

MAURO SERGIO DE OLIVEIRA LEITE

TAMANHO DA AMOSTRA PARA SELEÇÃO DE FAMÍLIAS DE CANA-DE-
AÇÚCAR

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2007

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

L533t
2007
Leite, Mauro Sergio de Oliveira, 1981-
Tamanho da amostra para seleção de famílias de cana-
de-açúcar / Mauro Sergio de Oliveira Leite – Viçosa,
MG, 2007.
xi, 51f. : il. (algumas col.); 29cm.

Orientador: Luiz Alexandre Peternelli.
Dissertação (mestrado) - Universidade Federal de
Viçosa.
Referências bibliográficas: f. 47-51.

1. Cana-de-açúcar - Melhoramento genético. 2. Cana-
de-açúcar - Seleção. 3. Amostragem (Estatística).
4. Simulação (Computadores). 5. Genética quantitativa.
I. Universidade Federal de Viçosa. II. Título.

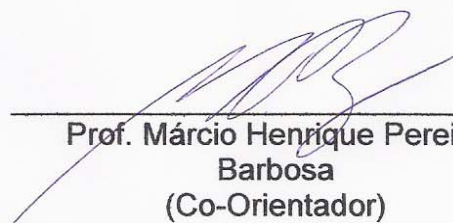
CDD 22.ed. 633.612

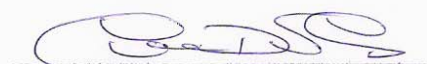
MAURO SERGIO DE OLIVEIRA LEITE

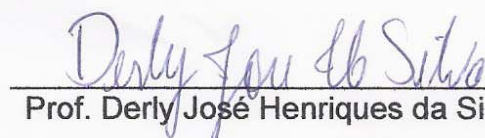
TAMANHO DA AMOSTRA PARA SELEÇÃO DE FAMÍLIAS DE CANA-DE-
AÇÚCAR


Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.


APROVADA: 16 de março de 2007


Prof. Márcio Henrique Pereira
Barbosa
(Co-Orientador)


Prof. Cosme Damião Cruz
(Co-Orientador)


Prof. Derly José Henriques da Silva


Prof. Sebastião Martins Filho


Prof. Luiz Alexandre Peternelli
(Orientador)

Ao meu bom e velho amigo Jesus, que sempre me amparou nas horas mais difíceis e participou das minhas maiores vitórias; e a Deus, por me olhar sempre com um amor incondicional!

*...pela alegria plena dos pequenos rasantes,
tenho certeza de que nasci pra voar...
Nas mãos de Deus posso voar alto!*

AGRADECIMENTOS

A Deus, por me impulsionar e acreditar sempre que eu sou capaz, não importa a situação.

À Universidade Federal de Viçosa e ao Programa de Pós-graduação em Genética e Melhoramento que, além da formação e de todo o conhecimento acadêmico, me proporcionaram crescimento pessoal e um ótimo ambiente de convívio.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudo que tanto nos ajuda.

Ao professor Luiz Alexandre Peternelli, pelo exemplo de responsabilidade e ética profissional, pela grande confiança em meu trabalho e na minha capacidade, por toda a amizade compartilhada nesses bons anos de convivência, pela paciência com os meus erros e pela empolgação com os meus acertos, pela ótima orientação e pelo exemplo pessoal, a minha maior gratidão.

Ao professor Márcio Henrique Pereira Barbosa, por acreditar sempre em meu trabalho e investir nele, pela oportunidade de participar do Programa de Melhoramento da Cana-de-açúcar da UFV, pela amizade, incentivo profissional e por todas as idéias.

Ao professor Paulo Roberto Cecon, por toda a ajuda essencial no decorrer deste trabalho, pela grande amizade e pelo exemplo profissional.

Aos professores da Estatística: Fabyano, Sebastião, Policarpo e Carlos Henrique, pela formação em estatística adquirida nos momentos de discussão sobre a tese no corredor, o meu mais sincero agradecimento.

Ao professor Cosme Damião Cruz, pelo aconselhamento e pela disponibilidade em sempre me atender quando preciso.

Aos meus queridos e grandes pais, Maria Emília e Sérgio Henrique, por me ensinarem a ser corajoso e a enfrentar as situações da vida com força, fé e principalmente alegria, pois na vida tudo passa, mas só o amor fica.

Ao meu grande irmão Paulo Henrique, por me ensinar sempre a alçar vôos cada vez mais altos, e por me incentivar a querer crescer.

Ao Délio, Denilce, Odilon, Taís, Janaína, Aline, Brenão e Dudu, aos meninos do 1412, e à equipe de churrasco, pelo carinho, apoio, amizade e pelos bons momentos de descontração.

Em especial, ao Márcio Pupin, pela valiosa ajuda neste trabalho, pelo companheirismo e pela grande amizade compartilhada em alguns anos de convivência.

A todos os meus companheiros e ex-companheiros de república, aos amigos da capela, da RCC e a todos que aqui não estão citados.

A todos que partilharam comigo de mais esta etapa vitoriosa.

BIOGRAFIA

MAURO SERGIO DE OLIVEIRA LEITE, filho de Sérgio Henrique Muraro Leite e Maria Emília de Oliveira Leite, nasceu em Três Pontas, Minas Gerais, no dia 29 de setembro de 1981.

Em março de 2000, iniciou o curso de Agronomia na Universidade Federal de Viçosa, graduando-se em julho de 2005.

Em agosto do mesmo ano, iniciou o Programa de Pós-Graduação em Genética e Melhoramento na Universidade Federal de Viçosa, em nível de Mestrado, tendo defendido a dissertação em 16 de março de 2007.

