – horímetro médio da colhedora, h;

$mês_k$ – mês safra;

$usina_j$ – usina do grupo.

com i=1,2,3 representando até 3000h, de 3000h-6000h e 6000h-12000h com valores de coeficiente 0,3594; 0,3440 e 0,1381 respectivamente;

k=1,2...,8 com valores de coeficientes da Tabela 6;

j=1,2...,20 com valores de coeficientes da Tabela 7.

4.6 Diagnósticos

As análises de diagnósticos dos modelos MLG e MLGM foram tratadas em 3 abordagens: a) gráfico *half normal* com envelope simulado; b) critério de Akaike – AIC; c) métricas MAEP e MPEQ. Para verificar a aderência dos modelos desenvolvidos, ambos foram comparados com mais dois modelos: o modelo nulo (média) e o modelo linear (supondo distribuição normal da variável resposta). No modelo linear, as mesmas variáveis do modelo MLG foram utilizadas e os valores dos coeficientes estão nos anexos.

As Figuras 11 a 14 mostram os gráficos *half normal* com envelope simulado para os modelos nulo, linear, MLG e MLGM respectivamente. Quanto mais pontos dentro dos limites do envelope, melhor ajustado é o modelo.

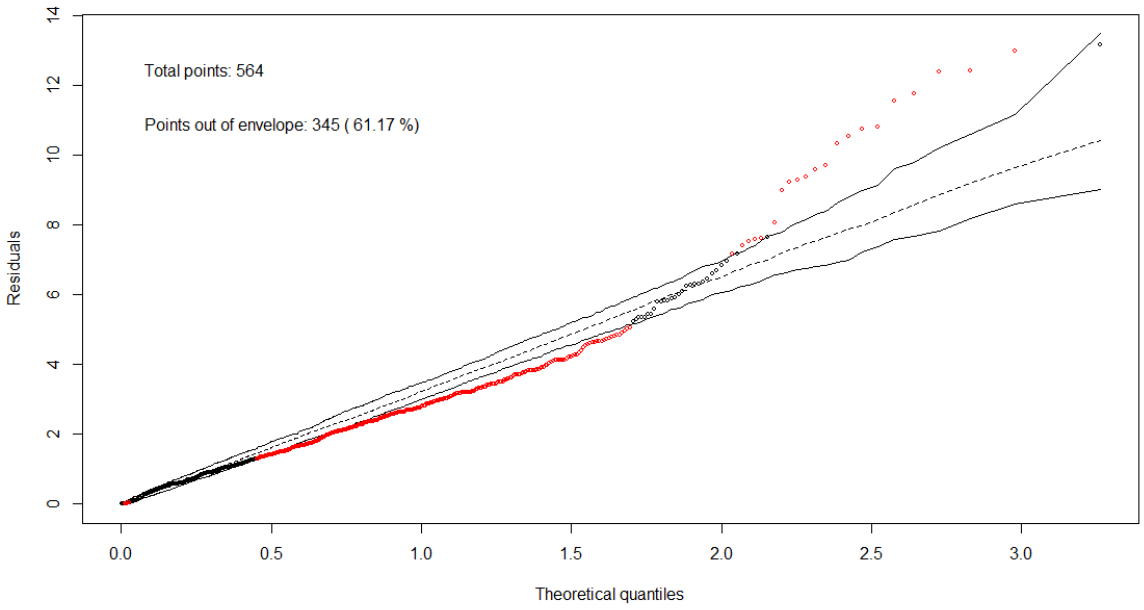

Figura 11 – gráfico *half normal* com envelope simulado para o modelo nulo

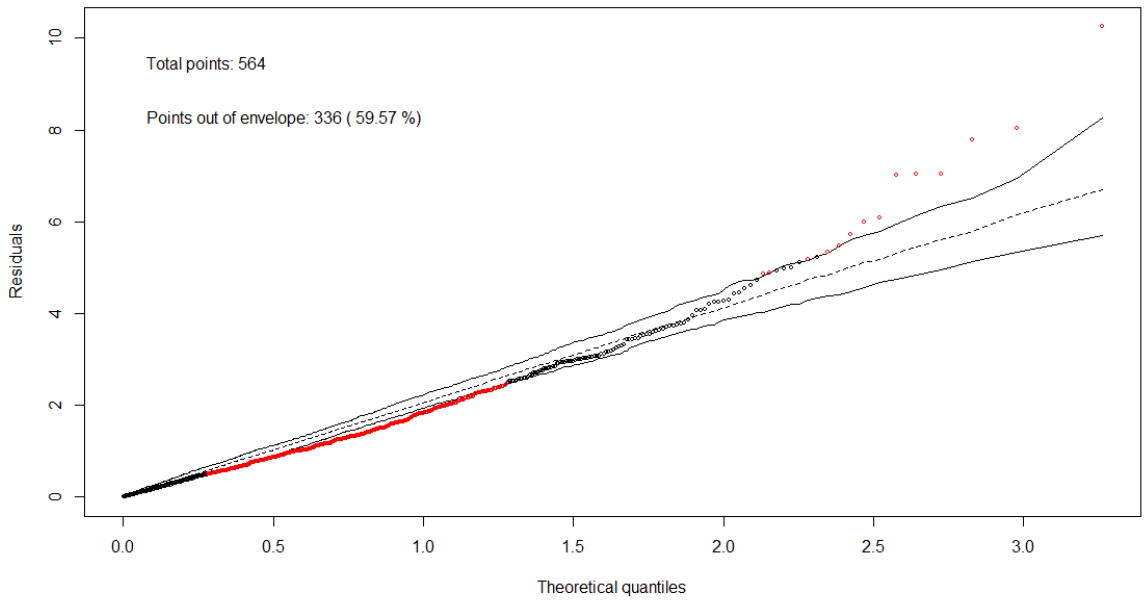

Figura 12 – gráfico *half normal* com envelope simulado para o modelo linear

