

# ANÁLISE DE CORRESPONDÊNCIA: UMA APLICAÇÃO NO MELHORAMENTO GENÉTICO DA CANA-DE-AÇÚCAR

Yury Vasconcellos da Silva<sup>1</sup>, Carlos Henrique Osório Silva<sup>2</sup>, Gemma Lucia Duboc de Araújo<sup>3</sup>, Luiz Alexandre Peternelli<sup>2</sup>, Gerson Rodrigues dos Santos<sup>2</sup>, Adair José Regazzi<sup>4</sup>

**Resumo:** *Aplicou-se a metodologia de análise de correspondência simples sobre um conjunto de dados proveniente de um experimento na área de melhoramento genético da cana-de-açúcar. Concluiu-se que o número de colmos é um fator determinante para antecipar a classificação de peso da touceira.*

**Palavras-chave:** *Variáveis categorizadas, análise multivariada, tabela de contingência.*

## 1 Introdução

A produção de cana-de-açúcar (*Saccharum spp.*) na safra de 2011/2012, estimada pela Conab (2011) foi de 571 milhões de toneladas e de acordo com Bertelli (2012) esta produção necessitará ser dobrada até 2020, para atender à fabricação de açúcar para o mercado doméstico e exportações, a produção de etanol, bem como na geração de 13 mil megawatts, o que equivale à produção da hidrelétrica de Itaipu.

Para satisfazer esta demanda de mercado fazem-se necessários estudos de melhoramento genético envolvendo a cana-de-açúcar visando obter aumento de produtividade e melhoria na qualidade da matéria prima para fabricação de álcool e de açúcar. Uma alternativa para agilizar a coleta de dados no campo é a categorização das variáveis componentes da produção, pois desta forma não é necessária a avaliação individual de plantas no experimento. Apesar de ser um processo menos eficiente de seleção, a categorização permite a avaliação de um maior número de famílias de cultivares por vez. (SOUZA; PETERNELLI; BERNARDES, 2012, p. 2). Assim, para análise desses dados categóricos foi utilizada a análise de correspondência sobre as variáveis das características físicas da cana-de-açúcar e peso da touceira.

A análise de correspondência (AC) é uma técnica estatística multivariada que permite realizar a análise simultânea de diferentes variáveis categóricas. Estas variáveis categorizadas podem possuir níveis de mensuração nominal e ordinal. A AC é essencialmente um método exploratório e descritivo que permite uma análise gráfica das inter-relações ou correspondências entre as variáveis, oriunda da redução de dimensionalidade. Tal gráfico é denominado mapa perceptual. Neles, é possível observar se as variáveis de interesse distanciam-se do pressuposto de independência convergindo para possíveis associações existentes entre elas, sendo possível ainda verificar como se dá esta associação, uma vez que níveis das variáveis de linha e de coluna assumem posições nos gráficos de acordo com suas associações ou com suas similaridades (GRE-ENACRE, 1984). Ou seja, o mapa perceptual gerado na análise de correspondência demonstra,

<sup>1</sup>Centro de Ensino Superior de Juiz de Fora, yury@numberonejf.com.br

<sup>2</sup>DET/UFV - Depto de Estatística da Universidade Federal de Viçosa

<sup>3</sup>FDV - Faculdade de Viçosa

<sup>4</sup>Depto de Ciências Florestais e da Madeira da UFES (ES)

de maneira simples, a existência de dependência entre as variáveis, bem como quais categorias estão relacionadas.

## 2 Objetivos

Definir, com a aplicação da AC, qual característica física da cana pode ser utilizada como indicadora do melhor peso da touceira.

## 3 Metodologia

### 3.1 Análise de correspondência

A análise de correspondência é obtida a partir da decomposição em valores singulares generalizada da matriz de resíduo obtida da diferença entre as frequências relativas observadas e esperadas de uma tabela de contingência.

A matriz  $P$  de frequências relativas observadas é denominada matriz de correspondência e é obtida dividindo-se cada termo  $n_{ij}$  da matriz  $N_{I \times J}$  de frequências absolutas pelo total geral  $n_{(..)} = \sum_{i=1}^I \sum_{j=1}^J n_{ij}$ .