SUMÁRIO

	Página
RESUMO	viii
ABSTRACT	x
1. INTRODUÇÃO	1
2. REFERENCIAL TEÓRICO	4
2.1. Seleção em cana-de-açúcar	4
2.2. Tamanho da amostra experimental	6
2.3. Métodos para estimar o tamanho da amostra	8
3. MATERIAL E MÉTODOS	11
3.1. Obtenção dos dados e detalhes experimentais	11
3.2. Características Avaliadas	13
3.3. Procedimentos Estatísticos	14
3.3.1. Análise de Variância	14
3.3.2. Obtenção das amostras	16
3.3.3. Métodos para a determinação do tamanho da amostra ideal	16
3.4. Processamento Computacional	19
4. RESULTADOS E DISCUSSÃO	21
4.1. Caracterização da população	22

	Página
4.2. Estudo das estimativas dos parâmetros.....	25
4.3. Método da Máxima Curvatura Modificado e Intensidade de Amostragem.....	40
5. CONCLUSÕES.....	45
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	47

RESUMO

LEITE, Mauro Sergio de Oliveira, M.Sc., Universidade Federal de Viçosa, março de 2007. **Tamanho da amostra para seleção de famílias de cana-de-açúcar.** Orientador: Luiz Alexandre Peternelli. Co-Orientadores: Márcio Henrique Pereira Barbosa, Cosme Damião Cruz e Paulo Roberto Cecon.

Qualquer que seja a forma de avaliar um experimento de famílias em cana-de-açúcar, figura a questão do número de indivíduos necessários para representar a performance destas famílias. A avaliação do tamanho da amostra ideal para representar precisa e eficientemente a média, variância e outros parâmetros em famílias é necessária e tem sido pesquisada, contribuindo para a otimização dos recursos disponíveis nos programas de melhoramento. Existe certa divergência e falta de consistência nos resultados obtidos para o tamanho de amostra ótimo em cana-de-açúcar. Dado o exposto, objetivou-se neste estudo a determinação do número mínimo de plantas que deve ser amostrado em experimentos de famílias de cana-de-açúcar, para características relacionadas à produção, que possibilite, sem perda de eficácia, a estimação de alguns parâmetros utilizados como auxílio na seleção das famílias. Para tanto, foram utilizados dados de um experimento conduzido no Centro de Melhoramento da Cana-de-Açúcar (CECA) da Universidade Federal de Viçosa, sediado no município

de Oratórios – MG, montado no delineamento em blocos casualizados completos composto por 18 famílias de irmãos completos de cana-de-açúcar em seis repetições e com informação individual de 20 plantas por parcela. Foi realizado um estudo de reamostragem com reposição, com posterior estimação de alguns parâmetros genéticos e fenotípicos. Também foram utilizados os métodos da Máxima Curvatura Modificado e da Intensidade de Amostragem. As estimativas do tamanho da amostra variaram de acordo com o parâmetro a ser estimado, de acordo com a variável a ser avaliada e de acordo com o método utilizado. O método da reamostragem permite uma comparação mais eficiente dos efeitos do tamanho da amostra na estimação de parâmetros genéticos e fenotípicos, o que não acontece nos Métodos da Máxima Curvatura Modificado e da Intensidade de Amostragem. Uma amostra de 16 plantas por família seria suficiente para obter estimativas fidedignas de todos os parâmetros discutidos, para todas as variáveis consideradas. Porém, sendo possível desmembrar a amostragem por característica, uma amostra de 10 plantas por família possibilitaria a estimação bastante precisa dos parâmetros para a variável Brix.

ABSTRACT

LEITE, Mauro Sergio de Oliveira, M.Sc., Universidade Federal de Viçosa, march, 2007. **Sample size for sugarcane family selection.** Adviser: Luiz Alexandre Peternelli. Co-Advisers: Márcio Henrique Pereira Barbosa, Cosme Damião Cruz and Paulo Roberto Cecon.

Regardless the kind of method used to evaluate an experiment of sugarcane families, it remains the question of the number of necessary individuals to represent the performance of these families. The evaluation of the ideal sample size to represent with precision and efficiency the mean, variance and other parameters in families is necessary and has been study, contributing to the optimization of resources available in a breeding program. The literature shows certain divergence and lack of consistency on results obtained for the best sample size in sugarcane. In this context, the aim of this study was to determine the minimum number of plants that should be sampled in experiments of sugarcane families, for yield traits, which would allow an effective estimation of some parameters used to assist the family selection. Data from an experiment carried out at Sugarcane Breeding Center (CECA) at Universidade Federal de Viçosa, in the municipal district of Oratórios – MG, were utilized. The experiment was planned as a randomized complete blocks desing with 18 full-sib families of sugarcane in six repetitions and individual information of 20 plants per plot. It was carried out a study of

resampling with replacement, followed by an estimation of some genetic and phenotypic parameters. Also the Modified Maximum Curvature Method and Sampling Intensity Method were used. The estimates of sample size varied accordingly to the parameter being estimated, to the trait being evaluated and to the method being utilized. The resampling method allows for a more efficient comparison of the sample size effects on the estimation of the genetic and phenotypic parameters, when compared to the Modified Maximum Curvature and Sampling Intensity Methods. A sample of 16 plants per family would be sufficient to obtain good estimates of all the parameters discussed, for all the traits considered. However, in the circumstance of a separate sampling by trait, a sample of 10 plants per family would allow an efficient estimation of the parameters for Brix.