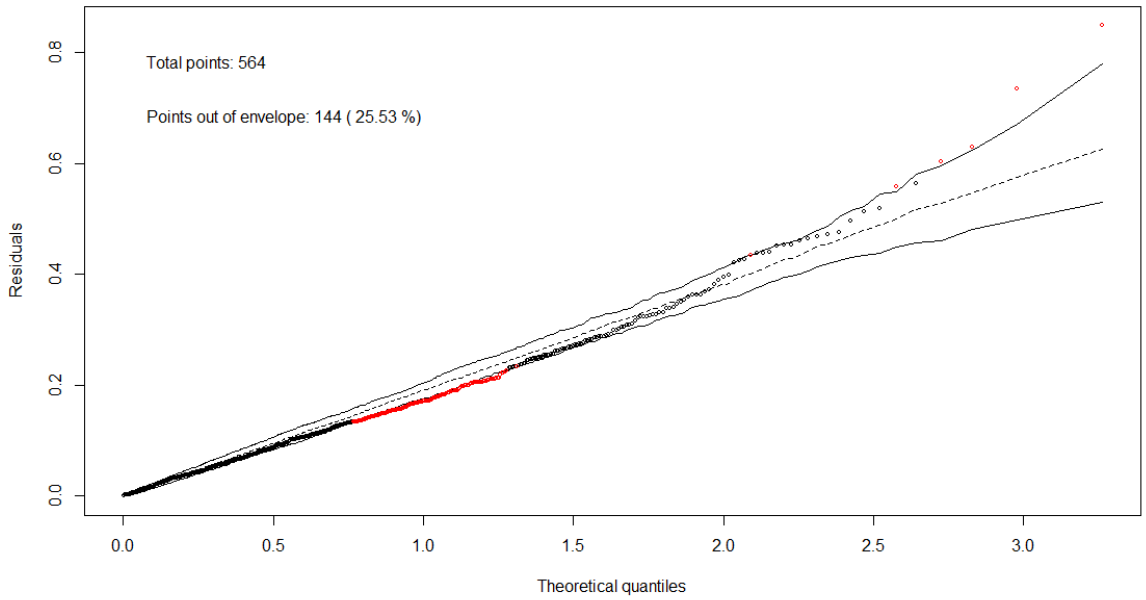

Figura 13 – gráfico *half normal* com envelope simulado para o modelo MLG

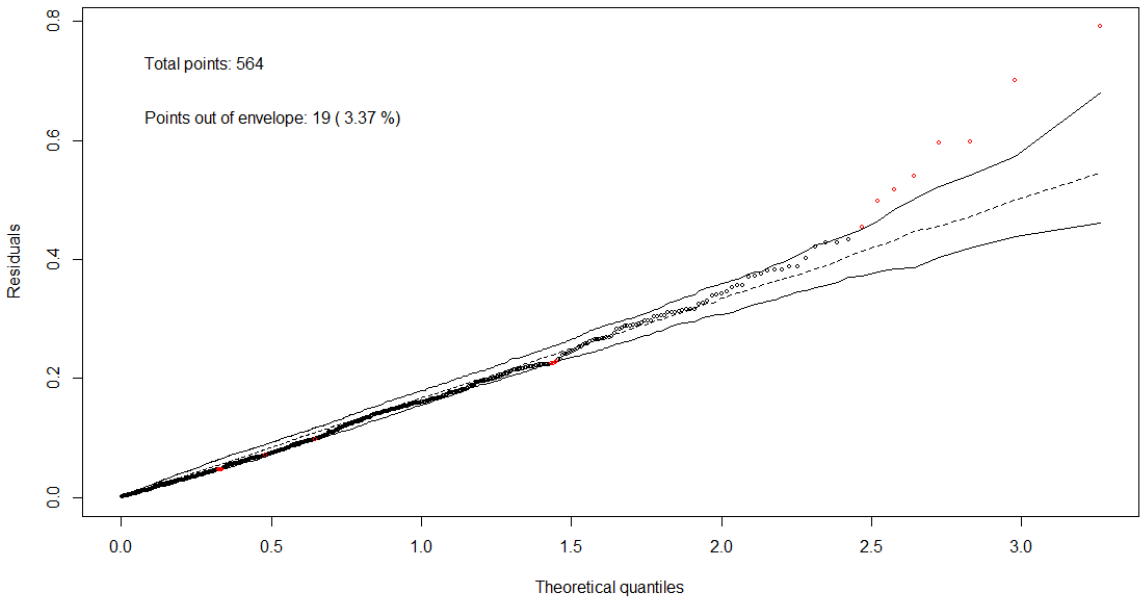

Figura 14 – gráfico *half normal* com envelope simulado para o modelo MLGM

Na análise gráfica *half normal* com envelope simulado, o modelo MLGM obteve melhor ajuste com apenas 3,37% pontos fora do envelope, seguido do modelo MLG com 25,53%, do modelo linear com 59,97% e por último o modelo nulo com 61,17%.

A Tabela 8 mostra os diagnósticos de critério e métrica dos modelos. Quanto menor o valor, melhor o ajuste do modelo.

Tabela 8 – Comparativo dos critérios e métricas dos diagnósticos dos modelos

Medidas	Modelos			
	MLGM	MLG	Modelo Linear	Modelo Nulo
AIC	2251,5	2325,2	2435,1	2937,1
MPEQ	3,5	4,2	4,3	10,6
MAEP	1,4	1,5	1,6	2,4

AIC – Critério de Akaike
MPEQ – Média predita do erro quadrado
MAEP – Média absoluta do erro predito

Assim como na análise gráfica *half normal* com envelope simulado, o modelo MLGM também obteve melhor resultado no critério AIC com 2251,5 e nas métricas MPEQ com 3,5 e MAEP com 1,4.

Tanto o modelo MLG quanto o MLGM possuem ajustes melhores que os modelos nulo e linear. No entanto, o grande ganho foi a inclusão dos efeitos

