

Seleção de descritores em cana-de-açúcar por meio de componentes principais

Selection of descriptors in cane sugar by means of principal components

Kleyton Danilo da Silva Costa; Islan Diego Espindula de Carvalho; Lucas da Silva Santos; Paulo Ricardo dos Santos; José Wilson da Silva.

RESUMO - Foram avaliados oito caracteres agrônômicos com o objetivo de selecionar os melhores para caracterizar os clones. Utilizou-se a análise de componentes principais objetivando o descarte de descritores considerados redundantes ou não discriminantes. Este método permitiu redução de quatro dos oito descritores agrônômicos, do conjunto inicialmente considerado. Foi possível selecionar, por intermédio desta técnica, número de colmos por touceira, altura do colmo, diâmetro médio do colmo e número de entrenós por colmo, considerados mais importantes na descrição da variabilidade presente na coleção de acessos dos clones estudados.

Palavras-chave: Variabilidade, Genótipos Superiores, Clones, Divergência Genética.

ABSTRACT - Eight agronomics characters with the objective had been evaluated to select the best ones to characterize clones. It was used analysis of main components objectifying the discarding of redundant or not discriminantes considered describers. This method allowed reduction of four of the eight agronomics describers, of the set initially considered. It was possible to select, for intermediary of this technique, number of colms for shoots, height of colms, average diameter of colm and number of entrenós for colms, considered more important in the description of the present variability in the collection of studied accesses of clones.

Key words: Variability, Superior genotypes, Clones, Genetic Divergence.

INTRODUÇÃO

Os programas de melhoramento genético de cana-de-açúcar avaliam em cada etapa de seleção diversos caracteres morfológicos, fisiológicos e industriais, sendo a seleção de variedades fundamentada em um conjunto de características. Esta prática, muitas vezes, pode gerar dificuldades principalmente quando alguns caracteres são difíceis de avaliar. Nesse aspecto, cada programa tem estabelecido as melhores alternativas de conduzir o processo seletivo. Em muitas situações, os pesquisadores e melhoristas de cana-de-açúcar tendem a avaliar um grande número de características, gerando acréscimo considerável de trabalho. Quando o número de características torna-se elevado, é possível que muitas características contribuam pouco para a discriminação dos indivíduos avaliados, representando, conseqüentemente, aumento no trabalho de caracterização, sem melhoria na precisão, e tornando mais complexa a análise e interpretação dos dados.

Assim, poderiam ser eliminadas aquelas características redundantes, o que reduziria o tempo e os custos com experimentos, sendo a análise por componentes principais uma alternativa, por possibilitar a identificação dos caracteres mais e menos representativos na explicação da variância total, bem como para a análise de divergência genética. Sua aplicação consiste em

transformar um conjunto original de variáveis em outro de dimensão equivalente, mas com propriedades importantes, que sejam de interesse em estudos de melhoramento.

Metodologicamente, componente principal consiste em uma técnica de análise multivariada que se encontra melhor descrita por Morrison (1981), Mardia et al. (1997) e Johnson e Wichern (1982). Esta análise consiste em transformar um conjunto de p variáveis originais X_1, X_2, \dots, X_p , pertencentes a n indivíduos ou populações, em um novo conjunto de variáveis, Y_1, Y_2, \dots, Y_p de dimensão equivalente, chamados componentes principais. Cada componente principal é uma combinação linear das variáveis originais, construídos de maneira a explicar o máximo da variabilidade total dessas variáveis originais e não-correlacionado entre si.