1. INTRODUÇÃO

O melhoramento genético da cana-de-açúcar baseia-se na seleção e clonagem de genótipos superiores de populações segregantes, obtidas por meio de cruzamentos entre indivíduos diferentes. Para maximizar a eficiência desse processo, são realizadas distintas etapas, envolvendo a escolha adequada dos genitores e a quantificação dos efeitos genéticos e ambientais na expressão de cada caráter sob seleção (MATSUOKA et al., 1999). A obtenção de clones altamente produtivos depende da detecção inicial de genitores que possam produzir progênes superiores e da posterior seleção dos melhores indivíduos dessas progênes.

Na tentativa de obter maior eficiência nos programas de melhoramento da cana-de-açúcar, uma nova etapa tem sido incorporada às fases iniciais. Essa etapa, chamada seleção de famílias, consiste em identificar cruzamentos superiores, dentre os vários cruzamentos testados, que serão utilizados para formar a população base sobre a qual se fará a seleção individual. Esta metodologia possibilita também o descarte de progênes inferiores já nas fases iniciais do programa (BASTOS, 2005).

Quando um caráter é controlado por vários genes, situação muito comum para a maioria das características de valor econômico, a

probabilidade de se associar a um indivíduo todos os alelos favoráveis é muito pequena. Várias pesquisas demonstraram que a combinação da seleção de famílias com seleção massal dentro das famílias selecionadas, na primeira fase de seleção dos programas de melhoramento, proporcionou maiores ganhos genéticos e maior frequência de clones superiores (MCRAE et al., 1993; COX et al., 1996).

A seleção de famílias tem sido empregada rotineiramente em alguns programas de melhoramento da cana-de-açúcar de outros países (COX et al., 1996; BRESSIANI, 2001), sobretudo para os caracteres cuja herdabilidade baseada nas médias de famílias tem sido superior à herdabilidade com plantas individuais.

Entretanto, para que a seleção de famílias seja eficiente é necessário conduzir experimentos com boa precisão. Há vários fatores que afetam a precisão experimental, entre eles a heterogeneidade do solo, do material genético e do sistema de manejo (GÓMEZ e GÓMEZ, 1984). Para minimizar os efeitos da heterogeneidade do solo existem algumas estratégias, entre elas a escolha do delineamento experimental, número de repetições, tamanho da parcela e da amostra dos caracteres.

Por outro lado o espaço, mão de obra e o tempo necessários para avaliar plantas individuais para componentes produtivos, por exemplo, são tão grandes que uma metodologia consistente para avaliar progênies em estágios iniciais de seleção seria muito importante (WU et al., 1978). Torna-se então fundamental conhecer o número de plantas que represente as famílias que estariam sendo avaliadas nos ensaios.

Para Wu et al. (1978), se um tamanho de amostra pequeno pode prover informações suficientes da capacidade de combinação de uma progênie ou família, um número bem maior de cruzamentos poderia ser testado todos os anos em etapas iniciais de seleção. Segundo Aguiar et al. (2000), para que os programas de melhoramento genético continuem obtendo sucesso é necessário que, cada vez mais, maior número de famílias seja avaliado, já que as diferenças a serem detectadas são gradativamente menores.

No entanto aumentar o número de famílias avaliadas implicaria em maior tamanho de experimentos e, conseqüentemente, maior demanda de

recursos para instalação e condução dos experimentos. Uma alternativa que contribuiria para não aumentar essas demandas seria a redução do tamanho das parcelas dos experimentos. No entanto, se o número de indivíduos amostrados para representar a família não for adequado, a precisão experimental será menor e, conseqüentemente, a obtenção de estimativas fidedignas de parâmetros genéticos e fenotípicos ficará comprometida.

Referente à cana-de-açúcar, verifica-se na literatura recomendações bastante variadas sobre qual deveria ser a amostra mínima representativa das propriedades genéticas das famílias. As sugestões variam de 14 a 500 plantas por família (WU et al., 1977; WU et al., 1978; MARIOTTI et al., 1981; PEIXOTO e RICCI, 1984; BARBOSA et al., 2001; LEITE et al., 2006). Estas sugestões de tamanhos de amostras variam em função do parâmetro a ser estimado, e também de acordo com o tipo de característica avaliada. Além destas diferentes recomendações, as metodologias empregadas para se estimar o tamanho da amostra ideal têm sido muito variadas.

Novos resultados de pesquisa a respeito do quanto é possível reduzir o tamanho da amostra nos ensaios de famílias seriam de grande importância. A economia em recursos, de forma geral, permitiria a avaliação de um maior número de famílias, ou também o teste das famílias em mais de um local. Com isso seria possível aumentar a população de clones levados às próximas etapas do programa, a probabilidade de seleção de clones superiores e, conseqüentemente, a eficiência de ganho nos programa de melhoramento de cana-de-açúcar.

2. REFERENCIAL TEÓRICO

2.1. Seleção em Cana-de-açúcar

No melhoramento da cana-de-açúcar, o processo de desenvolvimento de novas variedades envolve o que se convencionaram denominar de fases T1, T2, T3, FE e FM, isto é, primeira, segunda e terceira fase de teste, fase experimental e fase de multiplicação clonal, respectivamente. A descrição metodológica empregada em cada fase pode ser encontrada de forma bem detalhada e explicada em Matsuoka et al. (1999) e em Barbosa e Silveira (2000). Cada programa tem estabelecido as melhores alternativas para conduzir o processo seletivo com base nos recursos disponíveis e nas particularidades do sistema de produção de cana nas regiões ou países (HOGARTH et al., 1997; MATSUOKA et al., 1999; COPERSUCAR, 2000; BARBOSA e SILVEIRA, 2000).

