

Regressão múltipla e relacionamento entre variáveis para produção de cana-de-açúcar

Mateus Teles Vital Gonçalves¹, Pedro Henrique Silva Ferreira ², Luiz Alexandre Peternelli³

1. Introdução

A cana-de-açúcar (*Saccharum spp.*) é uma importante cultura agrícola cultivada em regiões tropicais e subtropicais do planeta (CHEAVEGATTI-GIANOTTO, 2011). O tamanho total da área cultivada com cana-de-açúcar no Brasil ultrapassa 8 milhões de hectares. Além disso, a produção brasileira de etanol e açúcar para a safra de 2017/2018 foi de 27,76 bilhões de litros e mais de 37 milhões de toneladas por hectare, respectivamente. Esses números colocam o país entre os maiores produtores e exportadores dessas duas commodities em todo o mundo (CONAB, 2018). Consequentemente, o cultivo de cana-de-açúcar tem um impacto significativo na economia brasileira.

O processamento de cana-de-açúcar produz principalmente açúcar e etanol (MATSUOKA, 2009). Além disso, o processamento da cana resulta em uma enorme quantidade de resíduo fibroso, conhecido como bagaço. O bagaço de cana é comumente queimado em usinas para geração de vapor e eletricidade. Além disso, o bagaço também pode ser usado na produção de etanol de segunda geração (BEZERRA, 2016). O desenvolvimento de variedades com alto teor de sacarose tem sido o foco para os melhoristas de cana-de-açúcar. No entanto, uma preocupação crescente com os impactos da atividade humana nas mudanças climáticas e a busca por novas fontes alternativas de energia renovável tem destacado a cana-de-açúcar como uma fonte potencial de biomassa para a produção de energia (SANTCHURN, 2012).

O ponto de partida de um programa de melhoramento genético de cana-de-açúcar é a seleção de genitores, buscando divergência genética e alelos desejáveis. Após os cruzamentos entre os genitores escolhidos, as progênies obtidas precisam ser avaliadas para serem avançadas ou descartadas em futuros ensaios. No entanto, restrições de recursos, como área de plantio, material vegetal (gemas) para propagação, e o grande número de genótipos obtidos após o cruzamento entre genitores durante as fases iniciais de um programa de melhoramento genético de cana-de-açúcar se tornam um dos principais impedimentos nesse processo são as fases iniciais. De fato, atualmente os programas de melhoramento genético de cana-de-açúcar podem levar 15 anos até o lançamento de novas cultivares comerciais. Portanto, ser capaz de classificar e selecionar eficientemente esses genótipos é fundamental.

A avaliação fenotípica em programas de melhoramento de cana-de-açúcar consiste na coleção e computação de um conjunto de características distinguíveis entre os indivíduos que compõem a população em estudo. A determinação de quais características serão avaliadas no processo de seleção é dependente das exigências do setor produtivo. Dentre essas características é possível destacar a tonelada de cana por hectare (TCH). Jackson (2005) ao avaliar ganhos obtidos através do melhoramento genético destacou que quase todo o ganho no rendimento de açúcar foi obtido através do incremento de tonelada de cana por hectare (TCH) ao invés de incrementos em porcentagem

¹ Afiliação. e-mail: mateus.vital@ufv.br

² Afiliação. e-mail: pedrohsferreira29@gmail.com

³ Afiliação. e-mail: peternelli@ufv.br

aparente de sacarose na cana-de-açúcar (PC). Além disso, Da Silveira (2015) com base em um estudo aplicando análise de trilha mostrou que o aumento nos rendimentos de fibra (FIB) ocorre principalmente devido ao maior acúmulo de biomassa (TCH). O objetivo desse trabalho foi ajustar uma regressão múltipla compreendendo variáveis relacionadas ao rendimento de açúcar e de biomassa para produção de cana-de-açúcar, definindo o grau de associação e avaliando os inter-relacionamentos entre essas variáveis.

2. Material e Métodos

2.1. Material vegetal

A população utilizada nesse estudo consistiu de 407 clones de 98 famílias de meios-irmãos derivados de clones elite e variedades comerciais. Os cruzamentos foram realizados na Estação de Serra do Ouro, Murici, Alagoas, Brasil (09 ° 13 'S, 35 ° 50' W, 450 m de altitude). As plântulas obtidas de cada família foram transplantadas em 2010 no Centro de Pesquisa e Melhoramento da Cana-de-Açúcar da Universidade Federal de Viçosa, em Oratórios, Minas Gerais, Brasil (20 ° 25 'S, 42 ° 48' W, 494- m altitude). As mudas foram conduzidas nos ensaios da primeira fase (T1) em 2011 (cana-planta) e 2012 (primeira-soca) (BARBOSA, 2012). A população utilizada neste estudo é derivada de mudas selecionadas em T1 e avançadas para a segunda fase (T2) do programa de melhoramento.