aleatórios no modelo MLGM resultando em performances muito superiores em relação a todos os outros.

4.7 Validação do modelo

Para a validação do modelo utilizou-se a base particionada com 10% das observações. Com as equações (13) do modelo MLG e (14) do modelo MLGM estimou-se o custo da colheita comparando com o custo real através das métricas MPEQ e MAEP. A comparação também foi realizada com os modelos nulo e linear.

A Tabela 9 mostra os valores das métricas calculados na base de validação do modelo. Assim como na análise de diagnóstico, o modelo MLGM também obteve melhor resultado quando testado em uma base diferente da que foi utilizada para desenvolvimento do modelo.

Tabela 9 - Comparativo das métricas da validação do modelo

Medidas	Modelos			
	MLGM	MLG	Modelo Linear	Modelo Nulo
MPEQ	3,2	3,5	3,5	9,5
MAEP	1,4	1,5	1,5	2,5

O único ponto de atenção para o modelo MLGM é que apesar de estimar bem o custo médio, algumas estimativas de valores extremos não ficaram bem ajustadas, conforme mostrado também por Dodd et al. (2006). Esse mau ajuste nos extremos podem ser devido a variáveis que não foram consideradas no modelo.

5 CONCLUSÃO

As variáveis que explicaram o custo foram produtividade, consumo de combustível, número de operadores por colhedora e horímetro.