Nas etapas iniciais dos programas são realizados muitos novos cruzamentos, cujas sementes constituirão famílias a serem avaliadas sobre o potencial de fornecer clones promissores. Tem sido relatado pelo Programa de Melhoramento da cana-de-açúcar da Austrália situado em Queensland (Bureau of sugar experiment stations - BSES) que o número de clones elite, avaliados nas fases posteriores de teste clonal, tem aumentado

significativamente quando se combina a seleção de famílias com a seleção massal (COX et al., 1996). Este programa avalia entre 350 a 450 famílias por ano em cana-planta, sendo as parcelas colhidas mecanicamente. A seleção de indivíduos é feita no ano posterior, na soca. Normalmente, cerca de 30 a 40% das famílias são selecionadas, e dentro destas famílias seleciona-se 40% dos indivíduos (KIMBENG e COX, 2003).

Hogarth (1971) já relatava maior eficiência, no programa de melhoramento da cana-de-açúcar, com a seleção praticada em famílias. Porém, naquele momento, a ausência de máquinas apropriadas para pesagem e corte dos experimentos inviabilizava pesquisas mais detalhadas.

De fato, aplicar a seleção massal em famílias superiores aumenta a probabilidade de identificar clones elite. Para tanto, faz-se necessário selecionar previamente essas famílias superiores. Tais inferências podem ser obtidas pela avaliação, em delineamentos experimentais, de um grupo de clones que representariam cada família, conforme empregado por Jackson et al. (1998). Outra alternativa consiste em avaliar as famílias em ensaios com repetição. Neste caso, as parcelas seriam constituídas de indivíduos ainda não clonados, os quais, juntos, proporcionariam informações sobre o valor genético das famílias avaliadas, como aplicado por Stringer et al. (1996).

Hogarth et al (1990) inferiram que as médias de famílias possuem maior utilidade que as variâncias dentro das famílias, na identificação daquelas com maior aptidão seletiva. Concluem também que a seleção de famílias não deve ser muito rigorosa, porque podem ser produzidos clones superiores também em famílias de baixa aptidão.

Quanto ao caráter utilizado para avaliar as famílias, Mariotti e Cuenya (1992) demonstraram que o número de colmos é o componente de maior estabilidade entre as mesmas, sendo importante para predizer ganhos esperados com seleção.

Skinner et al. (1987) desenvolveram estudos mostrando a herdabilidade, no sentido amplo, das principais características utilizadas no melhoramento da cana-de-açúcar. Esses estudos basearam-se em dados individuais e de famílias. Para o caráter produção a herdabilidade foi de 0,17 ou menor quando baseada em dados individuais. Com os dados de família

seu valor aumentou para 0,75 na Austrália e 0,48 em Fiji. Para Brix, as estimativas com base em dados individuais foram 0,27 (Havaí) e 0,65 (Austrália), enquanto para famílias estes valores subiram para 0,53 e 0,90 respectivamente. Resultado semelhante foi encontrado por Leite et al. (2006) em avaliação do tamanho de parcela em famílias de cana-de-açúcar. Estes autores encontraram valores de herdabilidade de 0,53 para a variável Brix, 0,61 para número de colmos e 0,73 para tonelada de cana por hectare

Segundo Falconer (1989) a herdabilidade é a propriedade mais importante de um caráter métrico. A função mais importante da herdabilidade no estudo genético é o seu papel preditivo expressando a confiança do valor fenotípico como um guia para o valor genético.

Os resultados mostrados sugerem que, na fase inicial do programa, a seleção individual realizada para Brix é bastante eficiente, enquanto para os demais caracteres a seleção com base nas médias de famílias é mais eficiente, em razão da baixa herdabilidade dos mesmos na seleção individual.

2.2. Tamanho da amostra experimental

Qualquer que seja a forma de avaliar as famílias, ainda figura a questão do número de indivíduos necessários para representar a performance destas famílias. Contribuindo para a otimização dos recursos disponíveis num programa de melhoramento, uma avaliação do número de indivíduos necessários para representar precisamente e eficientemente a média, variância e outros parâmetros nas famílias é também necessária e tem sido pesquisada.

Skinner (1971) sugeriu que 75 seedlings de um cruzamento biparental em cana-de-açúcar seriam satisfatórios na determinação do valor desse cruzamento. Após um estudo teórico sobre seleção nesta mesma cultura, Wu et al. (1977) concluíram que uma amostra de 500 indivíduos era requerida para estimar a variância de uma população, com pelo menos 90% de confiabilidade.

Wu et al. (1978) estudaram a amostra mínima necessária para estimar a média e a variância de uma progênie de cana-de-açúcar utilizando

amostras de 5 a 70 plantas e concluíram que 20 plantas estimam bem a média e que 40 plantas são necessárias para se estimar a variância entre os genótipos das progênes.

Avaliando características de rendimento e qualidade da cana-de-açúcar, Mariotti et al. (1981) utilizaram amostras de 20 a 30 plantas por progênie e concluíram que amostras tão pequenas como 20 plantas foram suficientes para avaliar a média da família para características como qualidade do suco, número de perfilhos, rendimento de cana e peso por colmo.

Assim também Peixoto e Ricci (1984), avaliando tamanhos de amostra de 50 a 200 plantas concluíram que uma amostra de 150 plantas seria suficiente para avaliar número de colmos, brix, produção, altura e diâmetro em cana-de-açúcar.

Em experimento conduzido pelo Programa de Melhoramento Genético da cana-de-açúcar da UFV Barbosa et al. (2001), partindo de uma população de 500 plantas, concluíram que 50 indivíduos seriam suficientes para se estimar a produção de colmos e que, para brix, 10 indivíduos por parcela poderiam ser amostrados para estimar o brix médio das famílias. Em outra pesquisa dentro do mesmo programa de melhoramento, Leite et al. (2006) concluíram que parcelas com apenas dois sulcos (14 plantas) poderiam ser usadas para estimar os parâmetros genéticos necessários em experimentos de famílias.