Morrison (1981) define o primeiro componente principal (Y_1) de um conjunto de p variáveis, X_1, X_2, \dots, X_p , contidas no vetor $X' = (X_1, X_2, \dots, X_p)$ como a combinação linear: $Y_1 = b_{11}X_1 + b_{21}X_2 + \dots + b_{p1}X_p = \mathbf{b}'_1\mathbf{X}$ cujos coeficientes b_{i1} são elementos do vetor característico \mathbf{b}_1 , associado à maior raiz característica da matriz de covariância amostral, \mathbf{S} , das variáveis X_i 's. O segundo componente principal (Y_2) estará associado a segunda maior raiz característica e, assim sucessivamente, até que toda a variância tenha sido explicada. Os autovalores (ou raízes características) ordenados são as

Recebido em 10 01 2013 aceito em 29 03 2013

Engenheiro Agrônomo e mestrando em Agronomia (Produção Vegetal) pelo CECA/UFAL. kleytondanilo@hotmail.com
Graduando em Agronomia pelo CECA/UFAL.

Engenheiro Agrônomo e Doutorando em Melhoramento Genético de Plantas pela UNESP.

Engenheiro Agrônomo e Mestrando em Melhoramento Genético de Plantas pela UFRPE.

Professor Doutor da UFAL. jwsamaral@hotmail.com

variâncias amostrais dos componentes principais. O primeiro componente também é definido como o de maior importância, uma vez que retém a maior parte da variação total encontrada nos dados originais.

A técnica envolve a matriz de covariância amostral **S** ou a matriz de correlação **R**, sendo que a maioria das aplicações tem envolvido esta última. Tendo em vista o grande número de variáveis medidas em unidades diferentes, foi necessária a padronização destas variáveis X_j ($j = 1, 2, \dots, p$). Nesse caso, a estrutura de dependência de X_j foi determinada pela matriz de correlação **R**.

A viabilidade de utilização dos componentes principais em estudos de descarte de características depende da possibilidade de resumir o conjunto de variáveis originais em poucos componentes, o que significa ter uma boa aproximação do comportamento dos indivíduos, (CRUZ e REGAZZI, 1997). Segundo Daher et al., (1997), esta técnica apresenta ainda a vantagem adicional de descartar os caracteres menos discriminantes, por já estarem correlacionados com outras variáveis.

A redução no número de variáveis é útil quando algumas dificultam a análise dos dados e fornecem pouca ou nenhuma informação adicional. Com base no princípio de que a importância ou variância dos componentes principais decresce do primeiro para o último, tem-se que os últimos componentes explicam uma fração muito pequena da variância total. Então, a variável de maior coeficiente no componente de menor autovalor deve ser a menos importante para explicar a variância total e, portanto, passível de descarte (PEREIRA, 1989). Jolliffe (1972 e 1973), afirmou que quando se dispõe de grande número de descritores, é possível que muitos deles sejam redundantes, tornando-se útil a sua eliminação, porque, além de pouco informativos, ocorre acréscimo no trabalho de avaliação e não apresentam informação adicional.

Cury (1993) realizou o descarte das variáveis através da técnica dos componentes principais, descartando a variável com maior valor absoluto do componente com menor autovalor. Dias (1994) descartou as variáveis que apresentaram os maiores elementos dos autovetores nos últimos componentes. Barbosa et al. (2002) estudaram o desdobramento dos coeficientes de correlação envolvendo a produção de colmos de cana-de-açúcar por hectare. Seus componentes indicaram que o número médio de colmos e o peso de colmos foram as principais causas da correlação com a produção de colmos por hectare. Barreto (2004) avaliando divergência em cana-de-açúcar, concluiu que as variáveis que menos contribuíram para o estudo da divergência foram: número de flores por touceira e altura do colmo.

Objetivou-se neste trabalho reduzir a dimensionalidade do conjunto original de variáveis com a menor perda de informação possível, eliminando as informações redundantes em decorrência da correlação entre variáveis, e descartar as que menos contribuem para explicar a variação total, evitando que sejam avaliadas em experimentos futuros.