Os efeitos aleatórios mês e usina contribuíram para explicar a variabilidade do custo, bem como as correlações existentes entre cada observação, possibilitando grande ganho e melhor ajuste para a técnica do MLGM.

O modelo apesar de estimar corretamente os valores em torno da média, não se ajusta para alguns valores extremos. Para trabalhos futuros, sugere-se investigar mais profundamente as dinâmicas de custo a fim de incluir novas variáveis ou utilizar novas técnicas para auxiliar na melhor estimativa dos custos.

REFERÊNCIAS

AGUIAR, R.G. **Balanço de energia em ecossistema amazônico por modelo de regressão robusta com *bootstrap* e validação cruzada**. 2013. 104 p. Tese (Doutorado em Física Ambiental) - Universidade Federal de Mato Grosso, Cuiabá, 2013.

AMERICAN SOCIETY OF AGRICULTURAL AND BIOLOGICAL ENGINEERS. Agricultural machinery management data ASAE D497.7. In: _____. **ASABE standards**. St. Joseph, 2011. p. 1-8.

ATKINSON, A.C. **Plots, transformations and regression.** Oxford: Oxford University Press, 1987. 296 p.

AUSTIN, P.C.; GHALI, W.A.; TU, J.V. A comparison of several regression models for analysing cost of CABG surgery. **Statistics in Medicine**, Calgary, v. 22, p. 2799–2815, 2003.

BALDI, I.; PAGANO,E.; BERCHIALLA, P.; DESIDERI, A.; FERRANDO, A.; MERLETTI, F.; GREGORI, D. Modeling healthcare costs in simultaneous presence of asymmetry, heteroscedasticity and correlation. **Journal of Applied Statistics,** Abingdon, v. 40, n. 2, p. 298-310, 2013.

BASU, A.; MANNING, W.; MULLAHY, J. Comparing alternative models: log vs Cox proportional hazard? **Health Economics,** Chicago, v. 13, n. 8, p. 749-765, 2004.

BAUTISTA, E.A.L. **Modelos lineares mistos e generalizados mistos em estudos de adaptação local e plasticidade fenotípica de *Euterpe edulis*.** 2014. 124 p. Tese (Doutorado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2014.

BOX, G.E.P.; COX, D.R. An analysis of transformation. **Journal of Royal Statistical Society.** Series B, London, v. 26, p.211–252, 1964.

BURLA, E.R.; FERNANDES, H.C.; MACHADO, C.C.; LEITE, D.M.; FERNANDES, P.S. Avaliação técnica e econômica do harvester em diferentes condições operacionais. **Engenharia na Agricultura,** Viçosa, v. 20, n. 5, p. 412-422, set./out. 2012.

CENTENO, A.S.; KAERCHER, D. Custo operacional das máquinas agrícolas. In: FNP CONSULTORIA & COMÉRCIO. **AGRIANUAL 2010**: anuário da agricultura brasileira. São Paulo: Agrafnp, 2010. p. 113-116.

CHAGAS, M.F.; BORDONAL, R.O.; CAVALETT, O.; CARVALHO, J.L.N.; BONOMI, A.; LA SCALA, JR., N. Environmental and economic impacts of different sugarcane production systems in the ethanol biorefinery. **Biofuels Bioproducts & Biorefining-Biofpr,** Malden, v. 10, n. 1, p. 89-106, Jan./Feb. 2016.

CONSELHO DE PRODUTORES DE CANA-DE-AÇÚCAR, AÇÚCAR E ETANOL DO ESTADO DE SÃO PAULO. Disponível em <www.consecana.com.br>. Acesso em: 30 abr. 2016.

CORDEIRO, G.M.; DEMÉTRIO, C.G.B. **Modelos lineares generalizados e extensões.** Piracicaba: [s.n.], 2014. 306 p.

CUNHA, J.P.A.R.; MARTINS, D.H.; CUNHA, W.G. Operational performance of the mechanized and semi-mechanized potato harvest. **Engenharia Agrícola***,* Jaboticabal, v. 31, n. 4, p. 826-834, jul./ago. 2011.

CUNHA, U.C.; RODRIGUES, J.R.F. A importância da contabilidade de custos na formação de preços em uma micro-empresa de uniformes profissionais. **REDIGE: Revista de Design, Inovação e Gestão Estratégica**, Rio de Janeiro, v. 3, n. 3, p. 1-24, dez. 2012.

DA COSTA, S.C. **Modelos lineares generalizados mistos para dados longitudinais.** 2003. 125 p. Tese (Doutorado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2003.