As massas de linha e coluna ( $m_r$  e  $m_c$ ) são probabilidades marginais obtidas a partir da matriz  $P$ , em que  $p_{i.} = \sum_{j=1}^J p_{ij} = \frac{n_{i.}}{n_{..}}$ ;  $\sum_{i=1}^I p_{i.} = 1$ ;  $p_{.j} = \sum_{i=1}^I p_{ij} = \frac{n_{.j}}{n_{..}}$  e  $\sum_{j=1}^J p_{.j} = 1$ . Estes vetores de massa de linha e coluna são obtidos a partir das seguintes operações matriciais:  $m_r = P_{I \times J} \cdot 1_J$  e  $m_c = (P^t)_{J \times I} \cdot 1_I$ . Assim a matriz contendo as frequências relativas esperadas para situação de independência entre as linhas e colunas de uma tabela de contingência, pode ser obtido por meio da matriz  $m_r m_c^t$ . As coordenadas principais e padronizadas conforme apresentado abaixo, são obtidas pela decomposição em valores singulares generalizada da matriz de resíduos em  $(P - m_r m_c^t) = AD_\sigma B^t$ , em que  $A = D_r^{-\frac{1}{2}} U$  e  $B = D_c^{-\frac{1}{2}} V$ . Vale lembrar que calcular  $F$  e  $G$  de forma independente pode levar a “erros” de sinais, como assegura Greenacre (1984, p. 93), neste sentido,  $G$  será calculado a partir de  $F$ .

- Coordenadas principais por linha:  $F = D_r^{-1} A D_\sigma$ ;
- Coordenadas principais por coluna:  $G = C F D_\sigma^{-1}$ .

Além das coordenadas principais, há também as coordenadas padronizadas que associadas as coordenadas principais possibilitam a análise entre linhas e colunas utilizando o mesmo mapa perceptual.

- Coordenadas padronizadas para linhas:  $\Phi = F D_\sigma^{-1}$ ;
- Coordenadas padronizadas para colunas:  $\Gamma = G D_\sigma^{-1}$ .

#### Tipos de mapas perceptuais

A conjugação dessas coordenadas permite obter os seguintes mapas perceptuais : (i) o **mapa perceptual simétrico** que utiliza as coordenadas principais por linha e coluna, (ii) o **mapa perceptual assimétrico principal por linha**, que utiliza a coordenada principal por linha e padronizada por coluna e (iii) o **mapa perceptual assimétrico principal por coluna**, que utiliza a coordenada principal por coluna e padronizada por linha.

Detalhes geométricos, com os princípios matemáticos e estatísticos da AC abordados passo a passo com um exemplo ilustrativo, são apresentados na dissertação de mestrado: ”Análise de Correspondência: uma abordagem geométrica”, de SILVA (2012).

### 3.2 Dados da cana-de-açúcar utilizados na AC

Foi utilizado um conjunto de dados proveniente de um experimento do Centro de Pesquisa e Melhoramento de Cana-de-Açúcar da Universidade Federal de Viçosa (CECA - UFV). As variáveis utilizadas foram o número de colmos por planta (NC), a altura média dos colmos (AC), o diâmetro do 5º internódio contado a partir da base do colmo (DC) e o peso da touceira.

Com base em critérios de prática de campo, os dados foram categorizados em pouco (pó), médio (me) e muito (um) para variável NC. Baixo (ba) e alto (al) para variável AC e fino (fin), médio (med), grosso (gro) e muito grosso (mgr) para variável DC. (SOUZA; PETERNELLI; BERNARDES, 2012, p. 3). Já o peso das touceiras (em TCH - tonelada de cana por hectare) foi categorizado para este trabalho utilizando como ponto de corte os percentis 20, 40, 60 e 80%. Portanto, a tabela de contingência submetida à AC foi formada por 24 categorias nas colunas (3×2×4) e 5 categorias de peso nas linhas.

As análises foram implementadas no software R versão 2.14.0 (R DEVELOPMENT CORE TEAM, 2011) com o pacote ca (GREENACRE, NENADIC, 2010), com mapa perceptual assimétrico principal por linha.

## 4 Resultados

Para facilitar o entendimento da interpretação da AC, tanto em relação às medidas descritivas dos perfis, quanto para a análise visual dos mapas perceptuais, buscou-se uma nomenclatura para as combinações entre as categorias das variáveis, utilizando as abreviações apresentadas na Tabela 1, na seguinte ordem: número de colmos, altura e diâmetro. Desta forma, mealfin quer dizer uma touceira com as características de número médio de colmos, altos e finos; e mubamgr, muitos colmos, baixos e muito grossos. Os pesos foram classificados em p\_pess (péssimo peso), p\_ruim (peso ruim), p\_reg (peso regular), p\_bom (bom peso) e p\_exc (excelente peso).

Com o objetivo de relacionar perfis de linha e coluna no mesmo mapa perceptual faz-se necessário o uso do mapa assimétrico. Para relacionar cada perfil de linha com o vértice de produção que mais se aproxima usa-se o mapa perceptual principal por linha.