MATERIAL E MÉTODOS

Os dados foram obtidos de experimento realizado no campo experimental da Usina Santa Clotilde, município de Rio Largo, Alagoas, com o apoio do Programa de Melhoramento Genético da Cana-de-Açúcar (PMGCA), no ano de 2011. O município está situado a uma latitude de 9° 27' S, longitude de 35° 27' W e uma altitude de 127,0 m, com temperatura máxima de 29°C e mínima de 21°C, com pluviosidade média anual de 1.267,7 mm. O delineamento experimental utilizado foi em blocos casualizados com 12 tratamentos e quatro repetições, sendo as parcelas constituídas de cinco sulcos de 10,0 m de comprimento, com espaçamento de 1,0 m entre sulcos. O solo era de topografia plana, classificado como Latossolo Podzólico Vermelho Amarelo. Os clones avaliados foram: RB72454, SP791011, RB92579, RB93509, RB931530, RB931003, RB931011, RB931611, RB9364, RB93503, RB951541 e RB961.

O preparo do solo para a instalação do experimento foi realizado por meio de uma aração e duas gradagens. Foram aplicados fertilizantes NPK, com base na recomendação da análise do solo. Também foi efetuado o controle de pragas, doenças e plantas daninhas, bem como a irrigação, objetivando garantir o bom desenvolvimento da cultura.

Avaliou-se o experimento por meio das seguintes características: Número de Colmos por Parcela (NCP), característica obtida através da contagem dos colmos formados e desenvolvidos nas fileiras da área útil de cada parcela, correspondente às três fileiras centrais; Número de Touceiras por Parcela (NTP), característica obtida pela contagem do número de touceiras formadas e definidas, por um conjunto de colmos do mesmo perfilho de origem; Número de Colmos por Touceira (NCT), obtida pela relação (NCP) / (NTP): Altura do Colmo (AC), em cm, medida da base do colmo até o último entrenó, com auxílio de uma trena de 3,0 m de comprimento; Diâmetro Médio do Colmo (DMC), obtida do centro do entrenó do colmo localizada a uma altura de 40 cm do solo, utilizando-se um paquímetro; Número de Entrenós por Colmo (NEC), obtido pela contagem do número de entrenós formados iniciando a partir da base ao ápice, descartando-se o palmito; Número de Folhas (NF), pela contagem de folhas totalmente desenvolvidas e apresentando aurícula visível e 80% da superfície da folha verde; Ângulo de Inserção da Folha (AIF) mediu-se o em graus, obtida nos quatro colmos tomando como folha padrão +3, seguindo a classificação de Dillewinj (1952), tomando-se o início a partir do dew-lap como zero grau. Utilizou-se nesta avaliação um transferidor. Todas estas variáveis agrônômicas foram avaliadas com nove meses após o plantio.

Para o descarte de variáveis, foi estabelecida a de menor autovalor (menor variância), que deve ser menos importante para explicar a variância total e, portanto,

passível de descarte (REGAZZI, 2002). As variáveis de maiores pesos nos últimos autovetores foram consideradas de menor importância. Entretanto, as variáveis de maiores pesos nos primeiros autovetores, são os de maior importância.

Todas as análises estatísticas foram realizadas no programa Genes, seguindo metodologia proposta por Cruz (2007).

RESULTADOS E DISCUSSÃO

Pela análise de variância multivariada houve diferenças significativas entre populações ($p < 0,01$) na maioria dos caracteres avaliados, exceto em relação à NTP, NCP e AI, indicando que, a princípio, os clones são divergentes (Tabela 1).

Tabela 1 - Resumo da Análise de Variância de Oito Caracteres Avaliados em 12 Clones de Cana-de-açúcar, Rio-Largo/AL, 2011.

GL	Quadrados Médios							
	DMC	AC	NTP	NCP	NE	NF	AIF	NCT
3	5,523	1220,562	626,062	13,254	4,854	0,243	39,324	380,687
11	199,633**	10099,062**	861,562 ^{ns}	40,623**	8,248**	3,657**	58,698 ^{ns}	2604,896 ^{ns}
33	105,823	6905,187	2206,687	50,603	1,460	0,546	28,320	412,596
	27,115	135,938	44,688	5,351	15,604	5,645	21,677	227,187
	6,60	10,64	18,29	23,14	7,74	13,08	24,55	8,94

** Significativo a 1% de probabilidade pelo teste F; ^{ns} = Não significativo a 5% de probabilidade DMC: Diâmetro médio do colmo (cm); AC: Altura do colmo (cm); NTP: Número de touceiras por parcela; NCT: Número de colmos por touceiras; NE: Número de entrenós; NF: Número de folhas; AIF: Ângulo de inserção da folha; NCP: Número de folhas.