DODD, S.; BASSI, A.; BODGER, K.; WILLIAMSON, P. A comparison of multivariable regression models to analyse cost data. **Journal of Evaluation in Clinical Practice,** Oxford, v. 12, n. 1, p. 76-86, Jan. 2006.

DOS SANTOS, N.B.; FERNANDES, H.C.; GADANHA JUNIOR, C.D. Economic impact of sugarcane (*Saccharum spp.*) loss in mechanical harvesting. **Cientifica** Jaboticabal, v. 43, n. 1, p. 16-21, 2015.

ENGBLOM, J.; SOLAKIVI, T.; TÖYLI, J.; OJALA, L. Multiple-method analysis of logistics costs. **International Journal of Production Economics,** Amsterdam**,** v. 137, n. 1, p. 29-35, May 2012.

FAO. Disponível em <http://faostat3.fao.org/browse/Q/*/E>. Acesso em: 03 maio 2016.

FAUSTO, M.; CARNEIRO, M.; ANTUNES, C.M.F.; PINTO, J.A.; COLOSIMO, E.A. Mixed linear regression model for longitudinal data: application to an unbalanced anthropometric data set. **Cadernos de Saúde Publica,** Rio de Janeiro, v. 24, n. 3, p. 513-524, Mar. 2008.

FORTIN, M. Population-averaged predictions with generalized linear mixed-effects models in forestry: an estimator based on Gauss-Hermite quadrature. **Canadian Journal of Forest Research,** Ottawa, v. 43, n. 2, p. 129-138, Feb. 2013.

FRYTAK, J.R.; HENK, H.J.; DE CASTRO, C.M.; HALPERN, R.; NELSON, M. Estimation of economic costs associated with transfusion dependence in adults with MDS. **Current Medical Research And Opinion,** Abingdon, v. 25, n. 8, p. 1941–1951, 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, v. 18, n. 3, p. 234-240, maio/jun. 2010.

GBUR, E.E.; STROUP, W.W.; MCCARTER, K.S.; DURHAM, S.; YOUNG, L.J.; CHRISTMAN, M.; WEST, M. KRAMER, M. **Analysis of generalized linear mixed models in the agricultural and natural resources sciences.** Madison: American Society of Agronomy, 2012. 283 p.

GOREN, A.; ZOU, K.H.; GUPTA, S.; CHEN, C. Direct and indirect cost of urge urinary incontinence with and without pharmacotherapy. **International Journal of Clinical Practice,** Malden, v. 68, n. 3, p. 336-348, Mar. 2014.

GRIEVE, R.; NIXON, R.; THOMPSON, S.G.; NORMAND, C. Using multilevel models for assessing the variability of multinational resource use and cost data. **Health Economics,** Chicago, v. 14, n. 2, p. 185-196, Feb. 2005.

HALPERN, R.; NADKARNI, A.; KALSEKAR, I.; NGUYENM H.; SONG, R.; BAKER, R.A.; NELSON, J.C. Medical costs and hospitalizations among patients with depression treated with adjunctive atypical antipsychotic therapy: an analysis of health insurance claims data. **Annals of Pharmacotherapy,** Thousand Oaks, v. 47, n. 7/8, p. 933-945, July/Aug. 2013.

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**, Beachwood, v. 7, n. 4, p. 2062–2080, 2013.

HIGGINS, A.; DAVIES, I. Capacity planning in a sugarcane harvesting and transport system using simulation modeling. **Proceedings of the Australian Society of Sugar Cane Technologists,** Queensland, v. 26, p. 1-9, 2004.

HOFFMANN, R.; SERRANO, O.; NEVES, E.M.; THAME, A.C.M.; ENGLER, J.J.C. **Administração da empresa agrícola.** 7. ed. São Paulo: Pioneira, 1992. 325 p.

JIANG, J. **Linear and generalized linear mixed models and their applications.** 2nd ed. New York: Springer, 2006. 257 p.

KRUEGER, D.C.; MONTGOMERY, D.C. Modeling and analyzing semiconductor yield with generalized linear mixed models. **Applied Stochastic Models in Business and Industry,** Malden, v. 30, n.6, p. 691-707, Nov./Dec. 2014.

KURUSU, R.S. **Avaliação de técnicas de diagnósticos para análise de dados com medidas repetidas.** 2013. 144 p. Dissertação (Mestrado em Estatística) – Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2013.