A Figura 1 ilustra quais características da cana-de-açúcar estão associadas a qual peso da touceira. Assim, por exemplo, pobamed e pobafin estão associados a peso de touceira péssimo.

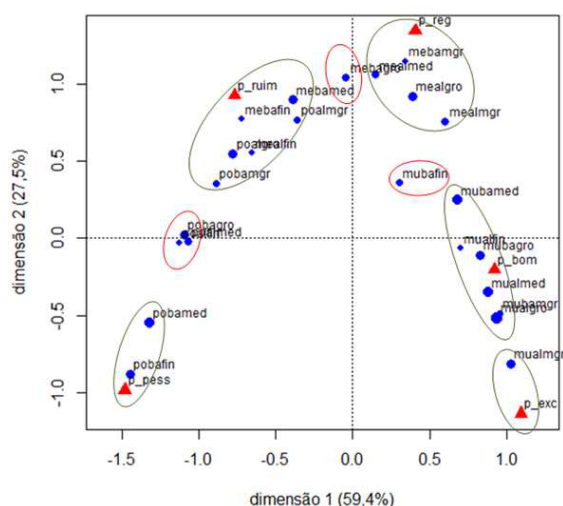


Figura 1: Mapa perceptual assimétrico principal por linha das características da cana-de-açúcar classificando-as de acordo com o peso da touceira.

Outra importante análise revelada pela Figura 2, é a associação do número de colmos com o peso da touceira. Assim percebe-se que o número de colmos da touceira é um fator determinante para que se possa avaliar seu peso. Touceiras com muitos colmos tendem a ter pesos maiores e touceiras com poucos colmos tendem a ter pesos menores, independentes da sua altura e diâmetro.

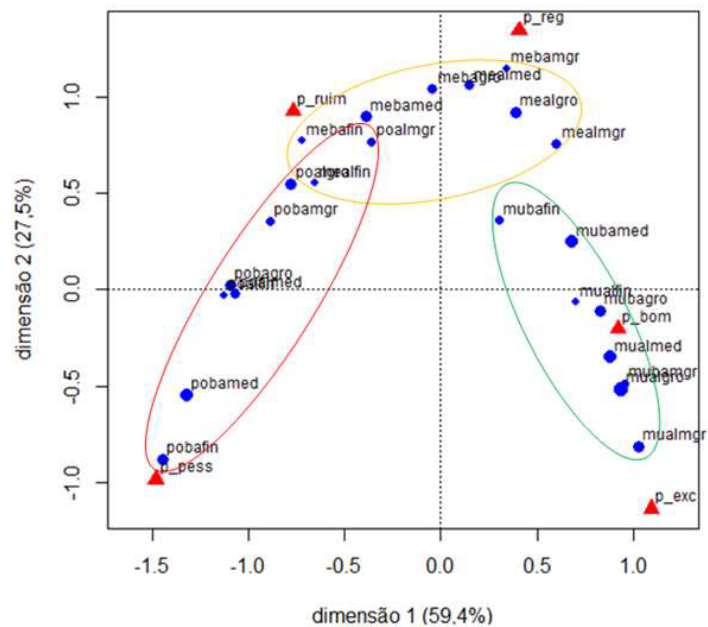


Figura 2: Mapa perceptual assimétrico principal por linha agrupando os perfis de linha de acordo com o número de colmos e, conseqüentemente, pelo peso de touceira.

## 5 Conclusões

Ao aplicar análise de correspondência aos dados da cana-de-açúcar verificou-se que o número de colmos é um fator determinante para antecipar a categoria de peso da touceira. Essa informação é de fundamental importância, uma vez que pode facilitar os aspectos operacionais na seleção de famílias para o melhoramento genético.

## Referências

- [1] GREENACRE, M. J., Theory and applications of correspondence analysis. Orlando: Academic Press, 1984.
- [2] GREENACRE, M. J., NEMADIC, O., Simple, multiple and joint correspondence analysis. Version 0.33, 2010. Disponível em: <<http://www.carne-n.org/>>.
- [3] SILVA, Y. V., Análise de Correspondência: uma abordagem geométrica. Dissertação (Mestrado em Estatística e Biometria) - DET - Departamento de Estatística, UFV - Universidade Federal de Viçosa, 2012.
- [4] SOUZA, V. B. C, PETERNELLI, L. A., BERNARDES, D. P., Seleção de famílias de cana-de-açúcar com base na categorização de variáveis. Resumo estendido. 2012