

Uso da regressão logística na determinação da influência dos caracteres agronômicos durante a seleção massal em cana-de-açúcar

Bruno Portela Brasileiro¹

Luiz Alexandre Peternelli¹

Lucas Santos Lopes¹

Mateus Teles Vital Gonçalves¹

Gustavo Felipe Ferreira Vieira¹

1 Introdução

Nas fases iniciais dos programas de melhoramento genético da cana-de-açúcar os genótipos não são avaliados com repetição, estão em um único ambiente e a seleção dos melhores genótipos é realizada de forma visual (massal), ou seja, baseada nos fenotípicos dos indivíduos. Os principais critérios de seleção adotados pelos melhoristas durante a seleção massal estão associados com os caracteres componentes da produção, como número de colmos, diâmetro e altura de colmos. Portanto, a seleção ou descarte de um indivíduo depende da combinação e magnitude dos componentes de produção (Zhou et al., 2011). Porém, a incidência de uma determinada doença, a saliência da gema axial, o chochamento fisiológico, a ocorrência de floração, o vigor da planta, teores de fibra e açúcar, assim como outros caracteres de importância agronômica, também podem influenciar a tomada de decisão do melhorista.

A escolha em selecionar ou rejeitar um determinado genótipo é uma resposta de natureza binária. Nesse caso a análise de regressão logística pode ser aplicada para avaliar o efeito ou poder de discriminação de cada um dos caracteres agronômicos usados como critérios de seleção, além de calcular probabilidades que podem ser usadas na predição dos indivíduos a serem selecionados (Agresti, 2007).

O objetivo desse trabalho foi avaliar a importância dos caracteres agronômicos durante o processo de seleção massal em cana-de-açúcar, além de avaliar o potencial do uso da regressão logística na identificação dos melhores genótipos.

¹ DET – UFV. e-mail: brunobiogene@hotmail.com. Agradecimento: FAPEMIG, CAPES, CNPq

2 Material e métodos

Foram avaliadas 7.719 seedlings de 128 famílias de cana de açúcar na fase de teste 1 (T1) e 659 clones durante a fase de teste 2 (T2).

Na fase T1, os caracteres avaliados por planta e utilizados como critérios de seleção foram: número de colmos (NC), diâmetro de colmos (DC), altura de colmos (AC), presença de rachadura (RC), saliência da gema (GE), comprimento do entre-nó (ET), arquitetura foliar (AF) e a presença de colmo tardio (CT).

Para todos os caracteres (exceto para RC e CT), foi atribuído o valor 1 (um) se a planta merecesse ser selecionada com base no carácter e 0 (zero) em caso contrário. Para RC e CT foi atribuído 1 para a presença de rachadura e colmo tardio, respectivamente, e 0 para a planta selecionada com base no carácter. Em especial, para NC foi atribuído 1 para as plantas que apresentaram pelo menos 6 colmos.

Na fase T2, os caracteres avaliados e utilizados como critério de seleção foram: NC, DC, AC, GE, RC, ET, AF, conforme descrito anteriormente, além da presença de joçal (JO), presença de carvão (CV), florescimento (FL), chochamento fisiológico (CF), ocamento (OCA) e vigor da planta (VIG).

Para as características NC, DC, AC, GE, ET, AF, foi atribuído o valor 1 (um) se a planta merecesse ser selecionada com base em cada carácter e 0 (zero) em caso contrário. Para JO, RC, FL, CV, CF e OCA foi atribuído 1 para a presença de joçal, rachadura, florescimento, carvão, chochamento e ocamento e 0 para a ausência dessas características nas plantas. Para VIG foram atribuídos valores discretos de 1 até 5, sendo 1 = menor vigor e 5 = maior vigor.

A seleção massal foi realizada nas fases T1 e T2 por um técnico com 25 anos de experiência no processo de seleção visual do programa de melhoramento genético da cana-de-açúcar. Antes da tomada de decisão sobre os genótipos a serem selecionados foi realizada a avaliação individual de todos os genótipos presentes no experimento a partir dos critérios de seleção supracitados.