KWAK, H.; LEE, W.K.; SABOROWSKI, J.; LEE, S.Y.; WON, M.S.; KOO, K.S.; LEE, M.B.; KIM, S.N. Estimating the spatial pattern of human-caused forest fires using a generalized linear mixed model with spatial autocorrelation in South Korea. **International Journal of Geographical Information Science,** London, v. 26, n. 9, p. 1589-1602, 2012.

KWONG, W.; DIELS, J.; KAVANAGH, S. Costs of gastrointestinal events after outpatient opioid treatment for non-cancer pain. **Annals of Pharmacotherapy,** Thousand Oaks, v. 44, n. 4, p. 630-640, Apr. 2010.

LANNA, G.B.M.; REIS, R.P. Influência da mecanização da colheita na viabilidade econômico-financeira da cafeicultura no Sul de Minas Gerais. **Coffee Science**, Lavras, v. 7, n. 2, p. 110-121, ago. 2012.

LIU, L.; STRAWDERMAN, R.L.; COWEN, M.E.; SHIH, Y.C.T. A flexible two-part random effects model for correlated medical costs. **Journal of Health Economics,** Amsterdam, v. 29, n. 1, p. 110-123, Jan. 2010.

MACHADO, C.C.; MALINOVSKI, J.R. **Ciência do trabalho florestal**. Viçosa: Universidade Federal de Viçosa, 1988. 65 p.

MANNING, W.G. The logged dependent variable, heteroscedasticity, and the retransformation problem, **Journal of Health Economics**, Amsterdam, v. 17, n. 3, p. 283–295, 1998.

MANOS, M.M.; DARBINIAN, J.; RUBIN, J.; RAY, G.T.; SHVACHKO, V.; DENIZ, B.; VELEZ, F.; QUESENBERRY, C. The effect of hepatitis c treatment response on medical costs: a longitudinal analysis in an integrated care setting. **Journal of Managed Care Pharmacy,** Alexandria, v. 19, n. 6, p.438-447, July/Aug. 2013.

MCCULLAGH, P.; NELDER, J.A. **Generalized linear models.** 2nd ed. London: Chapman and Hall, 1989. 511 p.

MIALHE, L.G. **Manual de mecanização agrícola.** São Paulo: Agronômica Ceres. 1974. 301 p.

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.

MITTAL, M.; HARRISON, D.L.; THOMPSON, D.M.; MILLER, M.J.; FARMER, K.C.; NG, Y.T. An evaluation of three statistical estimation methods for assessing health policy effects on prescription drug claims. **Research in Social & Administrative Pharmacy,** Amsterdam, v. 12, n. 1, p. 29-40, Jan./Feb. 2015.

MURAKAMI, Y.; OKAMURA, T.; NAKAMURA, K.; MIURA, K.; UESHIMA, H. The clustering of cardiovascular disease risk factors and their impacts on annual medical expenditure in Japan: community-based cost analysis using gamma regression models. **BMJ Open**, London, v. 3, p. 1-5, 2013.

NELDER, J.A.; WEDDERBURN, R.W.M. Generalized linear models. **Journal of the Royal Statistical Society.** Series A, London, v. 135, p. 370–384, 1972.

NEVES, S.; VICECONTI, P.V.E. **Contabilidade de custos um enfoque direto e objetivo.** 9. ed. São Paulo: Ed. Frase, 2010. 344 p.

NORONHA, F.N. **Projetos agropecuários.** 2. ed. São Paulo: Atlas, 1987. 269 p.

NOVACANA. **A produção de cana-de-açúcar no Brasil (e no mundo).** Disponível em <https://www.novacana.com/cana/producao-cana-de-acucar-brasil-e-mundo/>. Acesso em: 25 abr. 2016.

NUNES, J.A.R.; DE MORAIS, A.R.; BUENO FILHO, J.S.S. Modelagem da superdispersão em dados por um modelo linear generalizado misto. **Revista de Matemática e Estatística,** São Paulo, v. 22, n. 1, p. 55-70, 2004.

OLIVEIRA, E.; SILVA, F.M; SALVADOR, N.; SOUZA, Z.M.; CHALFOUN, S.M.; FIGUEIREDO, C.A.P. Custos operacionais da colheita mecanizada do cafeeiro. **Pesquisa Agropecuária Brasileira**, Brasília, v. 42, n. 6, p. 827-831, jun. 2007.