Em todos os programas de melhoramento de cana da Austrália a seleção de famílias tem sido amplamente utilizada, sendo que na primeira fase de seleção são usadas, como rotina, 20 plantas por família (KIMBENG e COX, 2003).

Como pode ser visto, ainda existe certa divergência e falta de consistência nos resultados obtidos para a amostra ótima em cana-de-açúcar. Para garantir o sucesso do programa de melhoramento, ao se estudar a progênie de um cruzamento é preciso trabalhar com o número mínimo de genótipos que permitam a total expressão da variabilidade para todos os caracteres agrônômicos de interesse (PEIXOTO e RICCI, 1984)

Igue et al. (1991) citam que parcelas ou amostras grandes implicam blocos grandes, e isso faz com que a possibilidade de encontrar blocos

homogêneos diminua. Por outro lado, parcelas pequenas têm as desvantagens de serem muito influenciadas pelas parcelas vizinhas e por fatores casuais, como falhas ou diferenças entre as plantas individuais.

Para Jackson e McRae (2001), aumentando-se o número de repetições, a precisão das estimativas das médias dos clones será aumentada; aumentando-se o tamanho da parcela será aumentada a correlação genética da amostra com o valor populacional e diminui-se o resíduo, porém a variância genética também será menor; enquanto aumentando o número de genótipos ou famílias a serem avaliados permitirá o aumento da intensidade de seleção.

É bem estabelecido o fato de que a precisão das estimativas tende a aumentar com o aumento do tamanho da amostra. Por outro lado, também é bem conhecido que, a partir de certo tamanho de amostra, aumentos adicionais não trazem ganhos compensatórios na precisão (SILVA et al., 1993).

A amostragem é uma técnica amplamente utilizada no estudo de populações, decorrente das vantagens que este processo proporciona como o menor custo e rapidez na obtenção e análise dos dados (BRAGA, 1986). Porém o tamanho da amostra em experimentos agrícolas é um dos principais problemas a ser definido pelos pesquisadores (STUKER e BOFF, 1998). A necessidade de identificação do melhor tamanho de amostra é evidente nos vários ramos da pesquisa, principalmente nos casos onde a capacidade de detecção de pequenas diferenças é essencial para o sucesso na identificação de técnicas mais adequadas.

2.3. Métodos para estimar o tamanho da amostra

Na literatura são relatadas as mais diversas maneiras para a estimação do tamanho de amostra ideal são relatadas na literatura (FEDERER, 1955; MÉIER e LESSMAN, 1971; WU et al., 1977; WU et al., 1978; PEIXOTO e RICCI, 1984; XIE e MOSJIDIS, 1999; JACKSON e MCRAE, 2001; BARBOSA et al., 2001; VIANA et al., 2002; LÚCIO et al., 2003; entre outros). Ao longo dos anos as metodologias foram se tornando cada vez mais eficientes e específicas, utilizando-se de ferramentas

estatísticas acuradas que possibilitam a estimação precisa de parâmetros de interesse do pesquisador.

Em trabalho com tamanho de amostra para algumas características do milho, Silva et al. (1993) utilizaram o coeficiente de variação e uma comparação entre estimativas de variâncias residuais das análises realizadas nos diversos tamanhos de amostra propostos, para definir a amostra ideal.

Xie e Mosjidis (1999) utilizaram o método bootstrap não paramétrico como uma alternativa para estimar o erro padrão e intervalos de confiança para a correlação genética em trevo vermelho. Através deste método puderam medir como o erro de amostragem é afetado pelo número de repetições, de famílias e de plantas por família utilizados e também puderam determinar, empiricamente, o tamanho ótimo do experimento para a estimação da correlação genética com precisão.

Amostrando o peso de frutos de abóbora italiana por fileira, a metodologia apresentada por Souza et al. (2002) para determinação do tamanho da amostra ideal faz uso do coeficiente de variação experimental (CV%), de semi-amplitudes do intervalo de confiança em porcentagem da média (D%) e de valores tabelados da estatística t de Student. Utilizaram também a correção para o tamanho final da amostra por se tratar de estudo de uma população finita de plantas.

Com o objetivo de estimar o tamanho ideal da amostra para massa fresca de frutos na cultura do pimentão, Lúcio et al. (2003) utilizaram a mesma metodologia citada em Souza et al. (2002), porém usando o termo “intensidade de amostragem” para designá-la.

Na determinação do tamanho da amostra para componentes de produção em cana-de-açúcar, Peixoto e Ricci (1984), propondo diferentes tamanhos de amostra como sendo os tratamentos, realizaram apenas uma análise de variância seguida pelo teste de Tukey e avaliaram as diferenças existentes entre as médias obtidas para cada tamanho de amostra, visando demonstrar o número de plantas necessárias para haver a estabilização da média.

Barbosa et al. (2001), também em experimento com cana-de-açúcar, propondo dois modelos diferentes para a análise de variância, sendo os

efeitos de família e de amostra dentro das famílias considerados fixos, prosseguiram com uma análise de variância para cada tamanho de amostra e estimação da correlação simples e parcial entre a produção total e a produção estimada pela amostra. Mais recentemente, Leite et al. (2006) utilizaram o método de máxima curvatura, proposto por Federer (1955), para determinar o tamanho da parcela ideal em experimentos de famílias em cana-de-açúcar. Considerando o efeito de famílias como aleatório, compararam também o efeito do tamanho da parcela nas estimativas de parâmetros genéticos. Porém ambos os estudos utilizaram dados totais por sulco e não dados de plantas individuais.