2.2. Avaliação Fenotípica e análise estatística

O ensaio foi instalado em maio de 2016 em um campo experimental localizado no Centro de Pesquisa e Melhoramento da Cana-de-Açúcar, Oratórios, Minas Gerais, Brasil (20 ° 25 'S, 42 ° 48' W, 494 m de altitude) e conduzido em um delineamento de blocos aumentados (FEDERER, 1961). Os clones (não replicados) e três cultivares de referência comuns (replicadas uma vez cada) foram dispostos em 22 blocos aumentados. As cultivares de referência foram RB867515 e RB966928 e RB92579. As parcelas experimentais foram constituídas de linha dupla de 3.5 m x 1,4 m entre fileiras (LEITE, 2006) e avaliadas em soca (após 24 meses de crescimento). As medidas fenotípicas foram obtidas em julho de 2018 (cana-soca). As variáveis mensuradas foram número de colmos (NC), peso médio de colmo (PMC), porcentagem de sólidos solúveis (BRIX), porcentagem de fibra (FIB) e tonelada de cana por hectare (TCH). As variáveis foram computadas seguindo os protocolos da CONSECANA. Os dados foram auto escalados e em seguida submetidos a análise de regressão múltipla com o auxílio do software R.

3. Resultados e Discussão

Nós ajustamos um modelo utilizando a variável TCH como resposta em função das variáveis explicativas NC, PMC, BRIX e FIB. A colinearidade é um potencial problema em análises de regressão múltipla, uma vez que torna difícil a partição de efeitos individuais de variáveis preditoras sobre a variável resposta. A multicolinearidade surge quando existe colinearidade entre três ou mais variáveis preditoras, i.e, as variáveis estão relacionadas entre si. A existência de multicolinearidade foi avaliada pela computação das estimativas dos fatores de inflação de variância (VIF) para as

variáveis PMC, BRIX, NC e FIB sendo obtidos os valores 1.25, 1.03, 1.14 e 1.15 respectivamente, indicando que essas variáveis são pouco correlacionadas.

Tabela I. Estimativas dos coeficientes de regressão múltipla (Coef), erro padrão (EP), estatística t (t) e de p-valor (p) para as variáveis peso médio de colmo (PMC), BRIX, número de colmos (NC), fibra (FIB), e tonelada de cana por hectare (TCH).

| | Coef | EP | t | p |
|-------------------|-------------|-----------|-----------|----------|
| Intercepto | -9.718e-17 | 8.054e-03 | 0.000 | 1.0000 |
| PMC | 8.142e-01 | 9.025e-03 | 90.214 | <2e-16 * |
| BRIX | -7.279e-03 | 8.193e-03 | -0.888 | 0.3748 |
| NC | 8.552e-01 | 8.639e-03 | 8.658e-03 | <2e-16 * |
| FIB | -2.153e-02 | 8.658e-03 | -2.486 | 0.0132 * |

Significância: ‘*’ 0.05

Na tabela I estão ilustradas as estimativas dos coeficientes de regressão múltipla das variáveis regressoras. As variáveis PMC e NC apresentaram estimativas de coeficientes de 0.8142 e 0.8552, respectivamente ($p < 0.05$), enquanto que FIB e BRIX apresentaram coeficientes de -0.0211 e -0.007, respectivamente, sendo que a variável BRIX apresentou ser não significativa ($p > 0.05$). Os resultados obtidos mostram que PMC e NC foram as variáveis mais importantes na determinação da TCH. Além disso, as variáveis FIB e BRIX tiveram pequeno ou nenhum efeito em relação a TCH.

A qualidade com a qual um modelo de regressão múltipla se ajusta aos dados pode ser mensurada por duas estimativas: o erro padrão da média (RSE) e o coeficiente de determinação ajustado (R^2). O RSE é uma medida da falta de ajuste e representa uma estimativa do desvio padrão do erro. Já o R^2 fornece uma medida da proporção de variância da variável resposta que pode ser explicada pelas variáveis preditoras. O modelo obtido apresentou valores de RSE de 0.1794, de R^2 de 0.9678 (Tabela II). Adicionalmente, a estatística F apresentou alto valor, o que sugere a opção de se rejeitar a hipótese de que os coeficientes de regressão múltipla são estatisticamente iguais a zero.