ONG, K.; LAU,E.; KEMNER, J.E.; KURTZ,S.M. Two-year cost comparison of vertebroplasty and kyphoplasty for the treatment of vertebral compression fractures: are initial surgical costs misleading? **Osteoporosis International,** Philadelphia, v. 24, n. 4, p. 1437-1445, Apr. 2013.

ORUETA, J.; ÁLVAREZ, A.G.; GOÑI, M.G.; PAOLUCCI, F.; SOLINÍS, R.N. Prevalence and costs of multimorbidity by deprivation levels in the basque country: a population based study using health administrative databases. **Plos One,** San Francisco, v. 9, n. 2, p. 1-11, Feb. 2014.

PAULA, G.A. **Modelos de regressão com apoio computacional.** São Paulo: USP, IME, 2011. 343 p.

PRATES, M.O.; ASELTINE, R.H.; DEY, D.K.; YAN, J. Assessing intervention efficacy on high-risk drinkers using generalized linear mixed models with a new class of link functions. **Biometrical Journal,** Weinheim, v. 55, n. 6, p. 912–924, 2013.

PRENZLER, A.; BOKEMEYER, B.; SCHULENBURG, J.M.; MITTENDORF, T. Health care costs and their predictors of inflammatory bowel diseases in Germany. **European Journal of Health Economics,** Heidelberg, v. 12, n. 3, p. 273-283, June 2011.

RIPOLI, T.C.C.; MIALHE, L.G. Custos de colheita da cana-de-açúcar no estado de São Paulo, 1981/82. **Álcool & Açúcar,** São Paulo, v. 2, n. 2, p. 18-26, 1982.

ROCKHILL, C.M.; JAFFE, K.; ZHOU, C.; FAN, M.; KATON, W.; FANN, J.R. Health care costs associated with traumatic brain injury and psychiatric illness in adults. **Journal of Neurotrauma,** New Rochelle, v. 29, p. 1038–1046, Apr. 2012.

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.

SILVA, A.A.T. **Influência local em modelos lineares generalizados mistos com variável resposta discreta.** 2014. 201 p. Tese (Doutorado em Estatística) – Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2014.

SILVA, F.M. **Colheita mecanizada e seletiva do café:** cafeicultura empresarial: produtividade e qualidade. Lavras: UFLA; Faepe, 2004. 75 p.

SOERENSEN, A.V.; DONSKO, F.; KJELLBERG, J.; IBSEN,R; HERMANN, G.G.; JENSEN, N.V.; FODE, K.; GEERTSEN, P.F. Health economic changes as a result of implementation of targeted therapy for metastatic renal cell carcinoma: national results from DARENCA study 2. **European Urology,** Sheffield, v. 68, n. 3, p. 516-522, Sept. 2015.

SWANSON, A.K.; DOBROWSKI, S.Z.; FINLEY, A.O.; THORNE. J.H.; SCWARTZ, M.K. Spatial regression methods capture prediction uncertainty in species distribution model projections through time. **Global Ecology and Biogeography,** Malden, v. 22, n. 2, p. 242-251, Feb. 2013.

TOURINO, M.C.C. **Arranjo populacional e uniformidade de semeadura na produtividade e outras características agronômicas da soja.** 2000. 139 p. Tese (Doutorado em Fitotecnia) – Universidade Federal de Lavras, Lavras, 2000.

UNÜTZER, J.; SCHOENBAUM, M.; KATON, W.J.; FAN, M.; PINCUS, H.A.; HOGAN, D.; TAYLOR, J. Healthcare costs associated with depression in medically ill fee-for-service medicare participants. **Journal of the American Geriatrics Society,** Malden, v. 57, n. 3, p. 506-510, Mar. 2009.

UPATISING, B.; WOOD,D.L.; KREMERS, W.K.; CHRIST, S.L.; YIH, Y.; HANSON, G.J.; TAKAHASHI, P.Y. Cost comparison between home telemonitoring and usual care of older adults: a randomized trial (Tele-ERA). **Telemedicine and E-Health,** Boston, v. 21, n. 1, p. 3-8, Jan. 2015.

VELOPULOS, C.G.; ENWEREM, N.Y.; OBIRIEZE, A.; HUI, X.; HASMI, 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**, New York, v. 184, p. 444-449, 2013.

VIEIRA, G. **Avaliação do custo, produtividade e geração de emprego no corte de cana-de-açúcar, manual e mecanizado, com e sem queima prévia.** 2003. 127 p. Dissertação (Mestrado em Agronomia) – Faculdade de Ciências Agronômicas, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Botucatu, 2003.