Segundo (RESENDE e BARBOSA, 2006), a melhor estratégia de seleção em cana-de-açúcar, seria através dos valores genéticos preditos pelo BLUP individual (RESENDE, 2002), que utilizaria simultaneamente informações de famílias e de plantas individuais para a seleção. Entretanto, este método não tem sido usado no melhoramento de cana devido às dificuldades em se obter dados de plantas individuais. Seria então de grande importância propor um tamanho de amostra pequeno e eficiente que possibilitasse o uso do BLUP individual.

Dado o exposto acima, ainda é importante definir em quanto é possível reduzir o tamanho da amostra nos ensaios de famílias, pois, além de reduzir custos e possibilitar que novas metodologias sejam aplicadas no programa, poderia ser testado um maior número de famílias com o mesmo recurso disponível. As razões para isto incluem o desejo de se avaliar grandes populações de clones, aumentando a probabilidade de identificação de clones elite.

Sendo assim, foi objetivo deste estudo a determinação do número mínimo de plantas que deve ser amostrado em experimentos de famílias de cana-de-açúcar, para características relacionadas à produção, que possibilite, sem perda de eficácia, a estimação de alguns parâmetros utilizados como auxílio na seleção das famílias.

3. MATERIAL E MÉTODOS

3.1. Obtenção dos dados e detalhes experimentais

Foram utilizados como genitores para o experimento clones elite e variedades comerciais em 22 cruzamentos biparentais direcionados entre clones produtivos e de maturação precoce de acordo com a Tabela 1.

Os cruzamentos foram realizados na Estação de Floração e Cruzamento da Serra do Ouro da Universidade Federal de Alagoas – AL, sediada no município de Murici – AL, geograficamente localizado a 09°13'S e 35°50'W, a 450m de altitude, com precipitação de 2.363mm anuais e URA de 93,2%. As sementes provenientes desses cruzamentos foram enviadas e plantadas no Centro de Melhoramento da Cana-de-Açúcar (CECA) da Universidade Federal de Viçosa-MG, sediado no município de Oratórios-MG, geograficamente localizado a 20°25'S e 42°48'W, com 494m de altitude, precipitação média anual de 1.250mm, URA 64,7%, e temperatura média máxima anual de 21,8°C e mínima anual de 19,5°C.

Tabela 1 – Relação dos clones e variedades utilizadas como genitores nos cruzamentos biparentais e as respectivas famílias de irmãos completos que foram geradas

Fêmea	Macho	Código da Família
RB768647	TUC717	14
RB835486	RB855453	19
RB835486	SP77-5181	22
RB855046	IAC862210	33
RB855156	RB855595	38
RB855156	RB925345	40
RB855156	SP80-3280	44
RB855511	RB855156	48
RB855595	RB855156	62
RB925211	RB835486	75
RB925211	RB8495	76
RB925345	RB855156	78
RB92579	TUC717	82
RB945954	RB835487	97
RB855595	RB945961	98
RB946903	RB925345	101
RB835054	RB931555	109
RB855046	RB966928	110
SP77-5181	RB835486	112
SP77-5181	SP80-1816	116
SP80-1816	RB925345	121
SP80-1816	SP77-5181	123

A topografia e o tipo de solo onde o experimento foi montado são, respectivamente, baixada e Latosolo vermelho-amarelo.

O experimento foi plantado no dia 12 de abril de 2004 no sistema de cana de ano e meio utilizando o delineamento em blocos casualizados completos, com 6 repetições e 24 parcelas, sendo 22 famílias e 2 testemunhas comuns: as variedades RB72454 e RB855156. A obtenção e transplântio dos “seedlings” foi realizada conforme descrito por Barbosa e Silveira (2000).

O experimento totalizou 144 parcelas em uma área total de 2,9 ha. Cada parcela foi composta por dois sulcos de 5 m, espaçados de 1,40 m,

sem bordadura, sendo cada sulco constituído por dez plantas com espaçamento de meio metro entre elas. Foi deixado um aceiro de 2 m entre as extremidades dos sulcos de parcelas diferentes. Foram utilizados dois sulcos da variedade RB855156 plantados ao redor do experimento como bordadura lateral. A adubação utilizada foi 500 kg/ha da fórmula 05-25-25 no plantio.

3.2. Características avaliadas

A coleta dos dados foi realizada em cana planta em julho de 2005 de acordo com os procedimentos descritos a seguir.

1. Medidas de brix tomadas em um colmo por planta utilizando refratômetro de campo, com amostragem feita no 5^o entrenó, a contar da base para o ápice.
2. Contagem do número total de colmos por planta;
3. Medidas de diâmetro do colmo tomadas com paquímetro no mesmo entrenó amostrado para leitura de brix.
4. Medidas de comprimento do colmo tomadas no mesmo colmo amostrado para leitura de brix. As leituras foram feitas com uma régua de madeira articulada de metro em metro, para facilitar o acompanhamento da curvatura do colmo;
5. Pesagem de toda a parcela em balança tipo dinamômetro, após o desponte dos ponteiros dos colmos.