A regressão logística foi aplicada visando determinar a importância dos caracteres agronômicos no processo de seleção massal realizado nas fases T1 e T2. Na predição do modelo, foi adotado o ponto de corte de 0,5, ou seja, foram selecionados os indivíduos i ($i = 1, 2, \dots, n$) que apresentaram uma probabilidade de seleção acima de 0,5. A probabilidade de seleção foi estimada usando o seguinte modelo de regressão logística:

$$\hat{p}_i = \exp\left(\sum_{j=0}^k \beta_j x_{ij}\right) / \left(1 + \exp\left(\sum_{j=0}^k \beta_j x_{ij}\right)\right), \text{ onde: } \beta_j = \text{coeficientes de regressão};$$

x_{ij} = valores das variáveis explicativas (critérios de seleção) j em cada indivíduo i ; k = número de variáveis. As taxas de erro aparente dos modelos de regressão foram calculadas com base no número de indivíduos classificados incorretamente pelo modelo logístico em relação ao número total de plantas avaliadas no experimento.

As análises de regressão logística foram realizadas no software R (R DEVELOPMENT CORE TEAM, 2013).

3 Resultados e discussões

Os modelos de regressão logística ajustados para as fases iniciais de seleção foram eficientes na predição dos indivíduos a serem selecionados. No T1, 153 indivíduos apresentaram probabilidade de seleção acima de 0,5. O modelo para a seleção no T1 apresentou capacidade de acerto próxima de 100%. Ao considerar a capacidade preditiva da condição seleção, esse percentual foi de aproximadamente 87% (Tabela 1).

Tabela 1 - Classificações pelos modelos de regressão logística usando como resposta a seleção massal aplicada no T1 e no T2.

| Seleção massal – T1 | Modelo de regressão completo | | Total | TEA * |
|---------------------|------------------------------|---------------|-------|-------|
| | Selecionar (1) | Descartar (0) | | |
| Selecionar (1) | 133 | 41 | 174 | 0,8% |
| Descartar (0) | 20 | 7.525 | 7.545 | |
| Total | 153 | 7.566 | 7.719 | |
| Seleção massal – T2 | Modelo de regressão reduzido | | Total | TEA * |
| | Selecionar (1) | Descartar (0) | | |
| Selecionar (1) | 79 | 34 | 113 | 5,1% |
| Descartar (0) | 0 | 546 | 546 | |
| Total | 79 | 580 | 659 | |

*TEA = Taxa de erro aparente em percentagem.

No T2, 79 indivíduos atingiram o limiar de seleção. A capacidade de acerto do modelo ajustado para a seleção massal no T2 foi de aproximadamente 95% e com uma capacidade preditiva da condição seleção de 100% (Tabela 1).

Todos os caracteres apresentaram efeitos significativos no modelo de regressão logístico ajustado para a seleção massal aplicada na fase T1. De acordo com os valores das

razões de risco (*odds ratio*) NC, GE, ET, DC, AC, AF, CT e RC foram, nessa ordem, os caracteres mais importantes no processo de seleção massal do T1. Conforme as razões de risco, uma planta com mais de 5 colmos tem 1.665 vezes mais chance de ser selecionada do que uma planta com até 5 colmos. Se a planta possuir uma gema pouco saliente, a chance de ser selecionada é 408 vezes maior do que uma planta de gema grande. Uma planta com entre-nó uniforme tem 90 vezes mais chance de ser selecionada em relação a plantas com entre-nó desuniforme. Se a planta apresentar DC considerado bom, a chance de seleção será 85 vezes maior do que uma planta de DC fino. Finalmente, para AC a chance de uma planta ser selecionada ao possuir uma boa altura é 77 vezes maior do que uma planta de baixa estatura.

Na fase T2, de acordo com os resultados obtidos, foi possível verificar a ausência de efeito significativo dos caracteres NC, RC, ET, AF, FL, JO e CV no modelo completo de regressão logística. Um novo modelo foi ajustado (modelo reduzido) e de acordo com os valores das razões de risco os caracteres: VIG, DC, AC, GE, CF e OCA, nessa ordem, foram os mais importantes no processo de seleção massal realizado na fase T2. No T2, a chance do indivíduo ser selecionado aumenta em 47 vezes a cada unidade a mais na escala de nota para o VIG. Uma planta com DC considerado grosso tem 9 vezes mais chance de ser selecionada do que uma planta de diâmetro fino. Uma planta considerada alta tem quase 5 vezes mais chance de ser selecionada do que uma planta de baixa estatura. As plantas com gema pequena apresentam quase 5 vezes mais chances de serem selecionadas do que as plantas de gema grande e as plantas que não apresentam CF possuem quase 3 vezes mais chances de serem selecionadas do que plantas com chochamento.