WAKEAM, E.; HYDER, J.A.; LIPSITZ, S.R.; DARLING, G.E.; FINLAYSON, S.R.G. Outcomes and costs for major lung resection in the United States: which patients benefit most from high-volume referral? **Annals of Thoracic Surgery,** Chicago, v. 100, n. 3, p. 939-946, Sept. 2015.

WITNEY, B. **Choosing and using farm machines.** Essex: Longman Scientific & Techical, 1988. 412 p.

YAU, K.K.W.; LEE, A.H.; NG, A.S.K. A zero-augmented gamma mixed model for longitudinal data with many zeros. **Australian & New Zealand Journal of Statistics,** Malden, v. 44, n. 2, p. 177-183, June 2002.

ZANLUCA, J.C. **A contabilidade e o controle de custos.** Disponível em: <http://www.portaldecontabilidade.com.br>. Acesso em 12 jun. 2014.

ZUCCHINI, W. An introduction to model selection. **Journal of Mathematical Psychology,** San Diego, v. 44, p. 41-61, 2000.

ANEXOS

Tabela 10 - Valores estimados dos coeficientes das variáveis para o modelo linear

Variáveis	Valores estimados				
	Coeficiente	Desvio	t value	Pr(>|t|)	Significância
Intercepto	69,6901	3,0680	22,7160	< 2e-16	***
Produtividade	- 6,4531	0,3169	-20,3610	< 2e-16	***
Consumo	1,6339	0,7241	2,2560	0,024	**
Número de operadores	2,1255	0,4065	5,2280	0,0000	***
horímetro [até 3000h]	- 3,6655	0,5853	- 6,2630	0,0000	***
horímetro (3000h, 6000h]	- 3,6405	0,4408	- 8,2590	0,0000	***
horímetro (6000h, 12000h]	- 1,6744	0,4364	- 3,8370	0,0001	***

Significância	*** 0,001	**0,01	*0,05	˙0,1

Código em programação R para testes estatísticos dos modelos.

```
library(hnp)
library(MASS)
library(hglm)
library(lme4)
setwd("C:/Users/cs213235/Desktop/Projetos/Mestrado/Dissertação")

base<-read.table("base5safras3.csv", header=T, sep=";")
attach(base)

'modelo nulo'
regn<-glm(r3~1, data=base)
hnp(regn, print.on=T, paint.out=T)
summary(regn)
predn<-fitted.values(regn)
xn<-cbind(c(predn))
MPEQn<-sum((xn-r3)^2)/nrow(base)
MPEQn
MAEPn<-sum(abs(xn-r3))/nrow(base)
MAEPn

'modelo normal'
reg<-glm(r3~log(prod)+horcat+log(l.t)+log(hct), data=base)
```

```
hnp(reg, print.on=T, paint.out=T)
summary(reg)
pred<-fitted.values(reg)
x<-cbind(c(pred))
MPEQ<-sum((x-r3)^2)/nrow(base)
MPEQ
MAEP<-sum(abs(x-r3))/nrow(base)
MAEP

'modelo glm'
reg3<-glm(r3~log(prod)+log(l.t)+log(hct),family=Gamma(link="log"), data=base)
hnp(reg3, print.on=T, paint.out=T)
summary(reg3)
pred3<-fitted.values(reg3)
x3<-cbind(c(pred3))
MPEQ3<-sum((x3-r3)^2)/nrow(base)
MPEQ3
MAEP3<-sum(abs(x3-r3))/nrow(base)
MAEP3

'modelo glmm'
reg8<-glmer(r3~log(prod)+log(l.t)+log(hct)+horcat+(1|mês)+(1|unidade),
family=Gamma(link="log"), data=base)
summary(reg8)
res8<-fitted(reg8)
MPEQ8<-sum((r3-res8)^2)/nrow(base)
MPEQ8
MAEP8<-sum(abs(r3-res8))/nrow(base)
MAEP8

d.fun <- function(obj) resid(obj)
s.fun <- function(n, obj) simulate(obj)[,1]
f.fun<-function(data)
```

```
glmer(y.~log(prod)+log(l.t)+log(hct)+horcat+(1|mês)+(1|unidade)-1,
family=Gamma(link="log"), data=data)

hnp(reg8, sim=20, newclass=TRUE, diagfun=d.fun, simfun=s.fun,fitfun=f.fun,
data=base, print.on=T, paint.out=T)
```