A análise de componentes principais permitiu a redução de oito descritores para quatro. O componente 1, explicou 32,95% da variância total, seguido dos componentes 2 e 3 com respectivos valores de 29,45% e 20,59% da variância total. No componente 1, observa-se que os descritores de maior peso foram ângulo de inserção (0,47) e altura do colmo (0,46). No componente 2 os maiores descritores em valor absoluto foram números de colmos por touceira (0,56) e diâmetro médio do colmo (0,52). No componente 3 os descritores de maior peso foram ângulo de inserção (0,48) e altura do colmo (0,46).

Pelo critério sugerido por Cruz e Regazzi (1997), conclui-se que as variáveis de menor importância, no

presente estudo foram: número de touceiras por parcela (NTP), número de folhas (NF), Número de colmos por parcela (NCP) e ângulo de inserção da folha (AI) Estas quatro características, de um total de oito, são dispensáveis ou redundantes, por apresentarem alta correlação com outras mais importantes. Adotando este critério, Choer e Silva (2000) identificaram 21 caracteres redundantes na divergência genética de *Cucurbita* ssp. Em determinados casos sem uso de delineamentos adequados os caracteres são tomados sem a existência de um estudo preciso da contribuição de cada caráter e suas respostas ao ambiente.

Tabela 2 - Variância de cada componente principal e sua importância em relação à variação total, Rio-Largo/AL, 2011.

Componente Principal	λ_j	$\lambda_j(\%)$	DMC	NE	AC	NF	AIF	NCP	NCT	NTP
Y ₁	2,63	32,95	-0,31	0,15	0,46	-0,40	0,47	0,22	-0,26	0,23
Y ₂	2,35	62,40	0,52	0,07	0,01	0,43	0,00	-0,06	-0,56	0,12
Y ₃	1,64	82,99	-0,17	0,05	-0,46	-0,2	-0,48	0,32	-0,15	0,25
Y ₄	1,25	98,74	0,01	-0,62	-0,19	-0,05	0,03	0,73	-0,11	0,43
Y ₅	0,04	99,32	0,46	0,54	0,18	0,04	0,20	0,44	0,15	0,29
Y ₆	0,02	99,59	-0,41	0,04	0,04	-0,03	-0,38	0,29	0,06	0,26
Y ₇	0,01	99,83	0,41	-0,11	0,32	-0,48	-0,31	0,02	0,54	0,23
Y ₈	0,01	100,0	-0,14	0,11	-0,39	-0,39	0,49	0,11	0,49	0,56

DMC: Diâmetro médio do colmo (cm); NE: Número de entrenós; AC: Altura do colmo (cm); NF: Número de folhas; AIF: Ângulo de inserção da folha; NCP: Número de colmos por parcela; NCT: Número colmos por touceiras. NTP: Número de touceiras por parcela.

Na análise da divergência genética, os quatro primeiros componentes principais explicaram 62,40% da variação (Tabela 2). Segundo Cruz e Regazzi (1997), nos casos em que este limite não é atingido nos dois primeiros componentes, a análise é complementada com a dispersão gráfica em relação ao terceiro e quarto componentes. No entanto em cana-de-açúcar, o interesse nessa avaliação e em outros trabalhos, obtendo resultados favoráveis com variação total abaixo da recomendada de 80% nos dois primeiros componentes.

Este fato evidencia a importância da utilização de técnicas multivariadas (componentes principais) na identificação de caracteres, que realmente devem ser avaliados com base em um estudo prévio da sua contribuição para a variabilidade (PEREIRA, 1989). O interesse nessa avaliação reside na possibilidade de se descartarem caracteres que contribuam pouco para a discriminação dos genótipos avaliados, reduzindo, dessa forma, mão-de-obra, tempo e custos despendidos, na experimentação agrícola (CRUZ e REGAZZI, 1997).