A importância dos caracteres agrônômicos sofreu influência das peculiaridades inerentes as fases T1 e T2. O NC, por exemplo, foi a variável de maior importância na seleção massal aplicada no T1 e não apresentou efeito significativo no modelo ajustado para a seleção praticada na fase T2. Isso se explica pelo fato da população da fase T2 apresentar menor variabilidade para o carácter número de colmos. A seleção de genótipos com elevado número de colmos na fase T1 colabora para a menor variabilidade para NC após o primeiro ciclo de seleção, conforme pode ser observado na Figura 1.

O fato dos genótipos no T2 serem representados em parcelas de 4 metros lineares também acaba gerando maior dificuldade para a classificação dos indivíduos quanto as suas capacidades de perfilhamento. Devido a essa dificuldade e a maior facilidade de classificação dos clones para DC, esse carácter ganhou maior importância e passou a ter mais influência na decisão do melhorista.

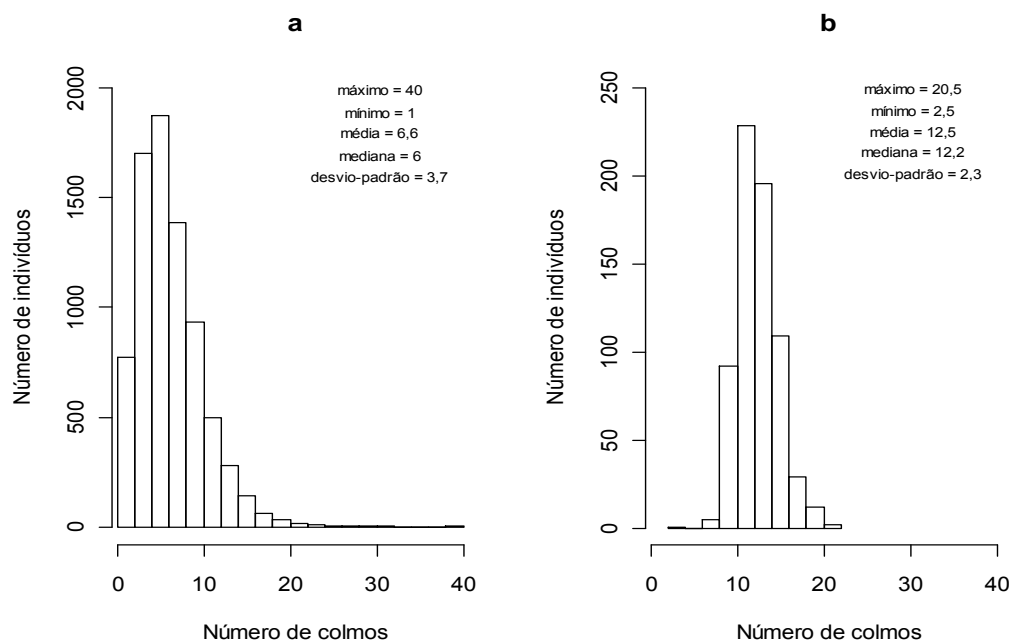


Figura 1 - Distribuições do número de colmos por metro linear nos genótipos avaliados nas populações das fases T1 (a) e T2 (b) do programa de melhoramento genético da cana-de-açúcar da UFV (PMGCA/UFV).

4 Conclusões

Com base na análise de regressão logística, foi possível avaliar a importância dos caracteres agrônômicos no processo de seleção massal e prever os indivíduos selecionáveis nas fases T1 e T2.

Os caracteres número de colmos, saliência da gema e comprimento do entre-nó foram os de maior importância no processo de seleção massal realizado no T1. Vigor da planta, diâmetro de colmo e altura de colmo foram os mais importantes na seleção do T2.

5 Bibliografia

- [1] AGRESTI A. **An introduction to categorical data analysis**. 2^a ed. Hoboken: John Wiley & Sons, 2007.
- [2] R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Áustria, 2013. Disponível em <http://www.R-project.org>.
- [3] ZHOU, M. M.; COLLINS, A.; KIMBENG, C. A.; TEW, T. L.; GRAVOIS, K. A.; PONTIF, M. J. Artificial neural network models as a decision support tool for selection in sugarcane: a case study using seedling populations. **Crop Science**. v, 51 p. 21-31, 2011.