Realizada a coleta dos dados, fez-se necessário obter as variáveis que iriam compor as análises. Das medidas individuais de brix obteve-se a variável Brix. Do número total de colmos por planta, obteve-se a variável número de colmos (NC). Utilizando as medidas de número de colmos (NC), diâmetro do colmo (DC) e comprimento do colmo (CC), foi estimado o peso de cada planta em Kg, pela fórmula $PE = \pi \times NC \times CC \times (DC/2)^2 \times d$ (Chang; Milligan, 1992), considerando-se o volume de um colmo igual ao do cilindro, e a densidade $d = 1 \text{ g/cm}^{-3}$. Estimou-se a tonelada de cana por hectare

(TCH), utilizando a expressão: $(PE \times 10)/0,7$ sendo 0,7 a área útil de cada planta em m^2 . A tonelada de Brix por hectare (TBH) foi obtida pela equação: $(TCH \times \text{Brix médio})/100$.

Logo, no total, foram analisadas sete variáveis, a saber: número de colmos (NC), diâmetro do colmo (DC), comprimento do colmo (CC), peso estimado de colmos (PE), brix (Brix), tonelada de cana por hectare (TCH) e tonelada de Brix por hectare (TBH).

3.3. Procedimentos Estatísticos

3.3.1. Análise de Variância

A análise de variância foi conduzida considerando o seguinte modelo estatístico (CRUZ et al., 2004):

$$Y_{ijk} = \mu + G_i + B_j + \varepsilon_{ij} + \delta_{ijk},$$

sendo:

Y_{ijk} : observação obtida no k-ésimo indivíduo da i-ésima família avaliada no j-ésimo bloco;

μ : média geral do experimento;

G_i : efeito aleatório da i-ésima família;

B_j : efeito aleatório do j-ésimo bloco;

ε_{ij} : efeito aleatório da variação entre famílias;

δ_{ijk} : efeito aleatório da variação entre plantas dentro da família.

O esquema da análise de variância e as esperanças dos quadrados médios para o modelo estatístico apresentado anteriormente, em que se consideram as informações de indivíduos dentro da parcela (CRUZ et al., 2004), está apresentado no Quadro 1.

Quadro 1 – Esquema da análise de variância para o delineamento em blocos casualizados completos mostrando fonte de variação (FV), graus de liberdade (GL) e esperanças de quadrado médio (E(QM))

FV	GL	QM	E(QM)
Blocos	b -1	QMB	$\sigma_d^2 + n\sigma_e^2 + ng\sigma_b^2$
Famílias	g -1	QMF	$\sigma_d^2 + n\sigma_e^2 + nb\sigma_g^2$
Entre famílias	(b -1)(g -1)	QME	$\sigma_d^2 + n\sigma_e^2$
Dentro de famílias	(n -1)bg	QMD	σ_d^2

A partir da análise, foram calculadas as estimativas dos seguintes parâmetros: variância de blocos ($\hat{\sigma}_b^2$); variância genotípica entre e dentro das famílias ($\hat{\sigma}_g^2$); variância ambiental entre médias de famílias ($\hat{\sigma}_e^2$); variância fenotípica entre médias de famílias ($\hat{\sigma}_f^2$); variância fenotípica dentro da família ($\hat{\sigma}_d^2$), coeficiente de variação experimental (%CVe), coeficiente de variação genético (%CVg), relação CVg/CVe, herdabilidade entre (h_{entre}^2) e herdabilidade dentro (h_{dentro}^2). Os estimadores utilizados para estes parâmetros são descritos a seguir:

$$\hat{\sigma}_b^2 = \frac{QMB - QME}{ng} \quad \hat{\sigma}_g^2 = \frac{QMT - QME}{nb} \quad \hat{\sigma}_e^2 = \frac{QME - QMD}{n}$$

$$\hat{\sigma}_f^2 = \frac{QMT}{bn} \quad \hat{\sigma}_d^2 = QMD \quad \%CVe = \frac{\sqrt{\hat{\sigma}_e^2}}{\bar{X}} \times 100 \quad \%CVg = \frac{\sqrt{\hat{\sigma}_g^2}}{\bar{X}} \times 100$$

$$h_{\text{entre}}^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_f^2} \quad h_{\text{dentro}}^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_d^2} \quad CVg/CVe = \frac{\%CVg}{\%CVe}$$

3.3.2. Obtenção das amostras

Como primeira etapa foi necessária a obtenção das amostras a serem comparadas, considerando como unidade básica cada planta. O procedimento utilizado para a simulação dos conjuntos de dados foi semelhante à técnica de reamostragem “bootstrap” (DAVISON e HINKLEY, 1997), e se resumiu nos seguintes passos:

- (i) De cada parcela dos dados originais foram retiradas amostras aleatórias de tamanho n , de maneira que as plantas amostradas nunca fossem repetidas em uma mesma amostra. Foram testados oito tamanhos de amostra, com n igual a 4, 6, 8, 10, 12, 14, 16 e 18 plantas;
- (ii) Reamostragens com reposição foram realizadas 500 vezes (XIE e MOSJIDIS, 1997 e XIE e MOSJIDIS, 1999), seguindo o exposto no item (i), gerando assim 500 novos conjuntos de dados para cada valor de n em cada variável avaliada;
- (iii) para cada conjunto de dados gerado realizou-se a análise de variância segundo o modelo definido no item 3.3.1. e estimaram-se os parâmetros de interesse, que foram armazenados em novos vetores de estimativas correspondentes.

3.3.3. Métodos para a determinação do tamanho da amostra ideal

Vários são os métodos empregados para estimar o tamanho da amostra experimental apresentados na literatura, variando quanto à sua complexidade. Foram utilizados alguns dos procedimentos que mais se adaptaram ao conjunto de dados disponível.