Tabela II. Valores do erro padrão do resíduo (RSE), coeficiente de determinação múltipla ajustado (R^2) e estimativa da estatística-F do modelo de regressão múltipla da variável tonelada de cana por hectare (TCH) sobre as variáveis peso médio de colmo (PMC), BRIX, número de colmos (NC), fibra (FIB), e tonelada de cana por hectare (TCH).

| | |
|-------------------------|--------|
| RSE | 0.1794 |
| R^2 | 0.9678 |
| F | 3724 |

A Tabela III representa a matriz de correlação fenotípica para as variáveis preditoras e a variável resposta. As estimativas dos valores de correlação entre a variável TCH e PMC e NC foram positivas e de alta magnitude, sugerindo que o aumento nas variáveis NC e PMC causaria uma mudança no mesmo sentido na variável TCH. Embora estejamos lidando com dados fenotípicos, quem podem ser confundidos com os efeitos ambientais, superestimando ou subestimando as verdadeiras relações entre essas variáveis os resultados encontrados por Brasileiro (2013) e Da Silveira (2015) aplicando análise de trilha com dados genotípicos sugerem que de fato, peso médio de colmo (PMC) e número de colmos (NC) são os principais determinantes das mudanças produtividade em cana-de-açúcar (TCH) indicando que a seleção de indivíduos de interesse pode ser realizada com base nessas duas variáveis.

Tabela II. Matriz de correlação fenotípica das variáveis peso médio de colmo (PMC), BRIX, número de colmos (NC), fibra (FIB), e tonelada de cana por hectare (TCH).

| | PMC | BRIX | NC | FIB | TCH |
|------|--------|--------|---------|---------|---------|
| PMC | 1.0000 | 0.0542 | -0.3149 | -0.2820 | 0.5505 |
| BRIX | | 1.0000 | 0.0047 | 0.1396 | 0.0379 |
| NC | | | 1.0000 | -0.0702 | 0.6003 |
| FIB | | | | 1.0000 | -0.3122 |
| TCH | | | | | 1.0000 |

Conclusão

Esses resultados corroboram com os resultados obtidos em diversos estudos biométricos com cana-de-açúcar disponíveis na literatura, onde o NC, por ser de fácil mensuração a nível de campo, tem sido recomendado como indicador de clones potencialmente mais produtivos.

Agradecimentos

Capes, CNPq, FAPEMIG, Ridesa.

4. Referências bibliográficas

BARBOSA, M. H. P. *et al.* Genetic improvement of sugar cane for bioenergy: the Brazilian experience in network research with RIDESA. *Crop Breeding and Applied Biotechnology*, v. S2, p. 87–98, 2012.

BEZERRA, T. L.; RAGAUSKAS, A. J. Review A review of sugarcane bagasse for second-generation bioethanol and biopower production. *Biofuels, Bioproducts & Biorefining*, p. 1–14, 2016.

BRASILEIRO, B. P.; PETERNELLI, L. A.; BARBOSA, M. H. P. Consistency of the results of path analysis among sugarcane experiments. *Crop Breeding and Applied Biotechnology*, v. 13, n. 2, p. 113–119, 2013.

CHEAVEGATTI-GIANOTTO, A. *et al.* Sugarcane (*Saccharum X officinarum*): A Reference Study for the Regulation of Genetically Modified Cultivars in Brazil. *Tropical Plant Biology*, v. 4, n. 1, p. 62–89, 2011.

CLAUDIO INÁCIO DA SILVEIRA, L. *et al.* Selection strategy in families of energy cane based on biomass production and quality traits. *Euphytica*, 2015.

CONAB (2018) COMPANHIA NACIONAL DE ABASTECIMENTO. Cana-de- açúcar. *Acompanhamento da safra brasileira de cana-de-açúcar. Terceiro Levantamento - Safra 2018/9, Brasília*, 2018. , p. 71.

FEDERER, W. T. AUGMENTED DESIGNS WITH ONE-WAY ELIMINATION OF HETEROGENEITY. *Biometrics*, v. 17, n. 9, p. 447–473, 1961.

JACKSON, P. A. Breeding for improved sugar content in sugarcane. *Field Crops Research*, v. 92, p. 277–290, 2005.

LEITE, M. S. D. O.; PETERNELLI, L. A.; BARBOSA, M. H. P. Effects of plot size on the estimation of genetic parameters in sugarcane families. *Crop Breeding and Applied Biotechnology*, v. 6, n. 1, p. 40–46, 2006.

MATSUOKA, S.; FERRO, J.; ARRUDA, P. The Brazilian Experience of Sugarcane Ethanol Industry The Brazilian experience of sugarcane ethanol industry. *In Vitro Cellular & Developmental Biology*, v. 45, n. January, p. 372–381, 2009.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

SANTCHURN, D. *et al.* From sugar industry to cane industry: Investigations on multivariate data analysis techniques in the identification of different high biomass sugarcane varieties. *Euphytica*, v. 185, n. 3, p. 543–558, 2012.