CONCLUSÕES

Os três primeiros componentes foram suficientes para explicar 82,99% da variação total existente, por meio da técnica de componentes principais;

Dos oitos descritores agrônômicos analisados quatro mostraram-se importantes para a descrição dos genótipos.

BIBLIOGRAFIA

BARBOSA, M. H. P.; BASTOS, I. P.; SILVEIRA, L. C. I.; OLIVEIRA, M. W. Análise de causa e efeito para a produção de colmos e seus componentes na seleção de famílias de cana-de-açúcar. **8º Congresso Nacional da STAB**. p. 366 – 370, 2002.

BARRETO, E. J. S. Avaliação de divergência genética de genótipos de cana-de-açúcar (*Saccharum spp.*) da estação de floração e cruzamento Serra - do- Ouro. **Dissertação (Mestrado em Produção Vegetal – UFAL)**, Maceió – AL. 48p, 2004.

CHOER, E.; SILVA, J. B. Avaliação da divergência genética entre acessos de *Cucurbita spp.* através de análise multivariada. **Agropecuária de Clima Temperado**, Pelotas, v. 3, n. 2, p. 213-219, 2000.

CRUZ, C.D. **Programa Genes-Applicativo computacional em Genética e estatística**. Versão 2007. Disponível em www.ufv.br/dbg/gene/genes.htm, acesso em 10 de jan. de 2008.

CRUZ, C. D.; REGAZZI, A. J. Modelos biométricos aplicados ao melhoramento genético. Viçosa, MG: **Imprensa Universitária**, 390p, 1997.

CURY, R. Dinâmica evolutiva e caracterização de germoplasma de mandioca (*Manihot esculenta* Crantz) na agricultura autóctone do sul do Estado de São Paulo. Piracicaba, SP: ESALQ. **Dissertação (Mestrado)** - Escola Superior de Agricultura “Luiz de Queiroz” /Universidade de São Paulo, 103p, 1993.

DAHER, R. F.; MORAES, C. F.; CRUZ, C. D. Seleção de caracteres morfológicos discriminantes em capim-elefante (*Pennisetum purpureum* Schum.). **Revista Brasileira de Zootecnia**, v.26, p. 247 – 254, 1997.

DIAS, L.A.S. Divergência genética e multivariada na predição de híbrido e preservação de germoplasma de cacau (*Theobroma cacao* L.). Piracicaba, SP: ESALQ. **Tese (Doutorado)** - Escola Superior de Agricultura “Luiz de Queiroz”/Universidade de São Paulo, 94p, 1994.

DILLEWINJ, C. Botany Sugarcane: Waltran : **Chromic Botânica**. 371p, 1952.

JOHNSON, R.A., WICHERN, D. W. Applied multivariate statistical analysis. **Englewood Cliffs: Prentice Hall**. 593 p, 1982.

JOLLIFFE, I.T. Discarding variables in a principal component analysis. I. Artificial data. **Applied Statistics**, v.21, p.160-173, 1972.

JOLLIFFE, I.T. Discarding variables in a principal component analysis. II. Real data. **Applied Statistics**, v.22, p.21-31, 1973

MARDIA, K.V.; KENT, J.T.; BIBBY, J.M. Multivariate analysis. 6.ed. London: **Academic Press**. 518 p, 1997.

MORRISON, D. F. Multivariate statistical methods. 2 ed . Tokyo, **Mc Graw Hill**, 415p, 1981.

PEREIRA, A. V. Utilização de análise multivariada na caracterização de germoplasma de mandioca (*Manihot esculenta* Crantz). **Tese (Doutorado)** - Escola Superior de Agricultura Luiz de Queiroz, Piracicaba. 180p, 1989.

REGAZZI, A. J. Análise multivariada. Viçosa, MG: Universidade Federal de Viçosa, (INF-766) (**notas de aula**). 2002.