3.3.3.1. Média, erro-padrão e intervalo de confiança para as estimativas dos parâmetros

Para este procedimento foi realizada uma adaptação da metodologia apresentada por Xie e Mosjidis (1997) e Xie e Mosjidis (1999), obtendo, com isso, as estimativas da média e do erro-padrão para os vetores de cada um dos parâmetros propostos no item 3.3.2. Foram construídos os intervalos de

confiança para os valores dos parâmetros estimados no conjunto de dados original (20 plantas) de acordo com a metodologia apresentada por Barbin (1993), exceto para as herdabilidades.

Nas estimativas dos intervalos de confiança para herdabilidade entre as famílias, foi utilizada a expressão apresentada por Knapp et al. (1985). Conforme esses autores esta herdabilidade é definida como $1-(\theta_2/\theta_1)$, e os intervalos podem então serem estimados pela equação $P\{1-[(M_1/M_2)F_{1-\alpha/2;g|2,g|1}]^{-1} \leq 1-(\theta_2/\theta_1) \leq [(M_1/M_2)F_{\alpha/2;g|2,g|1}]^{-1}\} = 1-\alpha$. Para o presente trabalho, M_1 foi o valor estimado para o quadrado médio do tratamento (QMT) e M_2 foi o valor estimado para o quadrado médio do erro entre parcelas (QME). θ_1 e θ_2 foram os reais valores dos quadrados médios do tratamento e do erro, respectivamente, e α é o nível de significância.

Para estimar os intervalos de confiança para a herdabilidade dentro das parcelas, foram feitas 1000 reamostragens do conjunto de dados original e obtidas as 1000 estimativas para este parâmetro. Em seguida foi construído um gráfico da distribuição destas estimativas e foram escolhidos os valores relativos aos quantis de 2,5% e 97,5% para os limites inferior e superior, respectivamente, do intervalo de confiança.

As variações ocorridas na dispersão das estimativas desses parâmetros foram discutidas com base nos intervalos de confiança obtidos.

3.3.3.2. Intensidade de amostragem

Outro procedimento para a estimação do tamanho ideal da amostra leva em conta a estimativa da média aritmética (\hat{m}), da variância (S^2) e do coeficiente de variação experimental em porcentagem (CVe%), estimados para cada conjunto de dados nos diferentes tamanhos de amostra. De posse destas estimativas foi definida a intensidade de amostragem (n_0) para as semi-amplitudes do intervalo de confiança (D%) de 10 e 20% (SOUZA et al., 2002; LÚCIO et al., 2003). Foi aplicado o teste de Bartlett (STEEL e TORRIE, 1997) para verificar se houve homogeneidade de variância entre os tamanhos de amostra.

Na determinação do %CVe, calculado pela fórmula mostrada a seguir, foram utilizadas as variâncias ponderadas, quando estas foram

homogêneas, e a maior variância nos casos em que as mesmas não apresentaram homogeneidade entre os diferentes tamanhos de amostra.

$$\%CVe^* = \frac{100\sqrt{S^2}}{\hat{m}},$$

*este cálculo não leva em consideração o CVe% apresentado com base na análise de variância proposta no item 3.3.1.

sendo,

S^2 : média ponderada das variâncias ou variância máxima;

\hat{m} : média geral.

A intensidade de amostragem (n_0), obtida de acordo com a fórmula apresentada por Cochran (1977), é determinada por:

$$n_0 = \frac{t_{\alpha/2}^2 (\%CVe)^2}{(D\%)^2},$$

sendo,

D%: a semi-amplitude do intervalo de confiança (10 e 20%) da média;

%CVe: coeficiente de variação experimental em porcentagem proposto acima;

t: valor da estatística t da distribuição t de Student.

Em se tratando de população finita igual a 20 plantas, foi aplicada uma correção para populações finitas conforme recomendado por Cochran (1977), a fim de se obter o tamanho final da amostra (nf), pela fórmula:

$$nf = \frac{n_0}{1 + \frac{n_0}{N}},$$

onde N é igual 20 plantas.

3.3.3.3. Método de máxima curvatura modificado

Foi também aplicado o tradicional método da máxima curvatura modificado (MEIER e LESSMAN, 1971) que consiste em representar a relação entre coeficiente de variação experimental (%CVe) e tamanho da parcela ou amostra, com uso de uma equação de regressão do tipo $y = a/x^b$, em que y representa o coeficiente de variação e x o correspondente tamanho de amostra em unidades básicas. Para este trabalho, foi usada a função $CV = aX^b$, em que o valor da abcissa, no ponto de máxima curvatura, é dado pela seguinte expressão apresentada por Chaves (1985), deduzida a partir de Meier e Lessman (1971):

$$X_{MC} = \left[\frac{a^2 b^2 (2b - 1)}{(b - 2)} \right]^{\frac{1}{(2-2b)}}$$

sendo,

X_{MC} = valor da abcissa correspondente ao ponto de máxima curvatura, ou seja, o tamanho da amostra recomendado;

a = constante da regressão;

b = coeficiente da regressão.

3.4. Processamento computacional

Devido ao grande volume de dados e análises a serem realizadas, foram desenvolvidos algoritmos usando a linguagem R de programação (R DEVELOPMENT CORE TEAM, 2005) para automatização dos procedimentos de simulação, análise e estimação de parâmetros propostos neste trabalho, bem como para a aplicação dos métodos propostos.

A linguagem R é uma linguagem orientada a objetos criada em 1996 por Ross Ihaka e Robert Gentleman que, aliada a um ambiente integrado, permite manipulação de dados, realização de cálculos e geração de gráficos, semelhante a linguagem S desenvolvida pela AT&T's Bell Laboratories e que já é utilizada para análise de dados, mas com a vantagem de ser de livre

distribuição. O programa, disponível na internet (<http://www.r-project.org>), que utiliza essa linguagem também recebeu o mesmo nome R.